

五十年來台灣人文學術研究叢書----文獻學與圖書資訊學

# 台灣地區中國古籍文獻資料數位化 的過程與未來的發展方向

元智大學·羅鳳珠

2000年11月，學生書局出版

## 一、前言

網際網路的發明，無疑的是近代文明發展的一件大事，資訊科技的影響力，也在這幾年之內涵蓋全人類的每一個層面，每一個角落。各學門、各行業引用資訊科技作為輔助工具，也幾乎到了無所不及的地步。文史學門引用資訊科技作為輔助工具，應該是各學門之中起步較晚的一個領域。

中文資訊技術有計畫的應用在中文文獻資料的處理上，當屬中研院於1984年7月開始的「史籍自動化計畫」為最早，該計畫為開發《廿五史》全文資料庫而成立，其基本目的是選擇對中國傳統人文研究具有重要價值的古代文獻，建立電腦全文資料庫，作為學術研究的輔助工具。在輸入文獻的選擇上，以中研院研究人員的專長與興趣為主要考量，第一年先將部份的《食貨志》建立全文機讀檔案，第二年持續將《廿五史》全部的《食貨志》建檔，在功能上也往前推進一大步，其後《廿五史》全部資料亦陸續建立。

網際網路的技術引進國內之後，筆者首先於1993年將《紅樓夢》、《全唐詩》正式上網，隨後中研院《廿五史》、《諸子百家》資料庫，台灣大學佛學資料庫上網。網路能傳輸多媒體資訊之後，筆者再將《紅樓夢》以多媒體形式上網。經過短短的五年，現在全球網際網路上由台灣地區研發的中國典籍文獻資料，已有數億字，也具備各種檢索功能，各種媒體資料，呈現一片欣欣向榮的景象。1998年起，由國科會主導的「數位博物館專案先導計畫」〔1〕，邀請中研院等單位進行數位博物館相關計畫，將文獻資料數位化的研發工作，從文字資料擴充到多媒體文獻資料，使用者定位於中小學生，企圖達到往下紮根的目的；1999年進行第二年度的「數位博物館計畫」〔2〕將範圍再次擴充，2000年國科會再次委託中央研究院全面規畫「國家數位典藏計畫」〔3〕，目前正在規畫中，預料可

以將文獻數位化的工作進行全方位的研發，中研院並且於 2000 年舉辦的第三屆國際漢學會議特別增加「漢籍數位典藏研討」議題，專門就「數位化的語文工具」邀請學者發表論文，中國古籍文獻資料數位化的工作，可說是全面有計畫的展開了。

中國古籍文獻數量龐大，在數位化的過程中，輸入方法與中文字碼不足是兩個大問題。中文的輸入、輸出、編輯、排版、全文檢索、編碼等技術，中研院張仲陶教授、謝清俊教授從 1980 年開始便已指導學生進行中文資訊有關文字問題的研究，奠立了基礎。經過近二十年的努力，文字輸入已有多種不同的工具供使用者選擇，鍵盤輸入之外也發展 OCR (optical character recognition) 技術，以滿足龐大的輸入需求。大五碼字種數不足的問題，從早期中研院於開發《廿五史》資料庫時，遇到缺字，逐一造字，以擴大字集的方式解決，到現今中央研究院文獻處理實驗室的「漢字字形資料庫」改用「以部件構字」的觀念，從缺字的輸入、檢索及呈現三方面解決，為解決中文缺字問題，又往前邁進一大步，對於中國古籍文獻數位化的工作貢獻極大。

中國古籍文獻數位化工作的主要目的有典藏、流通、研究、教學四方面，典藏與流通以保存與傳揚文獻為目的，研究與教學期使古文獻為現代人所用；典藏與流通重在管理，研究與教學重在應用，因此，如何設計出符合文史學界使用，使古籍數位化資料成為文史學界教學與研究的輔助工具，以提升其使用效能，成為古籍數位化方向的重要指標。

本文首先嘗試就文史學界與電腦學界對古籍數位化方向的交集與期待，為古籍數位化勾勒出一幅符合使用者需求的藍圖，其次再分述古籍數位化的幾個階段，最後提出未來的展望，以為將來發展的方向與目標。

## 二、文史學界與電腦學界對古籍數位化方向的交集與期待

筆者於 1987 年應《國文天地》編輯的邀請，以〈探一探文史資料自動化的路〉〔4〕為題，訪問電腦界張仲陶教授，文史界周何教授（經部）、毛漢光教授（史部）、王邦雄教授（子部）、王熙元教授（集部）等五位學者，從電腦界、文史界的角度，提出他們對古籍數位化的看法，並從中尋找一條文史資料數位化的路。張仲陶教授首先提出「不要問電腦能做什麼？而是問你要電腦做什麼？」的看法。張教授認為：「平面的資料輸入電腦，出來後還是平面的資料，頂多省去帶書的麻煩而已，我們要讓他具備其他的功能，也就是如何從中摘取需要的資料，這部份由電腦界負責設計程式，但文史界必須告訴我們「需要什麼？」文史界負責提出「需要什麼？」電腦界負責「怎麼滿足需要？」二方面配合起來，就可以做。……我們不希望文史界的人再花時間去學電腦，我們也沒有餘力從頭精

研文史。電腦運用的技巧我們知道，但要用在什麼地方，由各行各業，各學門的專家來決定，所以不要問『電腦能做什麼？』而是問『你要電腦做什麼？』這才是關鍵所在。」〔同4〕

文史學界的周師一田教授從經學的角度提出經學數位化首重訓詁資料的整理：「經學比較偏重思想方面，需要去體認、去領悟，以電腦目前的功能而言，並不能十分有效的處理思想層次的問題。所以，希望文史字義等訓詁方面的資料，能按時代分類，輸入電腦，才能很方便查到每一個字在各個時代的正確意義。……假使我們能透過電腦分析每一個字在各個時代的習慣用法及賦予的意義，便能更正確掌握經學文字的意義，進而瞭解經學內涵。……如果能由電腦來做字義的時代分類，很多問題都能解決。這麼做也許把電腦的使用功能縮得很小，但這是一個基礎，把文字字義都整理好，蒐集在資料庫，對研究所有中國文獻都是很重要的基礎工作。」周師還進一步提到以訓詁資料為基礎，再進而做考據辨偽與輯佚的工作，等到「電腦的功能越來越大，能處理較形而上的問題時，希望在經學意識觀念方面能提供一些消息（分析、判斷的能力）。……電腦是很呆板的東西，但怎樣使他具有高層次的功能，幫助人腦體會，這是我所期望的。」〔同4〕周師所提出的，其實就是現在所說的人工智慧的一部份。

從史部提出觀點的毛漢光教授，參與第一期的史籍自動化計畫，負責《食貨志》資料的分析、規畫、系統分析，比其他幾位教授，多了實際參與文史資料數位化的實務經驗，毛教授說：「就個人經驗言，在文史自動化的過程中，成敗的關鍵在文史界，不在電腦界，電腦本身很刻板，輸入什麼，便印出什麼。……電腦雖然替我們解決很多問題，他畢竟不是人，不能代替人腦，不能替人思考，一個是技術層面，一個是思想層面。」毛教授以參與《食貨志》數位化的經驗，進一步說：「我相信以眼前已有的基礎，文史界若肯通力合作，由文史界負責九分，一分交給電腦，文史自動化的工作，很快便可以完成。……（電腦）功能方面，依我的經驗，個人想到的，只要分析出來，電腦都可以做到。分析很重要，電腦不是萬能，還是得靠人腦控制。……所以文史自動化成敗的關鍵在文史界，不在電腦界。」〔同4〕

