

導論——什麼是數位人文

項潔*、涂豐恩**

摘要

數位人文的研究，近年來在國內外有著長足的發展。2009 年底，國立臺灣大學數位典藏研究發展中心舉辦的第一屆「數位典藏與數位人文」國際研討會，是這股趨勢在臺灣學術界的具體呈現。本文要圍繞「什麼是數位人文」的問題，提出觀察和討論。我們將從近年來的趨勢談起，簡單描述數位人文如何在全球的學界中萌芽，並迅速地發展，包括機構、期刊與會議的成立。接著討論數位人文發展的背景，即過去十餘年內數位化在各地所累積的豐碩成果，這樣的成果為人文研究帶來了挑戰，但同時也是另一波發展的契機。如何迎接這樣的挑戰，並將挑戰轉化為機會，有賴新的工具與新的研究方法。最後，將視野放大，數位科技改變的不只是人文學者研究的方法，更改變了整體的學術環境，改變我們傳遞、溝通和交換知識的方式，包括教學與學習，研究的呈現，還有知識社群的型態。

* 國立臺灣大學資訊工程學系特聘教授。

** 國立臺灣大學數位典藏研究發展中心碩士後研究員。

Introduction: What is Digital Humanities

Jieh Hsiang ^{*}, Feng-en Tu ^{**}

Abstract

The studies of digital humanities have been flourishing in recent years. We witnessed a growing number of projects, research papers as well as newly established journals and institutions. The First International Conference of Digital Archives and Digital Humanities, hosted by National Taiwan University in 2009, also reflected this significant leap toward the convergence of information technologies and a wide range of humanities in Taiwan. This article, as an introduction to the book, aims to depict the research trends around world and to discuss the possibilities of digital humanities. We provide a variety of examples to illuminate three different but related themes: (1) the achievements of digitization in the past decades; (2) new research tools and methodologies in digital humanities; (3) the way of presenting and sharing humanities studies in a connected world.

* Distinguished Professor, Department of Computer Science and Information Engineering, National Taiwan University.

** Research Associate, Research Center for Digital Humanities, National Taiwan University.

2009年12月1日至2日，國立臺灣大學數位典藏研究發展中心舉辦了第一屆「數位典藏與數位人文」國際研討會。會議內除了四場主題演講，一場圓桌論壇外，並有十八篇論文發表，及十餘篇海報論文。本書所輯錄的文章，都曾在該次會議宣讀過。這是臺灣第一個以「數位人文」為名的研討會，因此有著特別的意義。

在臺灣，「數位人文」(digital humanities)還算是個陌生的詞彙。但就在近年內，我們目睹數位人文的研究在國際間快速開展，相關的社群成立，甚至不乏專門的會議、期刊與教科書。在眾多學者共同努力下，「數位人文」作為一個新興的研究領域，愈來愈顯得豐富和厚實。

但究竟什麼是數位人文？如果要有個簡單的定義，那麼，它指的是那些唯有借助數位科技方能進行的人文研究。反過來講，數位人文的研究，即是企圖尋找在前數位時代中難以觀察的現象、無法想像的議題與無法進行的研究。這樣的定義或仍不免顯得寬泛，但正因如此，它能夠容納比較多的可能性——對於初生的領域而言，與其畫地自限，這樣的可能性應該更為重要。另一方面，這個定義提醒我們，數位人文研究應該專注於真正的突破，並帶來概念上跳躍性的發展 (conceptual leap)，而不只是小規模的改善。這當然不是個簡單的目標。但無論如何，我們希望這個簡要的定義，不是限縮想像力的圈套，而能成為研究者探索時的指引與參考。

從這個定義出發，本文要圍繞「什麼是數位人文」這個議題，提出觀察和討論，作為本書的導論。我們將從近年來的趨勢談起，簡單描述數位人文如何在全球的學界中萌芽，並迅速地發展。接著討論數位人文發展的背景，也就是在過去約十年內數位化在各地所累積的豐碩成果。這樣的成果為人文研究帶來了挑戰，但同時也是另一波發展的契機。如何迎接這樣的挑戰，並將挑戰轉化為機會，就有賴新的工具與新的研究方法——這是本文第三節的主題。

但數位人文也可以有更廣泛的指涉。如果將視野放大，數位科技改變的不只是研究方法，甚至影響著人文學界的生態。這樣的發展不難理解，當數位科技改變了人的生活，以生活為研究對象的人文學科，自然也會展現出新的面貌。以下就依次討論數位人文所牽涉的各種議題。

一、全球性的趨勢

2010年的11月17日，美國紐約時報刊出一篇報導，名為〈打開人文寶藏的數位鑰匙〉(Digital Keys for Unlocking the Humanities' Riches)，文章的起頭說，二十世紀的人文學歷經一波又一波「主義」的洗禮——形式主義、佛洛伊德主義、結構主義、後殖民主義，凡此種種。作者問道，下一個的關鍵概念 (big idea) 會是什麼？他的答案是：資料 (data)。因為我們生活在這樣的年代，研究者所能接觸的資料

量前所未有的，資訊技術則提供了處理大量資料的工具。兩者匯合，催生了「數位人文」的誕生。

這只是《紐約時報》一系列「人文 2.0」（Humanities 2.0）報導中的一個片段。從該年三月分開始，《紐約時報》已經就數位人文的主題，陸續發表多篇報導。報導的主題，涵蓋文學、戲劇、歷史等各種人文學的不同分支。這說明數位人文所帶來的衝擊，跨越了傳統學科的界線。¹

當然，數位人文的發展仍或存在爭議。在前述 11 月 17 日的報導中，美國普林斯頓大學的歷史學者 Anthony Grafton 就在受訪時提到「我們很容易忘記數位科技只是手段而非目的。」即便如此，這位 2011 年新任的美國歷史學會會長，於另一場簡短的訪談中，仍舊強調，在他的一年任期之中，最重要的任務就是盡可能提供管道，讓全美各地的歷史學者都能熟悉現有的數位資源。²

《紐約時報》的文章可以反映美國學界的趨勢，不過，數位人文的發展不局限於美國，也不止於英語世界。相反地，我們要說，這是一個全球學界的趨勢。在本次國際研討會中，我們邀請了來自英國、美國、澳洲、日本與南韓等地，帶領數位人文研究的學者，為與會來賓介紹各地的發展趨向。從他們的報告中，可以察覺到不同國家的數位人文研究各有側重。比如澳洲重視本土原住民文化的保存與研究；而南韓的數位人文，除了來自政府與學術機構的支援外，也與流行文化產業有密切的關係；來自日本立命館大學的研究群，則介紹他們為了重建千年古都——京都的空間，發展出的數位技術。³ 這些發展都與在地的文化結合在一起，因而各自有其面貌。但他們所共享的理念，則是透過數位科技，轉化人文研究的視野。