從子部提出觀點的王邦雄教授則語重心長的強調「文史自動化不能失去人的主導地位」，王教授說：「科學是『新的發現』，人文是『新的洞見』，是從內在生命裡發出的智慧之光，這要經年累月的孕育才能產生。……電腦畢竟不是人，無法做創發性的工作。……所以我很擔心，假使我們的學生很容易從電腦中得到資料，他們還會不會下工夫把原典一本一本的讀進心靈中、生命中，並且不斷去感受、去實踐，透過自己的體驗去發現新的東西。站在人文的立場，進行電腦化之前，必須先有這層顧慮與共識。」〔同4〕雖然有這一層顧慮，王教授認為可以「將記憶性的資料由電腦取代人力，學子們利用省下來的時間去思考、去發展學術。」王教授還進一步說：「文史資料電腦化之後，儘管有危機在，仍然可以事先防範，只要回歸到生活，除了思考力、洞察力的培養外，還要有生命的體驗，智慧的透顯。然後要認清任何資料都有其侷限性，有了資料，並不等於有

學問……無論電腦如何進步，人都居於主導地位，人若失去主導地位，生命便沒有意義……所以我認為必須回歸到生活，由人去運用電腦，人去發現意義，掌握相生的機會，這樣無論電腦如何進步，人才不會恐慌，要不然電腦越進步，人越後退，人文越萎縮，到最後由電腦來統治世界，那就斷了做學問的根本了，因此必須先做心理建設，認清人永遠是處於主導地位來運用電腦。」〔同4〕

從集部提出觀點的王師熙元教授則認為：「電腦發展至今，不僅可以處理一些機械性和數理方面的資料，人文學科中許多資料經過分析、歸納以後，也可以由電腦來處理。」王老師以為「研究工作最重要的是資料的運用，假如工具書不完整，資料又太龐雜，人力上浪費太多時間找資料，學術的進步必定很緩慢。」王老師並以文學批評為例，提出有系統的整理歷代詩話、詞話、曲話、文話、賦話等文學批評資料，建立有系統的文學批評資料庫，有助於文學的研究。其次為古典文學裡典故的用法繁複多樣，增加後人研究的困難，如能參考《藝文類聚》、《太平御覽》、《太平廣記》、《古今圖書集成》等類書，分門別類建立資料，以供研究時檢索參考。此外古典文學裡象徵和比喻技巧的應用，如能經過分析、整理、歸納，建立資料庫，才能符合文史研究的需要。電腦一旦介入文學界，我們如何善用電腦的長處，並避免可能帶來的隱憂，以及預防其中的弊端？王老師進一步提出：「用電腦處理資料，是為處理及應用方便，並節省後人蒐集資料的時間，從這個角度看，非常值得發展，但並不是有了電腦，一切人為的工夫便可以省下來，事先周密的考量以及設計非常重要，畢竟用電腦處理的最大目的還是在將來運用的方便。」〔同4〕

由上述可知，電腦學界與文史學界具有三個共同的觀點，其一為：電腦不能取代人腦；其二為：引用電腦節省處理資料的時間，人可以做更多思考性、創發性的工作；其三為：古籍數位化成敗的關鍵在文史學界，不在電腦學界，需要由文史學界提出需要，電腦學界滿足需要，二者的通力合作才是古籍數位化成功的關鍵。

從1987年到現在已有十二年的時間，這十二年當中，文史資料數位化的範圍從《食貨志》、《四史》擴充到包含經、史、子、集範圍的中國典籍資料，資料的形式也從平行的原文輸入資料進步到經過整理、分析、歸納的資料，如「中研院平衡語料庫」。資料的內容與媒體，也從原典增加到原典與周邊研究資料結合，文字與影像、聲音等多媒體資料並存，如【紅樓夢多媒體網路資料中心】、【漢代的墓葬與文化】。從文字的形、音、義，結合文字學、語言學、文學等方面的知識，從中建立語文知識網路的觀念也有初步的成果，如【國科會數位博物館先導計畫---搜文解字語文知識網路】。數量上累積到數億字，參與的單位從中研院到其他學術單位以及學者個人，使用者更是遍及全球。回顧十二年前這五位學者所提出的觀點，確實都屬於先知卓見。這十二年來，電腦學界的進步，何止是一日千里，但反觀文史學界，由於參與的人顯然是比電腦界少，還有很多成長的空間，亟待文史學界更多人投入。

### 三、中國古籍文獻資料數位化的過程

張仲陶教授在同一篇訪問提出「電腦是爲了處理資料而設計，資料是一種事實的記錄，除了文字之外，還可以用符號、圖片、錄音、錄影等方式記錄，用中文記錄的是中文資料，用英文記錄的是英文資料，電腦只是處理這些資料的工具。對他而言，沒有中英文之別，所以並沒有一個電腦是爲中文而設計。……但中文電腦發展的過程中，卻有幾個問題，一個是中文打字機的問題（即輸入與輸出）……一個是中文編碼的問題。」〔同4〕張教授這一段訪問，已是「史籍資料自動化」開始的第三年，在此之前，電腦學界爲中文資訊技術已經做了幾年的努力，奠定了相當的基礎。

中文資訊方面比較有系統的研究始於1972年，由國科會所倡導，隨後國內幾所大學與研究單位陸續進行相關的研究工作，廠商亦參與相關設備的開發，使中文資訊從學術研究階段進展到實務應用階段。

1991年以來，電腦環境的變化非常大，中國典籍資料庫也不斷的建立。尙未有電腦網路時，這些資料庫都以單機版製作，電腦網路發展之後，在網路上傳輸的資料庫也陸續建立。資料庫的內容也從原典資料擴充到研究資料、教學資料、多媒體資料。資料庫的內容方面也逐漸發展爲跨領域的整合型資料庫，多元媒體的綜合資料庫，以及具有人工智慧的語文知識網路。在功能方面，有純粹提供研究查詢使用的原典資料庫，有純粹提供教學與自學使用的教學資料庫，有教學與研究並用的綜合資料庫。在類別方面，從資料內容、建置形式、涵蓋媒體、使用功能等方面觀之，可以劃分爲含有全文檢索功能，包含原著典籍及注疏資料的資料庫，有包含音樂、影像等多媒體資料的多媒體資料庫，以工具書爲主的工具書資料庫，以研究論著資料爲主的圖書期刊資料庫，以教學爲主的教學資料庫，以文物書畫爲主的資料庫、以創作爲主的文學創作網站。各式各樣的文獻資料，或以網路版、或以光碟版傳承著中國的數位文獻資料。漢代發明紙張，爲中國文獻的保存與傳揚，提供重大的貢獻；從漢以前、漢代到當代，從竹簡、紙張到無紙的數位訊息，二千年之後，資訊科技的發展，爲文獻的保存與傳揚，提供了更便捷的保存方式、更快速的傳遞媒介，其中的影響力，不容小覷。綜觀十多年來的發展，可以約略分爲五個時期，分述如下：