過去幾年內，數位人文在世界各地成為愈來愈顯眼的研究趨勢。這股趨勢可以從三個具體的層面說明，分別是機構、會議與期刊的建立。

這幾年，不少大學已經陸續成立了以數位人文研究為目的之單位。（Zorich, 2008）其中歷史較久的是 George Mason University 的歷史與新媒體中心（Center for History and New Media）。此單位成立於 1994 年，從名稱中不難看出，彼時「數位人文」尚未登場，因此他們側重於歷史研究與新興媒體的結合。這樣的發展不是孤例，各地數位人文的中心許多是由原先類似機關演變而來。倫敦國王學院（King's College London）是另一個例子。該校原本有「人文計算中心」（Centre for Computing in the Humanities），而今則改名為「數位人文學系」（Department of Digital Humanities）。組織上也有相應調整，成立研究所，招收數位人文的研究生。（Spaeth, 2003）

1 包括 3 月 22 日的 Giving Literature Virtual Life、12 月 3 日的 Analyzing Literature by Words and Numbers、12 月 6 日的 In 500 Billion Words, New Window on Culture 等。

2 http://www.historianstv.com/conference/meeting_2011/

3 演講投影片可見於會議網站：<http://www.dadh.digital.ntu.edu.tw/Scope.php?LangType=tw&His=D09k2>

這種現象甚為普遍。以美國為例，西岸的史丹佛大學有史丹佛人文實驗室（Stanford Humanities Lab）、加州大學洛杉磯分校有數位人文中心；東岸的哈佛大學在 2008 年推出數位人文先導計畫（Digital Humanities Initiative），2010 年則再進一步成立人文 2.0（Humanities 2.0）實驗室。這些單位儘管經營型態各異，但都是以數位人文為研究核心。他們面向校內的人文系所，發展數位研究工具、建立討論平臺或是倡議跨領域的合作計畫。

若把眼光放回亞洲。在日本，前述的立命館大學，也有日本藝術與文化之數位人文中心（Digital Humanities Center for Japanese Arts and Cultures）。臺灣大學的數位典藏研究發展中心，英文名稱則是 Research Center for Digital Humanities（數位人文研究中心）——這應該是臺灣第一個以數位人文為名的研究單位。

除了機構外，定期性的會議也是另一個數位人文建制化的指標。一如臺大所舉辦的「數位典藏與數位人文」國際研討會，日本學界也發起類似活動，希望聯繫和匯聚國內數位人文的研究能量。但國際間最具規模的，應該是一年一度的數位人文年會。此年會的前身原是文學與語言學計算學會（Association for Literary and Linguistic Computing, ALLC）和計算與人文學會（The Association for Computers and the Humanities）每年定期的年會。自 2006 年開始，此會議正式改名為「數位人文」，在歐洲和美洲輪流舉行，到目前為止舉辦的地點分別為巴黎、伊利諾、芬蘭、馬利蘭、倫敦與史丹佛大學。⁴ 儘管以上名單仍局限於歐洲與北美，但從主辦國的分布，足以再次說明數位人文的發展，不是一時一地的孤立現象，而是國際社群共同關心的主題。

除了研討會，數位人文的期刊也在近年內陸續發刊。在這其中有些期刊直接拋棄傳統的紙本模式，而採取線上出版。《數位人文季刊》（*Digital Humanities Quarterly*）⁵ 就是一個例子。這份刊物自 2007 年至今，已經出刊十一期。發刊時間雖不算穩定，而涵蓋的議題則甚為廣泛，除了資訊探勘（data mining）等技術面討論外，也有如 2009 年春季號中「如何將數位人文的計畫完成（Done: Finishing Projects in the Digital Humanities）」這類的專題。（Kirschenbaum, 2009）後者反映數位人文的研究計畫往往規模龐大、牽涉多端，不只是要處理大批資料，更可能有眾多背景殊異的人員參與，這些因素對於計畫的完成在在形成挑戰。

當然，除了隨著新趨勢應運而生的期刊外，不少數位人文研究的文章，發表在其他較有歷史的刊物中。譬如牛津大學發行的《文學與語言學計算》（*Literary and Linguistic Computing*）是一個重要的發表場所。此期刊由文學與語言學計算學會、計

4 見其會議網頁：<https://dh2011.stanford.edu/>

5 <http://digitalhumanities.org/dhq/>

算與人文學會和加拿大的數位人文協會（The Society for Digital Humanities）共同主持，因此與前述數位人文年會有著密切關係，只不過在刊名上並未採取相應調整。

如果像《文學與語言學計算》這類由來已久的期刊，就能夠負載數位人文的研究成果，或許讓人想問：數位人文究竟是新的還是舊的？我們應該這麼說，數位人文畢竟不是無中生有的領域，回顧歷史，我們仍可以追索出它的前世今生與來龍去脈。本書所收錄由王曉光與稻葉光行共同撰寫的文章〈數位人文研究的結構與演化〉，就利用數字和圖表，描繪出「數位人文」一詞，如何逐漸取代原本人們慣用的「人文計算」（humanities computing）。從這個角度看，數位人文必然從後者的成果中，汲取不少概念與靈感。值得一提的是，在臺灣學界中，數位人文的發展或有著不同軌跡。「人文計算」在臺灣似乎算不上流行，反倒是「數位典藏」所累積的龐大資源和能量，成為數位人文在臺灣發展的動力。

除此之外，儘管「數位人文」有其發展淵源，但一個新詞彙的誕生必定承載了一些新的概念。與一度風行的「人文計算」相比，「數位人文」至少有三個有別於前者的特點：第一、它擁有的資料量更加龐大；第二、它所強調的不只是計算，不認為計算能解答所有人文的問題；第三、它重視資料的開放與分享，且努力降低進入領域的門檻，擴大影響層面。在以下的三節中，我們會進一步觸及這三個特點。

二、數位化的年代

大量的數位資料是數位人文研究的基礎。但除了原生數位（born digital）的網路資料外，許多研究所需的數位資料並不會憑空出現。這就有賴資源與人力的投入，將實體的資料加以數位化。2004年起，以歐盟為主體的「歐洲研究基礎建設策略論壇」（The European Strategy Forum on Research），結合英國、法國、德國、荷蘭、丹麥等眾多國家之力，合作推動「藝術與人文的數位研究基礎建設」（Digital Research Infrastructure for the Arts and Humanities，簡稱 DARIAH）。計畫的簡介中寫著，一如天文學家需要天文臺觀測宇宙，藝術與人文學者也需要相應的研究基礎建設。這是 DARIAH 成立的目的，也可以用來說明各國數位化工作的初衷。