#### （一）第一期：處理中文文字資料時期

1982年，有前中研院計算中心主任張仲陶教授指導，張永銘撰著的《中文書籍自動拼版系統之設計》、張仲陶教授指導，鄭一雄撰著的《中文字形輸出系統的設計》、張仲陶教授、謝清俊教授共同指導，曾士熊撰著的《中國文字特性資料庫的設計》等三篇碩士論文發表。這三篇論文從三個不同的角度研究，互相支援，試圖共同解決中文資訊的輸出、排版、資訊交換的問題。中文輸入的問題，則有1983年張仲陶教授、謝清俊教

授共同指導，王義科撰著的《中文文書編輯系統之研製》、高天助撰著的《國字資料庫之維護》等碩士論文；1984年有張仲陶教授、謝清俊教授共同指導，潘敏政撰著的《在時間域作中文語音合成的研究》、郭明仁撰著的《辦公室用的中文印製系統之設計》等碩士論文。〔5〕

除了學術研究之外，研究單位與產業單位也共同進行中文字的檢字法與編碼的研發工作，大致上可以歸納為五種檢字法：「一、部首及筆劃檢字法，二、筆順檢字法，三、形碼檢字法，四、字根檢字法，五、字音檢字法」〔6〕在中文文字的輸入方面，也有「字音、字根、形碼、筆順、混合（將兩種輸入法合併使用）」〔同6〕等五種字碼輸入法被研發出來。

這一段時間，可以說是一個文史資料數位化的萌芽階段，中文資訊的處理與應用都已逐漸邁向成熟的階段，而後張仲陶教授、謝清俊教授在中央研究院開始研發古籍資料庫。

## （二）單機版古籍全文資料庫的研發

中央研究院中國古籍全文資料庫發展的過程，謝清俊、林晰於〈中央研究院古籍全文資料庫的發展概要〉文中有詳細說明，茲擇其要略述如下，以明其發展梗概。

該文首先說明了中研院推動史籍自動化計劃的初衷：「爲了中華文化的延續，務必要使古籍能活出現代風貌，不可任其在科技的洪流中式微沒頂，而解決的方法，則是將古籍以電子媒體表達。這就是中央研究院（以下簡稱本院）在1984年7月1日開始推動史籍自動化計劃的初衷。」〔7〕其發展方向爲「本院處理古籍的計劃並不限於只使用全文資料庫技術，有許多資料是用關聯式資料庫處理的。諸如，1985年10月開始試做的【漢代墓葬綜合研究資料庫】，1986年2月的【台灣土著語言資料庫】，1986年4月的【台灣日據時代戶籍資料庫】，1987年1月的【清代竹塹地區土地申告書資料庫】，以及1989年計算中心所做的【說文解字和玉篇資料庫】等等。也有利用影像處理技術所做的古籍資料庫，如傅斯年圖書館發展的【善本書影像資料庫】，目前已完成該館近半數善本書的典藏，並已開放使用。這些資料庫雖非本文報告的重點，然而在語文處理技術上和全文資料庫是相輔相成的。」〔同7〕中研院於1988年推出《史記》、《漢書》、《後漢書》、《三國志》等前四史，1990年完成【二十五史資料庫】，內容上也經過分析、標誌、加工，使其學術研究上的用途更豐富。

除此之外，還有陳郁夫教授也陸續推出單機版的【《十三經》全文檢索資料庫】、【《宋儒學案》全文檢索資料庫】、【《明儒學案》全文檢索資料庫】等，以及筆者所研發的【《全唐詩》全文檢索系統】、【《紅樓夢》多媒體全文檢索系統】。〔8〕

## （三）網路版古籍全文資料庫的研發

網際網路引進國內之後，筆者首先將單機版的【《紅樓夢》多媒體全文檢索系統】改為網路版【《紅樓夢》網路教學研究資料中心】，於1994年上網，並陸續完成《全唐詩》、《全宋詞》、《宋代名家詩》（網址：<http://cls.admin.yzu.edu.tw>）等網路系統；其後中央研究院於1997年將《廿五史》及諸子百家資料庫改為網路版，訂名為【中央研究院漢籍電子文獻】資料庫（網址：<http://www.sinica.edu.tw/ftms-bin/ftmsw3>）如今累計的資料有「整部二十五史、整部阮刻十三經、超過兩千萬字的臺灣史料、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。」〔同7〕，另有台灣大學的【佛學資料庫】（網址：<http://ccbs.ntu.edu.tw/CBS-bin/userfrom/CHINESE>）陸續上網，成為1997年以前網路上最主要的中國典籍資料庫。爾後，陳郁夫教授亦將其單機版資料改為網路版【故宮「寒泉」古典文獻全文檢索資料庫】（網址：<http://210.69.170.100/s25/index.htm>），置於故宮博物院的網站。

#### （四）多功能、多媒體、多元化的文獻資料庫

古籍全文資料庫陸續上網之後，資料的形式從文字資料擴充到多媒體資料，文字資料也從文獻原典資料擴充為研究論著資料，其中有以蒐集研究資料為主的圖書、期刊系統，如國家圖書館【全國圖書書目資訊網】、【中文期刊聯合目錄】（網址：<http://www.ncl.edu.tw/>）、中研院【宋元明清資料庫】（網址：<http://www.ihp.sinica.edu.tw/database/index.htm>）等；以工具書為主的資料庫，如教育部【國語辭典】（網址：<http://www.edu.tw/mandr/clc/dict/>）；有將原典資料與研究資料結合，以提供研究與教學功能的資料庫，如台灣大學的【佛學研究中心】與筆者於元智大學主持的【《紅樓夢》網路教學研究資料中心】、【唐宋文史資料庫】。有提供網路教學、自學功能的，如中央研究院【搜文解字---語文知識網路】、元智大學【倚聲填詞格律自動檢測索引教學系統】、【依韻入詩格律自動檢測索引教學系統】、僑委會的【全球華文網路教育中心】（網址：<http://edu.ocac.gov.tw/index.htm>）等。文物書畫等文獻資料，在這個時期因為網路頻寬不足，傳輸速度緩慢，大多數以光碟形式製作，如故宮博物院的【龍在故宮】、【清明上河圖】、【境攬故宮】等。

#### （五）以三D（dimension）動畫技術呈現立體文獻資料

網路頻寬改善後，中研院於國家數位博物館專案先導計畫項下，研發網路版的【不朽的殿堂－漢代的墓葬與文化】，結合文字、影像、原典與研究資料的綜合文獻，引用三D動畫技術虛擬漢代墓葬文化，使用著隨著滑鼠的移動，進入虛擬的立體墓穴裡，觀賞漢代陵墓的擺設、壁畫、雕飾，為中國文獻數位化開闢另一個更逼近文獻原貌的數位博物館，展現

多樣的文獻資料風貌。台灣大學也以三 D 動畫技術製作【士昏禮】光碟，把《禮記》裡繁複的士昏禮禮俗以三 D 動畫呈現，使用者透過動畫，對於古代的昏禮習俗便可以一目了然，清清楚楚呈現眼前。