臺灣的讀者應該對國內「數位典藏」的計畫不會生疏，甚至曾經親身參與。數位典藏是臺灣幾個大型國家型計畫之一，自 2002 年發軔至今，已經有將近十年的歷史。最初計畫名為「數位典藏國家型科技計畫」（National Digital Archives Program, NDAP），初始有九個學術機構參與，分別為中央研究院、國立故宮博物院、國史館、國立臺灣大學、國立自然科學博物館、國立歷史博物館、國史館臺灣文獻館、國家圖書館及文化建設委員會。同時，亦公開徵選計畫，研究及學術機構的學者都能申請個別的計畫。國內許多機構的重要藏品，自此進入數位化程序中。自 2008 年起，

原本的計畫與「數位學習國家型科技計畫」結合，而成為「數位典藏與數位學習國家型科技計畫」（Taiwan e-Learning and Digital Archives Program, TELDAP）。⁶

數位典藏國家型計畫的參與單位均為政府機關及學術機構，數位化的內容也集中在珍貴的中華文化藏品與歷史檔案，但民間庶民文化的部分，就無法觸及。約莫同一時間，臺灣的文化建設委員會也啟動另外一項大型數位化計畫，名為「國家文化資料庫」，與數位典藏國家型計畫的不同點在於數位化的對象以臺灣的藝術作品、文獻與庶民文化為主。⁷ 結合數位典藏國家型計畫所涵蓋的中華文明「大傳統」和國家文化資料庫涵蓋的臺灣文化「小傳統」，基本上臺灣的文化資產和傳承已有大概的數位保存與呈現。

臺灣並不是特例。許多國家也同步展開大規模的數位化計畫。不過，臺灣的數位典藏計畫，最初以博物館和學術研究機構為中心；但在許多其他國家，大型圖書館扮演了數位化的先驅角色。比如美國國會圖書館（Library of Congress）的「美國記憶」（American Memory）就是一個好例子。自上個世紀 1990 年代開始，美國國會圖書館便有計畫地將館內藏品逐步數位化，並轉由 CD 儲存。不過，工作成本仍舊偏高，存取也相對不便。到了二十世紀末，隨著網路科技的進步，國會圖書館遂轉向建置以「美國記憶」（American Memory）為名的數位圖書館。顧名思義，「美國記憶」集中在與美國歷史和文化相關的資料。根據官方數字，至今已經有九百萬筆的資料上線，開放供讀者使用。⁸

日本的國會圖書館也於 1998 年（平成 10 年）提出「電子圖書館構想」，將館內明治、大正和昭和前期的圖書加以掃描，並提供數位的目錄索引。到 2010 年為止，總共有 47 萬冊的日文圖書完成數位化。由於著作權的問題，其中有 30 萬冊圖書只能在館內使用。即便如此，其餘 17 萬冊書籍已經全部上網，可以在世界各地免費連線閱覽。⁹

美國與日本的國會圖書館，只是數位圖書館發展中的兩個例子。在維基百科上所收錄的數位博物館計畫清單，項目就超過一百個。而這份清單顯然是不完整的，譬如臺灣在這方面的成果就完全沒有被提及；而日本，也只有 1997 年開始的「青空文庫」被列入，沒有提及日本國會圖書館，另一個重要的「亞細亞歷史資料中心」，也未被列入。¹⁰

6 這個計畫至今的成果，可見於其成果入口網：<http://digitalarchives.tw/>

7 舊版的網站位於：<http://nrch.cca.gov.tw/ccahome/>，臺大數位典藏中心則提供了另一個新版系統：<http://newnrch.digital.ntu.edu.tw/prototype/index.php>

8 <http://memory.loc.gov/ammem/index.html>

9 <http://kindai.ndl.go.jp/index.html>

10 「亞細亞歷史資料中心」（<http://www.jacar.go.jp/>）的內容，來自日本國立公文書館歷經兩年籌備後，開始數位化國立公文書館、外務省外交史料館以及防衛省防衛研究所圖書館，三所圖書館館內保管的檔案。涵蓋範圍從明治時代到第二次世界大戰結束為止。

除了官方所支持的數位圖書館或數位化計畫外，許多商業公司也投身其中。他們所建置的資料庫多半所費不貲，個別研究者不可能負擔得起。大學或其他公共的圖書館，因此成為主要的購買者。近年來電子圖書或電子資料庫的訂購，與實體書籍和期刊的購買，因為經費的排擠效應形成一些緊張的關係。這也是當代圖書館必須面對的棘手問題。（Darnton, 2008; Grafton, 2009）

在這方面值得一提的例子，是兩項中國古籍數位化工程：分別是「四庫全書」和「中國基本古籍庫」。文淵閣四庫全書的電子版是由香港迪志文化出版公司所推出。在傳統中國的大型叢書中，《四庫全書》是第一套被數位化的。推出之後，自然也引起了學界的注意。但對研究者而言，《四庫全書》固然名氣響亮、內容豐富，就版本而言卻有不少缺點。主要原因是《四庫全書》的編輯過程經過審查，不少內容遭到刪除或修改，對於歷史研究而言，反而是另一層使用上的障礙。這是商業公司或未考量的面向。

2001年開始，北京大學等校與北京愛如生公司攜手，推動另一項更大型數位化計畫，名為「中國基本古籍庫」，號稱涵括上萬本中國古籍，超過17億字的全文。相形之下，《四庫全書》的八億字也只能算小巫見大巫了。而「中國基本古籍庫」還只是愛如生所推出的各式產品中的其中之一。2008年主持這個計畫的北京大學歷史系教授劉俊文到臺北演講，分享數位化的經驗及其困難。演講中也提及未來將陸續推出諸如中國方志庫、敦煌文獻庫等不同的古籍資料庫，並因應古籍的內容，改進檢索與瀏覽的技術。

全世界以數位化為營運目的的商業公司，自是多不可數，無法在此一一介紹。但有一家特別的公司卻是不能略過的，那就是Google。Google原本以其獨特的演算法，在網路檢索的市場中獨占鰲頭，並且逐步擴張版圖，延伸到社會生活的各個面向。2002年，Google另一項前所未有的大規模計畫開始發軔，即後來的Google Books。按照原訂計畫，Google要建立世界上最大、最全面的數位圖書館，將人類有史以來出版過印刷書籍，全數掃描上網。2004年，Google與英美幾所大學圖書館——包括牛津、哈佛、史丹佛與密西根——簽訂合約，要將館藏書籍加以數位化。這個看似異想天開的行動，立刻引起了矚目，當然也有爭議。不少出版商對此大表不滿，認為這個行為牽涉到複雜的著作權問題，而Google並未提出妥善方案。這項烏托邦計畫，就在一片爭議聲中開始進行。時至今日，Google Books已經可以提供七百萬本圖書的全文檢索，儘管只有部分能提供整本書的瀏覽，但這個數量仍是十分驚人。（Darnton, 2009; Grant, 2006; Roush, 2005; Stross, 2008）

以上之所以花費這麼多篇幅，不厭其煩地細數各地數位化的成果，目的只在說明一個簡單的事實：我們生活在一個資料量前所未有龐大的時代。這個時代的數位人文，也因此與前一個世代的人文計算有顯著的差異。今天能取得大量資料的，不

再限於有龐大研究經費的研究者；任何一個人連上網路，就能利用如此龐大的資料，而且不只是文字資料，還包括了影像、音樂等各種媒體形式。對於人文研究而言，這應該是一個正面的發展。

一如王泰升教授在本書的文章所言，資料庫與研究者的關係，有如「你是我的眼」。在資訊技術的協助下，研究者宛如快速看過了大量的資料，並迅速找出所需的部分。這樣的改變，是經常要讓上一個世代、以土法煉鋼尋找資料的研究者大感驚訝的。¹¹ 臺灣清華大學歷史研究所教授黃一農，循此提出「E—考據」的概念。（黃一農，2005）現有的資料檢索功能，搭配以如此大量的資料，確實使得傳統曠日費時、披沙揀金的考據工夫，進入全新的境界。

但大量的資料也造成新的挑戰。王泰升教授在文章中指出，資料庫使用的本身可能存在門檻，如「日治法院資料庫」以日文寫成，有時筆跡甚為潦草，未經特別的訓練不易入手。研究者本身的基本功力是否足以駕馭史料，這是學術研究由來已久的問題了。但資料庫將這個問題帶到一個新的層次，借用 Roy Rosenzweig（2003）的話來說，這個時代的研究者面對不是資料匱乏，而是另一種「富裕的難題」（problem of abundance）。過量的資料要如何處理？究竟其中有多少是有意義的？或者，如何讓大量的資料產生新的意義？還有，過去逐筆細讀資料的方式，是否還能適用？如果不行，需要發展出什麼樣的新取徑？凡此種種，都是數位人文要處理的根本問題。

上述問題或許主要是人文學者需要面對的問題（O'Donnell, 2009），但負責建置資料庫、熟悉數位科技的資訊學者，在此也可扮演關鍵的角色。至少，從資料庫的設計上，就可以發展出更貼近使用者需求的技術。可惜的是，以往攸關建置資料庫的討論，多半局限於資訊技術本身的細節。但新穎的資訊技術，有多少能夠回應、貼合到研究者的需求？

關於這一點，我們在不同的文章中曾經反覆論及。（Chen, 2007；項潔，2009）大體而言，我們認為研究者的資訊檢索需求與一般的使用者不同，前者所尋找的經常不是單一的、片段的資訊。他們更在意的是資訊的內在和外在脈絡；此外，研究者也要觀察資料所透露出的「整體意義」。當研究者能取得的資料量愈大，這一點就更形重要。因此資料庫的建置者在這一點上，不能止於對資料內容做索引、對檢索結果排序，或者僅在使用界面上改進。換言之，Google 在網路檢索上的成功經驗，對特定領域或特定使用者並不一定適用。這並不意味著前面幾點不重要，只

11 臺大歷史系陳弱水教授講過這樣的例子：「一九九七年末，我在日本東京大學做研究時，一位年輕的中國思想史教授告訴我，他的一個學生在舉行學生論文口試時，一位年長教授對他的論文中表現的博學非常驚訝，他不知道，這些博學表現其實都是從電子文獻資料庫查來的。由此可見，資料庫的威力驚人，它可以使一位異國的大學部學生輕易掌握廣泛的中國古典資訊。」（陳弱水，1998，〈試談電子文獻資料庫在歷史研究上的應用〉，《古今論衡》，1，頁 53-57。）

是，資訊人員應該重新思考研究者的需求，並將資料的脈絡或整體意義，呈現在使用者面前。

一個能夠符合研究者需求的資料庫，或者，用我們的詞彙，一個「數位研究環境」，應該是數位人文進步的關鍵。也就是說，單單有資料是不夠的，資料需要有良好的工具，才能被確實運用、發揮潛力。

三、新方法與新視野

在這一節中，我們要介紹幾種具體的研究，都是利用不同的資訊技術，從大量數位資料挖掘出新的意義。他們所採用的方法，可以應用在不同的材料上。換言之，從方法論的角度，這些例子應該具有一些啟發性。

首先是「詞頻分析」，簡單的說就是計算文本中所出現詞彙的次數。在英語世界的期刊或會議中，常可以看見文學相關系所出身的學者，投身數位人文研究之中。但文學研究者可以從數位技術中得到什麼？「詞頻分析」是其中的一個例子。乍看有些不可思議，難道僅僅是計算文章中詞彙的頻率，可以提供研究者對於文學作品新的視野嗎？

以莎士比亞為例，所有以他為名的作品，是否都是出自莎士比亞一人之手？這向來是聚訟不休的問題。嘗試解答的方式有很多種，譬如分析莎士比亞的生平，尋找他代筆的證據；或者從文本本身分析寫作風格在各作品之間的差異。就後者而言，難免有陷入主觀困境的可能：寫作風格如何判定？這時詞頻的統計就可能派上用場。不同的作品、甚至不同的章節，是否有一致的慣用詞彙？雖然這類的統計不見得能直接導出結果，但至少可以提供討論的基礎。（Craig, 2004）

同樣的方式也可以應用在中國文學作品中，如《紅樓夢》後四十回是否為曹雪芹所寫，或是由他人續作——這是另一個聚訟不休的問題。而早在上個世紀便有學者嘗試用電腦統計紅樓夢的詞頻，試圖為這個爭論提出新的角度。

大體上看來，中文與英文的詞頻統計是同一套模式：列出所有文章中出現的詞彙，再統計其出現次數。但實際操作時，中英文間卻需要面對不同的技術性問題。比如在英文的同一詞彙，可能因為文法而有不同型態的變化，如單、複數的差別。在進行統計之前，就必須先對詞彙做正規化（如 *stemming*），以避免統計結果的錯誤。中文不存在這樣的問題，但卻需要先處理「斷詞」。在中文的脈絡中，怎樣算是一個「詞」？這個問題在中文世界裡十分複雜，無法直接挪用英語世界的作法；但又十分關鍵，因為這牽涉到詞頻統計的基礎。