由上述可知，文史數位化的發展已有十多年，然相關的研究大抵以發展中文資訊技術及中文字形問題為主，應用方面的研究以計算語言學佔的數量最多，應用於文學上的研究比較少，參與的研究人員也以資訊學門居多，文史學門較少。十多年來，資訊技術的進步，相關產品的配合，已經足夠於「滿足文史數位化的需求」，資訊學界的努力，已經為文獻數位化做好準備的工作，今後文史學界要努力的方向是「提出需求」，並進而拉近電腦「能」與「不能」之間的距離，讓電腦做得比人好的部份，交給電腦來做，人去做電腦還不能做的工作，並且透過語文知識的標示、建構，建立語文知識的人工智慧，使電腦更接近人腦，學習去做人能做的事。〔9〕

#### 四、未來的發展方向

資訊科技的發展，對各行各業造成前所未有的衝擊，對文史學門的研究與教學，也帶來全新的方向。利用電腦作為記錄與傳承典籍的工具，目的在於取其方便性〔同 7〕；利用電腦作為研究工具，取其強大記憶、儲存、分析、檢索能力的優點；利用電腦作為教學工具，取其不受時空限制的特性。以電腦作為文獻儲存的工具，只涉及儲存媒介的改變，但是作為研究與教學的工具，涉及情境的部份，電腦仍然難以取代人的地位。中國典籍數量龐大，全面改建成數位系統，提供傳承、研究、教學使用，非三年五載可以完成，那麼，在古籍數位化的過程中，其優先順序如何？如何兼顧文獻儲存、研究、教學等功能，在電腦「能」與「不能」之間，如何拉近兩者的距離，使電腦處理資訊時具備人工智慧？這是古籍數位化工作進行十多年之後，必須思考的問題。

古籍數位化工作的資訊技術已可滿足需要，中文字形的問題也有比較好的解決方法，十多年來累積的文獻也有數億字，若再加上中國大陸方面的【四庫全書】計畫以及國內即將進行的【善本書】、【古今圖書集成】計畫，古籍數位化的數量正快速而穩定的成長之中。然而，相較之下，在教學及研究上的使用，進度比較緩慢，內容的分析及語文知識庫的建構，仍在起步階段。未來的發展應朝資料庫的有效整合，以擴充使用功能，提供更多元化的使用需求，以及內容的標注與分析，建立語文知識網路、引用電腦作為新的研究工具以開拓新的研究方法而努力。

##### （一）資料庫的建立方向與功能的擴充

以電腦作為文獻儲存的工具，只涉及儲存媒介的改變，但是作為研究與教學的工具，涉及情境的部份，電腦仍然難以取代人的地位。中國典



籍數量龐大，全面改建成數位系統，提供傳承、研究、教學使用，非三年五載可以完成，那麼，在古籍數位化的過程中，其優先順序如何？如何兼顧文獻儲存、研究、教學等功能，在電腦「能」與「不能」之間，如何拉近兩者的距離，使電腦處理資訊時具備人工智慧？如何建立中國語文的類神經網路系統？這是古籍數位化工作進行十多年之後，必須思考的問題。

古籍數位資料庫的建置，電腦技術已經足以滿足需要，各種不同文體的系統架構模式也已成熟，資料庫的數量，也有數億字，因此，循同樣模式建置新的資料庫，已經是輕而易舉的工作。未來的發展，若仍以此為唯一目標，只是在數量的寬廣度方面增加，並不能在質的深度上提升。

如何在質的深度上提升，依筆者近幾年來觀察所得，提出幾點淺見：

### 1. 廣泛建立工具性書籍資料：

在研究的過程中，研究者對於研究範圍的原典資料需要仔細研讀，反覆咀嚼、推敲，以電腦檢索所得之資料，只是片段的、零星的，所以需要熟悉原典資料，才不至於見樹忘林。工具書一般作為查詢使用，屬於參考性質，因此工具書改製成電腦系統，使用效率高於其他原典資料。電腦超強的搜尋、檢索資料能力，其效率千萬倍於人腦，也不至於遺漏。工具書的開發對使用者的用處最大，例如字典、辭典等，以人工翻檢的紙本字辭典，只能從字首查詢，以電腦作為查詢工具，可以從任何一個關鍵字查詢；又如《宋人傳記資料索引》之類的工具書，如能改以電腦檢索，使用效率當可大大提高。

### 2. 工具書資料庫與原典資料庫結合查詢：

研究唐宋詩詞的學者，大概都有過這樣的體驗，唐宋詩人喜以詩詞交往，詩人於詩題、詞題中提及某人時，常以官銜、字號、別名、排行等為題，後之研究者要查明該人物之確實身份，往往大費周章仍不一定可得。以宋代為例，雖有《宋人傳記資料索引》可供參考，但該書只能從資料之字首查詢，使用不便。

唐詩亦如是，江蘇吳汝煜、胡可先二位學者有鑑於此，集數年時間心力，完成《全唐詩人名考》，該書「共搜輯別人考訂成果約 3440 餘人次，自己考出的人名約 3860 餘人次。合起來總數約有 7300 餘人次。」（《全唐詩人名考》，前言）本書作者吳汝煜與胡可先二位先生，在其《唐五代人交往詩索引》（上海古籍出版社）的基礎上，進一步做《全唐詩人名考》。主要考證：「《全唐詩》題目、序、注中以官職、封爵、諡輩號、科第、行、地望、職業及字號等相呼稱的人物的姓名，並扼要介紹其生平資料，某些唐詩作者姓名有誤者，亦加以辨正，旨在為唐詩研究者和欣賞者提供參考。」〔10〕（《全唐詩人名考》，凡例一）

《全唐詩人名考》以人力搜輯資料，以《全唐詩》題目、序、注為基本資料，參校其他考證資料，逐一以人力核對，耗費大量的人力與時間。吳先生也因積勞成疾，英年早逝，殊為可惜。如果能以電腦為工具，廣泛建立相關資料，藉助電腦強大的蒐集、整理能力，當可達到更好的研究效率。唐德剛先生在使用中央研究院【廿五史系統】後，憶及胡適先生當年埋首於考據工作，上窮碧落下黃泉的尋找資料，耗費很多時間，唐先生因而有「人才浪費不起」的感嘆。將工具書與原典資料結合查詢，可以解決類似的問題。

### 3. 縱向與橫向資料的結合檢索

不同年代的縱向資料，設計可以跨越時間檢索縱向檢索的功能，在研究上提供了清晰的縱向演進軌跡，有其重要的意義，如中央研究院【廿五史系統】，便可以選擇單獨檢索斷代史，也可以選擇以全部《廿五史》為檢索範圍。

橫向跨領域資料的檢索，目前較少，是未來發展的重要方向。以文史而言，文史自來相通，文學作品引用大量史料，或擷取史料化為典故以豐富文學作品內涵，唐宋詩詞的典故便有很多來自史書以及其他神話、小說典籍，後人讀詩讀詞，若不瞭解典故出處、意義，便無法完整掌握文意。筆者於進行《晁補之及其文學研究》之研究過程中，因晁補之詩沒有任何註解本，詩作中引用典故的句子，在現有工具書查詢不到典出何處時，嘗試以該句字詞為關鍵字，到中研院【瀚籍資料庫】查詢，往往有意想不到的收穫，解決了很多典故出處的難題，橫向跨越不同資料庫的交叉查詢功能，有其重要意義。