解決方式之一，是直接忽略語意的問題，而讓任意相鄰的字都視為統計單元，

即所謂的 n-gram。以「數位人文」為例，語意的判斷會將這四個字斷成「數位」「人文」，但如果採取 bi-gram 的斷詞，即其中任兩個連續出現的兩個字都會被視為一個詞，則「數位人文」會被切成「數位」「位人」「人文」等三個詞。這方法的優點是可透過電腦自動處理全文，而減少人力；另一個好處是避免人為主觀的判斷，以較為中立的方式進行斷詞，並進行統計。

依照詞頻統計所做的研究，不只可以用於風格分析，還有可能進一步發展。試想，每個人寫作（或發言）時所使用的詞彙，不只是寫作風格的差異，也反映著每個人思考的傾向和方式。從這個角度來看來，藉著寫作風格分析，其實也是在捕捉思考的風格。

有人採取這種方法，分析歷任美國總統就職或各種重要演說的內文，再以標籤雲（tag cloud）的方式呈現。¹²但僅僅以單場演講的內文，內容未免還嫌少，並不真的能發揮電腦處理和呈現複雜資料的能力。如果我們要將這個方法應用到不同研究中，或可以著眼在更大量的資料上。比如，如要分析臺灣總統的演說內容，則不只可以計算就職演說一文，還可以將其他場合的演說辭一併納入。又或者，如果我們計算近五到十年來臺灣四大報社論所出現的詞彙，很有可能看見一些有趣的趨勢或彼此的差異。再進一步想像，這樣的作法不僅可以適用於當代文本，甚至可應用於歷史資料。比方說，既然可以分析當代總統的演講詞，是否也可能循此模式，分析清朝歷任皇帝的「上諭檔」？

當然，在此必須要強調，詞頻統計的結果，並不能、也不該直接導致任何結論。它最多是呈現出某種現象，現象背後潛藏了什麼意義，如何詮釋，都需要人文研究者的介入。只是，借用電腦計算出的詞頻，時常有出現意想不到的結果，舉例而言，我們曾嘗試在臺灣歷史數位圖書館（Taiwan History Digital Library, THDL），檢索「朱一貴」三字，卻意外發現人名的詞頻統計中，除了與朱一貴事件直接相關的人物外，竟有不少林爽文事件的相關人物。這似乎暗示二者之間有些關聯。可是這關聯是什麼？如何解釋這樣的現象？就需歷史研究者提出更為深入的論述。

近年寓居臺灣的學者金觀濤與劉青峰兩位教授，也是運用詞頻統計的方式，重新探索中國近代思想的發展。他們先是將大量近代發行報刊內容進行數位化，再針對近代中國一些重要的詞彙，諸如民主、科學等，利用資料庫進行詞彙頻率統計。用他們的說法，這是用「關鍵詞」去捕捉「觀念史」的發展；這樣的方法本身並非新發明，而是傳統觀念史研究中熟悉的取徑。只是在資訊科技的協助下，他們所能分析的文本大幅度擴張，所掌握的現象也和過往的研究不可同日而語。循此，他們也對一些對於近代中國思想印象式的說法，提出商榷和修正。（金觀濤、劉青峰，2008）

12 <http://chir.ag/projects/preztags/>

同一時間，在英語世界也有人在進行類似的研究——不過規模更大，因為他們所運用的實驗資料來自 Google Books，共計五百萬本書籍。在此基礎之上，他們探索的範圍也更廣。特定的觀念詞彙當然仍是觀察重點，但他們用了琳瑯滿目、類型各不相同的詞彙。比如 God，比如 men 與 women 的比較，又比如 evolution、cell、bacteria 與 DNA 的比較。同時，他們可以此探索人物名聲的起落，比如他們在研究中舉的例子：Galileo、Darwin、Freud、Einstein。（Michel et al. 2011）這些詞彙呈現曲線，是否都有意義？恐怕不盡然，不同的讀者或可能想到更富深意的例子。不過從方法論的層次，他們提出了一個有趣的嘗試。不久前 Google 也推出 Books N-gram Viewer 的功能，對於這類研究方法感興趣的研究者，都可以輕易地在網路上操作。¹³

數位人文的另一種可能，是發現大量資料內部隱含的關係，也就是應用資訊科技中的資訊探勘（data mining）或文本探勘（text mining）的技術。這類技術應用的範圍很廣。比如在商業範圍內，資料探勘可以用來分析顧客的行為模式，以利對後續的服務，提供參考的基準。比如網路書店中常用的商品推薦機制，即是利用大量的購買紀錄，去推測出可能的購買模式或偏好。

這種尋找關聯式規則的研究，只是資料探勘的其中一種方式。在資訊工程的教科書中，資料探勘的主題還包括諸如分類（categorization）與分群（clustering）、推估與預測（estimation and prediction），還有連結分析（link analysis）等。背後各自有其演算的理論與模型。在此我們無需重複資工教科書的內容，因為要直接套用現有的技術，諸如分類、分群或是關聯式規則等，然後到人文研究中尋找題目，其實並不容易，也未必是好的模式，反而可能往往會形成削足適履。可是，如果從材料的角度出發，再嘗試不同的技術，或許是比較好的方式。以下我們從幾個角度來討論這件事。

以我們自己進行的研究為例，過去我們數位化臺灣的土地契約文書與明清中央政府的行政檔案。從典藏的角度，可以計算出他們個別的數量，如土地契約文書有約三萬件，而明清行政檔案有超過三萬五千件。可是從研究的角度，更重要的是個別文件之間的「關係」。例如，土地契約反映土地買賣的交易關係；明清行政檔案，反映中央與地方之間訊息流通與政策商議的過程。循此，我們設計出兩種不同的自動方式，先是擷取文件中的特徵詞彙，再進一步重建這兩種獨特檔案間的關係。這種方式可以串連出檔案之間的故事，歷史學者則可以把這些檔案間的故事，置於更大的歷史脈絡之中，進行學術性的分析和討論。這類數位人文的取徑，並無法取代歷史研究者的角色，但可以縮減原先研究工作中，比較重複和無趣的部分。（黃子鳴，2009）

13 <http://ngrams.googlelabs.com/>

除此之外，資訊科技也可以協助研究者找出複雜的人際網絡關係。哈佛大學與北京大學等校合作的計畫——中國歷代人物傳記資料庫（CBDB）就是一個例子。這個計畫最早是由美國的歷史學者 Robert M. Hartwell 所提出，以宋代人物的傳記為中心，想在累積大量的資料之後，進行群體傳記學（prosopography）的研究。¹⁴ 群體傳記學是想找出某一個群體所共享的身分，比如他們的教育程度、出身背景乃至宗教信仰等，藉此對社會現象進行分析。（Stone, 1971）這個作法至今仍有一些學者在不同的議題上應用，比如清史學者 Kent Guy 就利用群體傳記學，研究清朝中葉以前的巡撫，他統計巡撫的籍貫、教育背景等，提出了一些觀察，比如這些巡撫大多在科舉考試中不是太成功；而對於擔任巡撫的人而言，他們差不多都在官職生涯中期擔任這個職位，是他們進一步升官的中途站。（Guy, 2011）對於這些結論，以往的研究者或有一些模糊的印象，但唯有透過數據，才能讓這些論斷更為堅實。

不過隨著資訊技術的進步，CBDB 從群體傳記學，又轉移到了另一種研究取向，也就是人際網路的分析。後者看中的不是人物群體共享的特性，而是許多一對一關係組構而成的複雜網絡。人物間的關係，是歷史研究向來關注的問題。不過，以人工分析，往往只能局限在比較明顯的、單純的關係，一些比較隱晦的、複雜的關係，可能就難以兼顧。可是透過電腦的輔助，歷史學者有可能去觀察人際網路中不同節點的互動，並提出新的論斷。比如參與 CBDB 計畫的研究者，有人以此描繪宋代思想學派間的互動，並指出一些在思想史上不見得突出的學者，可能反而扮演了不同學派間溝通的橋梁。

在 CBDB 發展的同時，Peter K. Bol 也領導了另一個計畫，即中國歷史地理資訊系統（Chinese Historical GIS）。¹⁵ 結合群體傳記學，地理資訊系統可以指出一些有趣的歷史現象，比如他們依照南宋中後期進士登科的出身背景，繪製地圖，意外地發現其中有許多來自四川盆地東部的進士。又或者，他們以宋代中國各地的商稅額繪製地圖，再與北宋科舉進士出身的地理分布相比較，進而探索商業與地方教育文化的關係。¹⁶ 這些都是尚在發展中的研究，還有發展的空間，但他們說明結合了 GIS，歷史學者可以開始著手許多傳統難以處理的議題。

過往歷史學者擅長從時間上觀察歷史演變，對於地理空間分析則相形見绌。但有了 GIS 的結合，由地理角度的觀察會更為容易，相信可以鼓勵學者往此方向前進。（Gregory, 2007; Knowles, 2002）

以上我們舉了一些數位人文研究的例子，當然並不完整。不過，從上面的討論

14 <http://isites.harvard.edu/icb/icb.do?keyword=k35201&pageid=icb.page145374>

15 <http://www.fas.harvard.edu/~chgis>

16 以上見 CBDB 網頁上的介紹：<http://isites.harvard.edu/icb/icb.do?keyword=k35201&pageid=icb.page149054>

可以看出，數位人文並不是單純而粗糙地希望藉著電腦來回答問題，而是嘗試以資訊技術找出某些值得進一步深入研究的現象。在這樣的過程中，人文學者的角色不能被取代。如何從這些現象中，找出具有深刻意涵的問題，仍然有賴人文學者的眼光。換言之，數位人文是一個資訊技術與人文不斷互動的過程，二者缺一不可。

四、連結與分享的世界

數位科技改變的不只是人文學者研究的方法，更改變了整體的學術環境。數位時代產生變化的不只是研究，它也改變我們傳遞、溝通和交換知識的方式，這包括教學與學習，研究的呈現，還有知識社群的型態。而以上種種其實是互相聯繫一起的。在臺灣，數位典藏與數位學習的匯流，就是例證之一。

在這其中，網路無疑扮演了極為重要的角色。今日的網路使用者，可以輕易登上 YouTube 或類似的網站，觀看成千上萬來自世界各地的開放課程與演講錄影。網路使用者，也可以在維基百科中，找到無數的詞條資料，從此他們無需購買大部頭的百科全書，也無需擔心百科全書內容無法反映最新的知識。他們還可能找到各式各樣的專題網站，介紹人文（乃至科學）的各種知識。透過網路線，人們可以連結到了一個無限寬廣的世界。

但除了這種樂觀的宣稱之外，對於網路以及數位時代，當然也有不那麼正面的意見。比如，人們擔心網路缺少了深入的知識，充斥片段的、甚至錯誤百出的資訊。這些錯誤百出的資訊，卻可能反而有著極高的影響力和傳播性。種種反對意見，有些合理，有些可能是杞人憂天，但網路對於人類生活的持續影響，恐怕是很難停止的。因此，與其用悲觀或懷舊的姿態哀悼傳統世界的消逝，或許不如思考人文學者對於未來的學術環境能有什麼正面的介入和影響，而不只是被科技的進步推著走。

以維基百科而言，這個 web2.0 時代的產物，應該在教學中扮演什麼角色，時常引起正反兩極的爭議。不過幾年前，教師抱怨學生從維基百科上抄襲答案，甚至有學校禁止學生使用維基百科。弔詭的問題是，查閱網路版百科全書，與查閱一本紙本的百科全書——比如著名的大英百科全書——究竟差別何在？為何過去教師們大力擁抱紙本百科全書，卻會對維基百科產生質疑？這樣的問題，隨著維基百科內容的日漸豐富和使用的日益普及，顯得愈來愈尖銳，挑戰著傳統百科全書的編著者、出版商與使用者。（Rosenzweig, 2006）

但維基百科在教學上也可能扮演不同的角色。比如，一門歷史課程的作業，可否讓學生共同撰寫維基百科的相關詞條？在學習撰寫詞條的過程中，學生不只是學習查閱資料，同時要練習如何把知識系統性地呈現出來。而因為維基百科本身是開放式的媒介，學生所面對的讀者因此不只是教師一人，還是全世界的讀者。從另一

個角度看，學生在撰寫作業的過程中，也可能修正原本網路上錯誤的資訊，而讓後來的讀者在查閱維基百科時得到更豐富且正確的內容。這份作業因此不只是為了學生個人的成績，還可能有更大的貢獻。

維基百科的例子說明，與其抱怨網路上的知識不盡正確，學院裡的研究者不如採取正面的行動，讓網路成為知識分享的平臺。在美國近年來發展的數位歷史（digital history）中，就十分重視這一點。他們所謂的數位歷史，更接近帶著推廣性質的公眾史學（public history）。藉由設立專題網站，平時蝸居學院的歷史學者，嘗試讓一般人也能接觸到第一手的史料，或是專業史家提出的論述。前陣子兩位美國喬治梅森大學（George Mason University）的歷史學者 Daniel J. Cohen 和 Roy Rosenzweig（2006），合撰的導論書《數位歷史》，最能說明這一點。書中內容幾乎全都集中於資料的數位化，以及網站的建置。關於前一節所介紹的，以資訊技術對大量史料進行分析和研究的取徑，反倒少有著墨。不過，他們對於公眾史學的重視仍有其重要性。學者就指出，過去設立的美國歷史網站，在九一一事件過後，流量頓時大幅增加。對此，他們認為類似這樣特別的事件，讓人們對於自我認同感到疑惑，刺激他們開始探索過去，以面對眼前看似難以理解的問題。也是在這樣的時刻，公眾歷史最能發揮其作用。這一點對於臺灣歷史的研究者或許格外有啟發。當臺灣研究在學術界內已經站穩腳步，如何把豐富而嚴謹的成果，提供給感興趣的公眾？這應該是個重要的問題。（Ayers, 2004）