前項所言，詩詞題中人名資料的查詢，跨領域資料庫的交叉查詢，亦有助於文學資料的澄清。文學資料也常常可以校正史料的錯誤，這些工作透過電腦檢索比對，可以省卻許多人力，又可得到更好的成效，因此橫向、縱向跨領域資料的綜合交叉檢索，可以拓展出新的研究領域，也可以進行一些單憑人力很難完成的研究工作。

筆者有鑑於此，自 1998 年起，進行唐宋代文史綜合資料庫的實驗計畫，資料庫包含《新唐書》、《宋史》、《全唐詩》、《全宋詞》、《北宋名家詩》、《宋人傳記資料索引》、唐宋地名等資料，進行跨領域結合檢索的實驗。未來再增加《全唐文》、《全宋文》、《舊唐書》、詩話、詞話等資料，以作為唐宋文史研究的資料庫。

### 4. 原典資料與後人研究資料的結合

原典資料固然是研究過程中最重要的依據，後人研究資料也不可或缺。建立一個可以結合二者查詢的資料庫，能提供給使用者更大的便利性。

後人研究資料包含專書著作資料、期刊論文資料、會議論文資料、以及其他相關的周邊研究資料。並且將這二種資料建立關鍵字詞、建立參見檔，使用者在查詢某一範圍的原典資料時，如果已經有後人研究資料，系統能主動出示，如此便可以清楚掌握研究的動態，避免重複。

## 5. 系統架構朝向開放式的系統架構

封閉式的系統架構會影響資源的交換、共享，也會影響系統的擴充、發展；對系統與程式設計者而言，增加開發的成本，對使用者而言，增加學習的困擾。因此，採用國際標準的通訊協定、共用平台、開放式的系統架構，取得資源比較容易，這將是必要的趨勢。系統架構雖然會因為資料內容的不同而有些微的差異，但只要作小幅度的修正即可。筆者所主持的【網路展書讀----中華典籍網路資料中心】（網址：<http://cls.admin.yzu.edu.tw>）便是採用開放式的系統架構。以【全唐詩系統】為例，該系統完成後，如果要再建立【宋詩系統】，因為唐宋詩的資料結構一樣，同樣可以從「作者」、「詩題」、「詩句」三個檢索點檢索，所以只要抽掉《全唐詩》文字資料，換上《宋詩》資料，便成為另一套【宋詩系統】，系統與程式不必重新設計，對管理者而言，節省開發與維護的成本，對使用者而言，只要學會了使用【全唐詩系統】，便同時能夠使用【宋詩系統】，不必重新學習。《宋詞》同樣屬於韻文，但是與詩比起來，句子的長短有更多的變化，形式上比詩多了詞牌、宮調，在檢索點的設計上，多出「詞牌」與「宮調」二個檢索點。小說、散文也可以循同樣的模式，所以古籍數位化的發展，必須注意到架構的開放、資源的共享、系統的擴充等因素。

## 6. 從單向的查詢擴展為雙向互動的系統

早期傳統的自動化系統，以「管理的自動化」為主要的目標，系統的設計以「單向被動」的等待使用者前來查詢。「雙向互動」的觀念，以「服務的自動化」為導向，其內涵精神包括「主動性的資訊服務」，意即從「單向被動」的等待使用者提出需求，進步到主動的提供服務。及「互動性的資訊交換」，使用者可以回饋資料到資料庫，讓資料的蒐集從管理者單向蒐集，擴大到所有使用者多向回饋。這種作法可以保持資料的完整性與新穎性。〔11〕

## 7. 自助式的功能設計

任何一個開放性的資料庫，都是同時提供給多人使用，每個人的需求不同，再完整的系統設計，都難以滿足所有使用者的需要，因此，引用自助式（Do it by yourself）的觀念，製作一個簡單便捷、易學易用、人性化、個別化的個人工作平台，提供可以整理個人資料與網路資源的工具，讓使

使用者在使用資料庫的資料時，可以將資料庫的資料、個人外加資料、其他網路資源等，以複製黏貼（Copy paste）或連結（Linker）的方式建立個人的資料庫或個人專屬網頁，提供個人蒐集、整理研究資料，或編纂教材，或與他人交換資料之用，可以滿足不同使用者的個別化需求。〔12〕

## （二）語文知識的分析、標示與建構方向

中國文獻數位化的研究工作，在字形的表述與語言學方面的研究取得最多的成果，應用於文學方面的研究較少。中文字形的演變有一定的脈絡可循，語言學也可以歸納出一定的規則，而文學則複雜得多。文學的形式，經過許多的演變，文學的內容，因為作者的不同、時代的不同、環境的不同、甚至於君王好惡態度的不同，而呈現多樣的風貌。文學的內容又包含了人的思維與感情，藝術表現手法等。以電腦作為文學研究的工具，全文檢索的功能已不能滿足文史學門的需求。

文史數位化的發展，經過資訊學門十多年的努力，在技術上已經完全可以滿足文史學門的需求，全文檢索的功能，經過十多年的發展，技術已經完全成熟，以之應用於資料的全文檢索，其正確率與效率都令人滿意。然而，中國學術研究如果要引用資訊科技作為研究工具，除了全文輸入之後提供全文檢索功能之外，還需要做很多標示的工作，這些標示可以建構一個電腦的語文知識網絡，使電腦具備人工智慧，提高資料檢索的完整性，可以大大提升中國學術研究的應用範圍。然而在電腦不具備思維能力、感發能力之前，電腦對於文史研究的幫助，只限於資料的儲存、搜尋、分析、整理，而且其正確率與速度都比人工快上千萬倍，但是對於內容的研析與判斷，距離人的判斷仍有相當大的距離。

各種不同的文學形式中，又以詩學最為精緻凝練，所以，引用電腦科技作為文學的研究工具，有一定的困難度，作為詩學的研究工具，困難度更高。因此，電腦要作為文史研究的輔助工具，如果只使用電腦強大的記憶、搜尋、分析、檢索能力，當然已經足足有餘，如果要進一步涉及文史內容等語文知識的範疇，需要朝人工智慧發展，還需要一段時間的努力，需要更多文史學者的參與。

以電腦作為漢學研究的輔助工具，目前最大的侷限是缺字問題以及字詞的形音義表述問題，前已述及。對電腦而言，每一個不同的字形都是一個獨立的符號，不代表任何意義。但是對人而言，每一個字形含有字音、字義。每一個字形映入腦中，都能立即將形音義三者串連，產生一個適當的意象，那怕是一字多形、多音、多義的字，都能有不同的聯想反應，甚至於由字進一步組成詞、組成句子，組成一篇文章，都能給予讀者不同的感發，感發的情境還會因人、因時等各種因素的不同而有差異，但是對電腦而言，除非給予每一個字必要的定義，否則都只是符號。每一個字進入