從上述數位歷史的發展，我們可以進一步提出兩點。第一，公眾歷史的概念未嘗是新的，一直以來，都有學者企圖跨出象牙塔，將歷史學的研究成果普及化。過去，許多國家也都出版不少普及性的歷史期刊或書籍，至今仍是如此。可是數位時代給了歷史學者一些新的可能性，多媒體的運用即是其中之一。網路媒介可以提供紙本書籍更多感官性的素材，諸如視覺或是聽覺的材料。同樣是在美國，一些研究者已經開始探索這些新穎的發展。比如以美國麻省理工學院為首的「目擊文化」（visualizing cultures）計畫，就召集了一批東亞歷史的研究者，利用各式各樣的圖片，呈現出近代中國和日本的歷史發展。由於成果精緻，內容又可靠，非常受到關注與好評。¹⁷ 而曾任美國歷史學會會長的 Robert Darnton（2010），則是對聽覺材料情有獨鍾，他最新的著作《詩歌與警察》（*Poetry and the Police: Communication Networks in Eighteenth-Century Paris*），描述批判政府的地下詩歌，如何在十八世紀的巴黎流傳。為此，他還特別找來歌手，重新錄製當時的地下歌曲，試圖讓讀者更能感受到書中描述的現象。

隨著通訊媒介的發展，這種以視覺、聽覺，乃至於其他材料，搭配傳統文字敘

17 <http://ocw.mit.edu/ans7870/21f/21f.027/home/index.html>

述的表達方式，或許會愈來愈流行。換句話說，不只歷史學者，還有廣義的人文學者，都要面對一種呈現研究的新方式——只要想想各種會議場合無所不在的投影片就好了。儘管這種新方式的影響，可能優劣互見，端看個人運用，但顯然是無法反轉的趨勢。

利用聲光化電的方式來表述研究成果，不單只是比較花俏，或者只是吸引觀眾的注意力——儘管這也十分重要——可能還有更深一層的影響，也就是我們對於研究方法和學科本質的認知。以歷史學研究為例，研究者愈來愈注意到圖像在各類題目中的重要性。它們不只是藝術研究的對象，或者是文本描述的附屬品，而是本身需要詮釋的材料。圖像所記錄和傳達的訊息，與文字之間不見得合縫合隙。圖像和文字之間，可以既是互相補充，又是充滿張力。另一方面，視覺與聽覺等各方面的材料，可能喚起讀者或聽眾在理性外的感官經驗，這和強調客觀中立的研究立場，或許有些衝突，可是如果歷史學強調是對過去世界理解和體會，那麼，圖像與聲音很可能比單調的文字敘述更為有效。（Fogu, 2009）

除了上面提及的現象外，公眾歷史給我們的另一個啟發，是網路時代核心的觀念：開放。這是一個愈來愈難閉關自守的時代。前一節中我們提到日本與美國國會圖書館的數位化工作，其成果都是完全開放，免費下載，除了著作權的因素外，幾乎不多加以限制——這應該是所有公立單位進行數位化的前提。甚至許多大學也是以此為指導原則，就我們眼見所及，如澳洲國立大學圖書館、日本京都大學圖書館、美國普林斯頓大學東亞圖書館。這幾所大學所數位化的古籍，都可以在網路上直接下載。還有我們前一節所提及，哈佛大學的兩個計畫：CBDB 與中國歷史地理資訊，也已經全數公開在網路上，供使用者下載使用。臺大數位典藏研究發展中心也是秉持同樣的想法，我們所建置的資料庫，全數免費使用；而除非資料提供者要求，我們亦盡量讓資料庫開放，使各地的研究者都能方便連線查索。

1970 年代，臺灣銀行經濟研究室的周憲文先生，發起《臺灣文獻叢刊》的編纂計畫。這套共有 309 種的叢書，至今仍是研究臺灣歷史重要的參考書。周憲文回顧其出版計畫時，這樣寫著：

研究歷史，一要有史料，二要有史觀；前者有賴資料公開的風氣，後者得憑個人獨特的修養。我們十多年來的工作方針，嚴格說來，就在盡量發掘並提供有關台灣經濟的研究資料。因為有充分的史料，社會上自然會有高明之士，運用其正確的史觀，深入研究，有所造就。我們願意為多數的學人服務，而絕不關心到小我（私人或機關）的成績。說明白些，我們固亦從事個別的研究，但願與大家在同樣的資料基礎上進行。我們堅信，個人能力畢竟有限，資料的公開是學術進步的前提條件。（川口長孺，1957：1）

這種信念的重要性，在網路時代更加明顯。尤其，前一節所介紹的資料探勘或文本探勘，總是必須建立在大量的「資料」或「文本」之上，否則也只是巧婦難為無米之炊。因此，研究者彼此的合作、連結，是非常關鍵的。在 2009 年的會議上，來自各地的數位人文研究者，對此都有相當的共識。相信未來在此領域中，不只是跨領域（數位、人文），還有跨地域的合作，都將愈來愈普遍。

五、代結語：數位人文的開端

本文廣泛地介紹了數位人文研究的成果，希望提供出一個背景圖像，讓讀者理解目前這個領域的趨勢。一如我們在前言中所說，這是個年輕的領域，充滿活力，還在不斷成長。

本書所收錄的 12 篇文章，都是由 2009 的會議論文，經過審查、修改後的成果。這些文章的內容，呼應著我們前面所敘述的數位人文趨勢。第一個部分的三篇文章，分別來自臺大法律系的王泰升教授、臺大歷史系博士生陳志豪、以及臺大資工所博士生陳詩沛。這三篇文章從不同的角度回應同一個問題，就是在數位時代，歷史研究會走向何處。王教授的文章透過三個不同的資料庫，利用具體的實例，討論資料庫對歷史研究帶來的機會與挑戰。他特別強調許多資料庫的使用是存在門檻，但此門檻並非來自資訊技術，而是史料本身樹立的障礙。比如，以日文寫成的日治法院檔案，對於沒有受過訓練的讀者而言，往往不易解讀，這因此是任何研究者自身必須培養的基本工夫。