電腦變成單一的符號，作為保存文獻的工具，足足有餘，但是，作為文史研究的工具，仍有落差。

中國文字屬於方塊字，每一個字具有形音義三個要素，而且大多數的字一字多形、一字多音、一字多義，組成詞之後的變化更大。對於電腦而言，電腦只認得字形，不同的字形對電腦而言都屬於不同的符號，也就是電腦會將不同字形的異體字定義為不同的字，除非以人工標示二者為同一個字。電腦無法辨認異體字之間的關係，更遑論辨識通同用字之間的關係，如果再加上字音、字義、詞彙的變化，就成為複雜的語文知識網。而這些語文知識網路的建立，需要靠人工去標示，也可以藉助電腦為工具，使標示的效率提高。

資料的加工標示，解決了異體字的對應關係之後，再標示字音與字義。文字、聲韻、訓詁之學為基礎之學。在中國文字具有一字多形、一字多音、一字多義的特色下，電腦只能辨識字形，無法辨識同義字詞時，正確的資料（字形完全符合）往往不等於完整的資料（含同義字詞）。

提升中國學術研究的應用範圍與應用效率，首先要將文字之形、音、義關係標示清楚，其次標示詞語、語法、詞性，再其次標示專有名詞，再其次標示典故，其標示方法如下：

## 1. 文字形音義的標示

### (1) 字形標示：

中國文字經過千百年的演變，字形有甲骨文、金文、大篆、小篆、隸書、楷書等差別，字形上有變化，加上後世使用之後，產生所謂的通俗體字、古今字、通用字、簡繁體字等異體字並存，而有一字多形的現象。

對於人而言，大多數的異體字，人之肉眼所及，立即能辨認是相同的字，但對於電腦而言，每一個字形都是符號，不同的字形，有不同的符號，不同的內碼。在電腦的辨識上，「不同的符號」便代表「不同的字」。而在詩詞裡，常使用通用字或通用詞，例如「遊」與「游」、「由」與「猶」、「強」與「彊」、「穠」與「濃」、「間」與「閒」、「仔細」與「子細」、「蝴蝶」與「胡蝶」在詩詞裡常通用，因此我們必須要先建立一個「通同異體字詞」資料庫，教會電腦辨識異體字與通同用字之間的關係，甚至於加上詞彙資料庫輔佐，更正確辨識異體字之間用在何種情況下相等（相通），用在何種情況下不相等（相通）。這些通用字通常以單字詞的姿態出現，所以更需要標示，如果以二字以上的詞出現，還可以藉著詞語的標示區別。

### (2) 字音標示：

常用的中國字讀音約有一千四百種，因此有極多的同音異義字，而大多數的中國字又具有一字多音的特性、音隨義轉的特性。因此讀音的標示，會影響到字義與詞義的標示，也就是說音與義存在密切的關係，因此正確的標示讀音，有助於提升字義與詞義標示的正確性。

### (3) 字義與詞義標示：

中國文字大多一字多義，因此同義字的標示，可以提高資料檢索的完整性。詩詞為求修辭之美，同義字詞的變化較之其他文體更為豐富多樣。多義字往往有其特定的用法，亦即某一種字義的字，必定與另一個字組成一個特定的詞，這個詞使用這個字義。因此字音、字義、詞義是相關的。

根據徐超著《中國傳統語言文字學》〈第五章語源學〉曰：「所謂『語源』，實際上就是『詞源』。但不是『詞語在文獻裡最早出處』的詞源，而是指詞語的音義來源。既然講『來源』，就不是一個詞語的問題，至少他要討論到此詞與彼詞的關係，及涉及到兩個或兩個以上的詞。……因此，語源學所要研究的，實際上主要還是同源詞（又叫同族詞）的問題，及研究一個或多個詞語的歷史的淵源關係，屬歷史語言學的範疇。他的任務是上推語源，下求流變，藉此聯繫同源詞，進而探求詞語的孳生規律，建立科學的語言學理論等。傳統語言文字學裡的『聲訓』和『推源』等訓詁方式大致可以看成是語源探究的性質。」

〔13〕徐超還進一步指出：「傳統語言文字學裡的語源學研究，是以『聲訓』的名目和形式出現的。什麼是『聲訓』呢？聲訓是指就字（詞）音來探求語源和貫通詞義的訓詁方法。……這就從研究內容（語源、同源詞）和研究手段（借助於語音分析）上跟語源學差不多取得了完全的一致。他們都是以研究詞彙的同源系統為主要目的的。」〔同 13，頁 282-283〕由徐超這段文字可知，語源與流變之探求以語音分析為主要手段（方法），也可以在探求了語源與流變的軌跡中，分析語義的流變與關係。因此，字音與字義、詞義的標示息息相關。把語源與流變標示出來，建立語源流變資料庫，有助於更正確的標示字音與字義、詞義資料庫，這個資料庫除了提供語義學、詞彙學的研究之外，還可以分析出同義字詞，提升檢索的完整性。

### (4) 同義字詞的標示

同義字包含前面所講的通同異體字，是普遍通用的字，而同義字詞是指為了修辭上的變化而使用不同的字詞，但其字詞義相通者，如詩人寫酒，以「杜康」代替。同義字詞還包含引用同一典故，但是使用的字詞不同，因典故來源一樣，其引用的意義一樣，而成為同義字詞，

例如：以秦代東陵侯種瓜事入典的東陵瓜典故，出現在詩詞裡的用法就有很多種，東陵侯指秦代召平，《史記》卷五十三〈蕭相國世家〉第二十三載曰：「召平者，故秦東陵侯。秦破，爲布衣，貧，種瓜於長安城東，瓜美，故世俗謂之『東陵瓜』，從召平以爲名也。」（頁2017）召平又作邵平。這個典故，在唐宋詩詞裡變化成「邵平瓜」、「邵平」、「邵平園」、「邵平田」、「東陵侯」、「東門瓜」、「東陵瓜」、「東陵」、「東園瓜」、「東陵圃」等不同的詞，但其意義都用作爲歌詠遺民或隱士。這一類的同義字詞如能經過標示，一者可以明修辭之變化，再者檢索相關資料時能夠更完整。

## 2. 詞語、詞性、語法的標示

### （1）詞語標示：

詞語標示即建立詞彙資料庫。詩詞是比較特殊的語言，詩詞詞彙資料庫的建立，在古籍數位化的龐大工作中可以提供極大的助益，例如提高掃描軟體的辨識率、提高自然輸入法（聯想輸入法）的選字正確率、可以提供文章的除錯功能等等。筆者現在正以唐宋詩爲實驗對象，建立詞語自動標示的功能，標示的方法另有專文論述。

### （2）詞性的標示：

在多義字詞的中國文字裡，字義、詞義往往與詞性相關，詞性的不同往往又牽涉到讀音的不同，例如很多字當作動詞使用時改念爲第四聲便是。詞性的標示還可以提供語法的研究與律詩對仗的研究。