陳志豪與陳詩沛的文章，則分別從使用者和建置者的角度，集中討論臺灣歷史數位圖書館。陳志豪詳細地介紹了資料庫內容、功能以及使用方式。陳詩沛則在論文中，說明如何藉由前置的資料處理流程，提供使用者不同層次的檢索結果。這類雙向的對話，相信無論對使用者或是對建置者，都會有所助益。

本書的第二部分，是關於資料探勘技術的討論。前文曾經提及王曉光等人的文章，他們運用大量的資料，描繪出「數位人文」的概念在學術界的演變。劉吉軒與賴隆平的文章，則是另一種資料探勘研究的展演。他們利用 1970 年代到 1990 年代臺灣文官的人事資料，藉由資訊技術計算文官升遷的模式。

葉智豪等人的文章，則偏向資料探勘的技術面。前文我們曾經提到詞頻的計算及其應用，不過，如果用人工對文章做標注，將文章中重要的詞彙標示出來，無疑將耗費大量的人工。資訊科技在此可以成為協助工具，藉由人與電腦的合作，縮短這樣工作的時間。此文就介紹如何以半自動化的方式，將中文文章中的詞彙擷取出來，同時予以意義上的分類，如人名、地名、官名等。這項技術應該可以適用在不同的電子全文之上。

本書第三部分的兩篇文章，都在原本的數位資料之上，建設了進一步的研究平臺，讓使用者可以輕易地在其中探索和研究的。洪振州等人的文章，介紹了臺灣法鼓佛教學院近年來在數位典藏與數位人文領域的進展。他們不只數位化了大量的佛教典籍，從這一篇文章中，我們也看到進一步的應用。比如將《高僧傳》，或臺灣佛寺的影像，與地理資訊系統結合。齊藤進也等人的文章，則是介紹了日本立命館大學的成果。他們發展了一套名為 *kachina* 的系統，透過視覺化的方式，協助文化資產的數位化以及後續的分享。

當然，發展數位人文的同時，數位典藏的工作仍然必須繼續下去。本書的最後一部分，收錄了來自國內國外的主題典藏系統介紹，內容涵蓋阿富汗的特藏資料、西方典藏的日本陶器、臺灣的政府公報與人事異動資料，以及日治時期戶口調查簿的檢索系統。使用者可以從其中了解數位化的過程，或尋找所需的資料；建置的人則可以從中讀到不同類型的數位化經驗。

無論是上述本書所收錄的文章，抑或是本文所介紹的趨勢，相信都只是未來數位人文研究的一部分。他們是嘗試和摸索的經驗分享，但我們希望此書可以作為臺灣數位人文的開端，將來則期待看見更多更精彩的，甚至遠遠超過眼前所能預期的成果。

參考文獻

- 川口長孺，1957，《臺灣割據志》，臺北：臺灣銀行。
- 金觀濤、劉青峰，2008，《觀念史研究：中國現代重要政治術語的形成》，香港：中文大學出版社。
- 項潔、陳詩沛、杜協昌，2009，〈台灣古契約文書全文資料庫的建置〉，《第三屆台灣古文書與歷史研究學術研討會論文集》，臺中：逢甲大學出版社。
- 黃一農，2005，《兩頭蛇：明末清初的第一代天主教徒》，新竹：清華大學出版社。
- 黃于鳴，2009，〈臺灣古地契關係自動重建之研究〉，臺北：國立臺灣大學資訊工程學研究所碩士論文。
- Ayers, E. L. (2004). Doing Scholarship on the Web: Ten Years of Triumphs - and a Disappointment. *Journal of Scholarly Publishing*, 143-147.
- Chen, S.P., Hsiang, J., Tu, H.C., & Wu, M.C. (2007). On Building a Full-Text Digital Library of Historical Documents. In D. Goh, T. Cao, I. Sølvberg & E. Rasmussen (Vol. Eds.), *Lecture Notes in Computer Science: Vol. 4822. Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*. Berlin, Heidelberg: Springer.
- Craig, H. (2004). Stylistic Analysis and Authorship Studies. In R. Siemens & S. Schreiban (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell.
- Darnton, R. (2008). The Library in the New Age. *The New York Review of Book* 55, 10.
- Darnton, R. (2009). Google & the Future of Books. *The New York Review of Book* 56, 2.
- Darnton, R. (2010). *Poetry and the Police: Communication Networks in Eighteenth-century Paris*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Fogu, C. (2009). Digitalizing Historical Consciousness. *History and Theory* 48, 103- 121.
- Grafton, A. (2009). Apocalypse in the Stacks? The Research Library in the Age of Google. *Daedulus* 138(1), 87-98.
- Grant, J. (2006). Google Book Search: An Introduction. *Google Librarian Center Newsletter*.
- Gregory, I. N. & Ell, P. S. (2007) *Historical GIS: Technologies, Methodologies, and Scholarship*. Cambridge: Cambridge University Press.
- Guy, R. K. (2010). *Qing Governors and Their Provinces: the Evolution of Territorial Administration in China, 1644-1796*. Seattle: University of Washington Press.

- Kirschenbaum, M. G. (2009). Done: Finishing Projects in the Digital Humanities. *Digital Humanities Quarterly* 3(2). From <http://digitalhumanities.org/dhq/vol/3/2/000037/000037.html>
- Knowles, A. K. ed. (2002). Past Time, Past Place: *GIS for History*. Redlands, CA: ESRI.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team et al. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331, 176-182.
- O'Donnell, J. J. (2009). Engaging the Humanities: The Digital Humanities. *Daedalus* 138(1), 99-104.
- Rosenzweig, R. (2003). Scarcity or Abundance? Preserving the Past in a Digital Era. *American Historical Review* 108(3), 735-762.
- Roush, W. (2005). The Infinite Library: Does Google's Plan to Digitize Millions of Print Books Spell the Death of Libraries; or Their Rebirth. *Technology Review*.
- Spaeth, D. (2003). Research and Representation: the M. Phil in History and Computing. *Computers and the Humanities* 37(1), 119-127.
- Staley, D. J. (2003). *Computers, Visualization, and History: How New Technology Will Transform Our Understanding of the Past*. Armonk, New York: M.E. Sharpe.
- Stone, L. (1971). Prosopography. *Daedalus* 100, 46-79.
- Stross, R. E. (2008). *Planet Google: One Company's Audacious Plan to Organize Everything We Know*. New York: Free Press.
- Srinivasan, P. (1997). Digital Library Projects in the United States. *Bulletin of Information Technology* 17(6), 15-21.
- Zorich, D. (2008). *A Survey of Digital Humanities Centers in the United States*. Washington, D.C.: Council on Library and Information Resources.