### （3）語法的標示：

文言文及詩詞常用倒裝句，語法的標示有助於了解詩文含意。

## 3. 專有名詞的標示

人名、地名、帝王年號、官職稱謂、山川草木、鳥獸蟲魚、風雲雪雨等專有名詞的標示，除了有助於提升詞語與詞性標注的正確性之外，還可以提供研究的需要。人名的標示可以提供文人交遊考的研究、地名的標示可以窺見城市文化的榮衰、帝王年號及官職稱謂的標示，可以查考作者及作品的時代背景、山川草木及鳥獸蟲魚的標示可以探討文人的文化活動，人與外物的互動、風雲雪雨的標示可以提供人與自然的互動研究，以及修辭的技巧等等。除此之外，這些專有名詞先標示出來以後，對於以電腦自動標示詞語的工作有極大的幫助。

#### 4.典故的標示

典故的應用在文學中應用比較多，韻文尤其普遍以典故修辭，豐富詩意，典故的解釋又與單純的字詞義不同，因此典故的標示也是不可或缺的。除此之外，典故的標示，對於以電腦自動標示詞語的工作有極大的幫助。

#### 五、結論---對未來文史研究的意義

前已言及，古籍數位化工作發展迄今，相關的研究以發展技術及文字問題為主；應用方面的研究，成果最豐碩的是計算語言學方面的研究，中央研究院語言研究所及北京大學計算語言學研究所，在這方面取得很好的成績；應用於文學方面的研究，最早見於《紅樓夢》的研究，不過也是偏於《紅樓夢》的語言風格研究，其次有應用於詞學方面的研究，以詞律方面的問題為主，筆者於〈在網際網路建立漢學研究環境的重要性及可行性---就中國文學而論〉一文有詳細的論述〔14〕。

同文也論及多位學者對於以電腦作為文史研究輔助工具的看法。然而事隔二、三年，電腦的技術再往前躍進一大步，語文知識網路的觀念建立了，相關的研究也有「國科會數位博物館先導計畫---搜文解字語文知識網路計畫」〔15〕正在進行，類神經網路的觀念與實際應用的研究也都如火如荼的展開，與當年不可同日而語，在這樣的環境下，我們再來檢視以電腦科技作為文史研究的輔助工具，其意義何在？

如同文結語所言：「電腦是不是能成為帶領中文研究工作突破傳統窠臼的萬靈丹？在電腦還不具備思考功能之前，這個答案顯然是否定的。電腦的某些特性優於人腦，可以協助從事以人力無法做到的統計工作，但是無法完全取代人在研究中扮演的角色。」〔同 14〕筆者於〈中國古典詩詞教學與習作的新嘗試---網路作詩填詞系統---兼及其可行性與侷限性〉一文之結語述及電腦的「能」與「不能」時，以作詩填詞這種純粹文學創作的電腦輔助工具為例，提出以下看法：「詩詞創作是純粹的心靈思維活動，電腦誠然難以取代理人類心靈的思維活動，以及情意的感發、美感的審查、經驗的共鳴。但是對於規則性的資料，電腦的處理能力千萬倍於人，藉助電腦工具，可以為人類節省很多時間。這方面的功能，對於中國古籍研究的幫助不容小覷。但文史研究的自動化，若以此為終極目標，卻也辜負了資訊學界在電腦科技方面所做的努力。文史學界從文史專業的角度，引導文史數位化的方向，讓電腦科技為文史研究作更好的服務，將平面的文史資料加上必要的標注，藉著知識結構的改變，研究工具的改變，產生新的研究方法，拓展新的研究領域，必能產生新的研究成果。處理單一規則的訊息，檢索平面的資料，對電腦而言，只是雕蟲小技。如何在電腦『能』作的與『不能』作的中間，尋找一條突破的管道，使電腦更接近人工智慧，讓電腦科技與人文研究作最好的



結合，對於資訊界與人文界，這條從『不能』到『能』的過程，對雙方而言，無疑的，都是一個很大的挑戰。」〔16〕

純粹心靈活動、意念感發的創作，都有可以縮短從「不能」到「能」的空間，文史的研究當然也有一些方法可以嘗試。

在網際網路上建構一個可以「多元整合、交互參照」的網路綜合資料中心，每一種資料可以單獨成爲一個主題資料庫，不同資料庫可以透過適當的系統設計，提供跨資料庫多元組合的方式交互檢索，成爲一個綜合型資料中心，如此當可以提供給使用者更多元化的研究與教學使用需求。多種資料庫結合的系統設計，對學術研究的效益，數倍於單一資料庫。

在使用功能上，提供具備人工智慧的語文知識結構功能，使電腦的分析判斷能更接近人的品質。

在系統開發與管理上，以使用者的需求作爲系統設計的導向。研究論著資料需隨著新增資料的增加不斷更新，原著典籍資料也可能因漏收或因文物出土而增加。傳統資料庫的作法往往無法提供隨時更新資料的空間，功能的設計也只能提供使用者從資料庫中單向取得所需資料，系統製作人與系統管理員都很難在「資料蒐集的完整性」與「資料更新的時效性」兩方面做到立即而面面俱全的地步。從人性化的角度來看，並不能完全符合不同使用者的需求。因此，這兩項工作假若由系統管理者單方面承擔，轉而由所有使用者全面分攤，可以得到改善的機會。由系統提供簡單、便捷、人性化的介面與工具，引用DIY(Do It by Yourself)的觀念，從使用者需求的角度，以「使用方便」爲導向，兼顧「管理方便」的原則，設計一個多向、靈活、人性化的網路資料系統，並建立一個可供個人蒐集、儲存、整理、編纂資料的個人工作平台，使用者可以從資料庫取得符合個人研究需要的資料，或者資料庫漏收、未收的資料。系統亦允許使用者自行建立具有個人風格的工作平台，在其個人工作平台上進行研究工作，讓系統功能對個別使用者的侷限降到最低。除此之外，提供一個資源共享的空間，使用者有新增資料時可以回饋到資料庫，資料庫的更新可以從管理者擴充到每一個使用者，如此一來，資料的更新便由管理者單向增補擴充到由使用者全向增補，資料蒐集的完整性與更新的時效性必能提高很多。

在使用功能上，提供具備人工智慧的語文知識結構功能，進而建立中國語文的類神經網路系統，是未來發展的方向。全文檢索的功能，經過十多年的發展，技術已經完全成熟，以之應用於資料的全文檢索，其正確率與效率都令人滿意。然而，文史研究如果要引用資訊科技作爲研究工具，除了全文輸入之後提供全文檢索功能之外，還需要做很多標示的工作（已如前言），這些標示可以建構一個電腦語文知識網絡，使電腦具備人工智慧，使資料的檢索從正確性提升到完整性，可以大大提升文史研究的使用效率與應用範圍。全文檢索技術的發展，在資料檢索的正確性與效率性方面已經毫無問題，但是在中國文字具有一字多形、一字多音、一字多義的特色下，電腦只能辨識字形，無法辨識同義字詞時，正確的資料（字形完全符合）往往不等於完整的資料（含通同義字詞）。透過各種語文

知識的標示，使電腦具備人工智慧，和自動學習的能力，提高資料檢索的完整性與正確性，更能為文史研究注入一股新的生命力。

跨資料庫的聯合檢索，可以使封閉式的研究邁向跨領域的開放研究；自助式的使用功能設計，可以使系統提供更人性化的功能；文字形音義的標示，更是促使電腦具備人工智慧的必然條件，這些標示工作的累積、相關知識的連結，有朝一日，一定能形成一個中國學術研究的類神經網路系統。

在電腦不具備思考能力之前，在電腦無法涉及文學研究中的感情活動與美學欣賞之前，電腦誠然無法完全取代人腦，人文學界的研究如此，在其他領域的研究上，電腦也只是一個「工具」，研究工作的主導者是「人」。

值得期待的是在數位化網路上建構漢學的研究環境，電腦學界的努力，使技術上已能符合文史學界的需求，文史學界如何與電腦學界結合，使電腦技術能充分為人文研究服務，以消彌科技與人文的疏離隔閡，有賴於二個領域的合作。

十年前文史界與電腦界已有這樣的共識：張仲陶先生強調「不要問電腦能做什麼？而是問你要電腦為你做什麼？」；文史界的學者也強調「文史資料自動化的過程中，不能失去人的主導地位，以及學門的主導地位。」、「如何善用電腦的長處，並避免可能帶來的隱憂及弊端。」是文史界的期許。「由文史界負責提出『需要什麼』？由資訊界負責『如何滿足需要？』」是文史與資訊兩個學門的交集與共同努力的方向。

新的科技，新的方法，可以開拓新的研究領域，研究方向，這是毋庸置疑的。然而如何使用工具？並且進一步建構新的研究環境，產生新的研究主題，創造新的研究成果，在在都值得文史研究工作者深思，也需要文史學門參與，開創一個有別於傳統的研究方法、研究領域。使用新的工具時如何調整研究方法，讓工具能為從事研究之「人」所用，而不是「人」被工具所限制，甚至於過度依賴、膨脹了新工具的功能，是過程中必須留意的。

走過十多個年頭，典籍資料的數位化，累積的數量已經相當可觀，全文檢索的技術也完全成熟，透過知識結構的建立，使電腦在提供人文研究時，能夠更接近人力判斷的品質，對研究的幫助將有突破性的進展，讓電腦做得比人好的部份交給電腦做，讓人做得比電腦好的部份，想辦法讓電腦也能做，在電腦還不能做之前，人去做電腦還無法做的工作，這將會是文史研究引用資訊科技作為輔助工具時，最有意義的一個方向。

#### 附註：

- 一、「數位博物館專案先導計畫」是國科會 1998 年委託執行的計畫，第一年度的執行期間為 1998 年 12 月 1 日至 1999 年 10 月 31 日，共有【不朽的殿堂－漢代的墓葬與文化】、【搜文解字－語文知識網路】、【臺灣原住民－平埔族群】、【淡水河溯源】、【蝴蝶生態面面觀】、【臺灣的本土植物和魚類】、【火器與明清戰爭】、【資源組織與檢索之規範】、【人文與自然資源地圖】、

【數位典藏系統技術研發】、【系統評估—以「淡水河溯源」為例】等十一個計畫。

二、「數位博物館計畫」是國科會繼「數位博物館專案先導計畫」的第二年計畫，執行期間為1999年12月1日至2000年10月31日，共有【故宮文物之美系列】、【玄奘西域行】、【淡水河溯源（Ⅱ）】、【台灣民間藝術家之建置—以楊英風為例】、【台灣建築史】、【生命科學—人體奧秘展覽館】、【台灣本土魚類（Ⅱ）—尋回台灣本土的淡水魚類】、【語文知識網路（Ⅱ）—生活語文、唐宋流行歌】、【平埔文化網路之維護與管理】、【中醫藥、針灸數位博物館】、【蘭嶼生物/生物多樣性數位博物館】等十一個計畫。

三、「國家數位典藏計畫」依據1999年7月行政院『電子、通訊、資訊策略會議』通過，會後行政院將此案交由國科會協調執行，預計2000年1月1日正式開工。內定參與此計畫機構有：故宮博物院、國家圖書館、省立歷史博物館、臺灣省文獻委員會、台中自然科學博物館、臺灣大學、中央研究院七個單位。

四、本段引文均引自拙著，〈探一探文史資料自動化的路〉，《國文天地》第3卷第3期（1987年8月），頁33-43。

五、論文詳目如下：

1. 張永銘著，張仲陶教授指導，《中文書籍自動拼版系統之設計》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十一年。
2. 鄭一雄著，張仲陶教授指導，《中文字形輸出系統的設計》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十一年。
3. 曾士熊著，張仲陶教授、謝清俊教授共同指導，《中國文字特性資料庫的設計》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十一年。
4. 王義科著，張仲陶教授、謝清俊教授共同指導，《中文文書編輯系統之研製》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十二年。
5. 高天助著，張仲陶教授、謝清俊教授共同指導，《國字資料庫之維護》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十二年。
6. 潘敏政著，張仲陶教授、謝清俊教授共同指導，《在時間域作中文語音合成的研究》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十三年。
7. 郭明仁撰著，張仲陶教授、謝清俊教授共同指導，《辦公室用的中文印製系統之設計》，國立台灣工業技術學院工程技術研究所電子工程技術組碩士學位論文，民國七十三年。

六、引自曾士熊著，《中國文字特性資料庫的設計·第一章概論》，頁3-6。

- 七、謝清俊、林晰著，〈中央研究院古籍全文資料庫的發展概要〉，收錄於中央研究院資訊科學研究所文獻處理研究室謝清俊論文區，網址：<http://www.sinica.edu.tw/~cdp/>，發表日期：1997年3月。
- 八、筆者所研發的單機版《全唐詩》全文檢索系統完成於1992年，單機版《紅樓夢》多媒體系統完成於1992年，並於1993年10月於北京市舉辦的海峽兩岸中國古籍整理研究現代化技術研討會發表〈《紅樓夢》多媒體系統〉論文一篇。
- 九、參見拙著，〈中國古典詩詞教學與習作的新嘗試---網路作詩填詞系統兼及其可行性與侷限性〉，《教學科技與媒體》，1997年12月15日，頁2-11。
- 十、見吳汝煜、胡可先著，《全唐詩人名考》，江蘇教育出版社，1990年8月出版。
- 十一、參見拙著，〈以「互動觀念」建立「紅樓夢網路資料中心」對紅學發展之影響〉，《紅樓夢學刊》，1997年增刊，總第75期，頁532-541。
- 十二、參見拙著，〈DIY 唐宋詞多媒體網路系統〉，《中國古籍整理研究出版現代化國際會議論文集》，1995年7月22-24日，頁287-302。
- 十三、徐超著，山東大學文史書系---《中國傳統語言文字學》，1996年6月，山東大學出版社出版，頁281-282。
- 十四、參見拙著，〈在網際網路建立漢學研究環境的重要性及可行性---就中國文學而論〉，《漢學研究通訊》，1997年2月，第16卷第1期。
- 十五、黃居仁、鍾柏生、羅鳳珠，1998年12月1日至1999年9月30日，《數位博物館專案先導計畫--「搜文解字 I」--語文知識網路》，國科會（NSC88-2745-P-001-011），網址：<http://www.dmpo.sinica.edu.tw:8000/words/main.html>。
- 十六、參見拙著，〈中國古典詩詞教學與習作的新嘗試-網路作詩填詞系統---兼及其可行性與侷限性〉，《教學媒體與科技》，1997年12月15日，頁2-11。