



數位人文取徑下歷史研究的轉折與開展—— 以《自由中國》為例

薛化元、江子揚、蔡銘峰、劉吉軒*

數位保存概念導入臺灣後，推動數位典藏累積的研究資源及學術能量，成為開創數位人文（Digital Humanities）發展有力動能¹。數位科技應用，不僅轉換成知識傳遞、溝通與交換模式，同時亦逐漸形塑出整合研究與教學，以及構築知識社群的數位學習系統²。數位典藏整體計畫在 90 年代的啟動，進一步豐蘊其後數位人文學的根基³。數位人文研究在 2008 年前後，由於數位化典藏與資料庫系統建立等後，獲致突破性發展，透過文獻特性的描寫，即詮釋資料（Metadata）；以及文獻內容全文打字、校對的數位化前置工程，整合人文計算與數位底藏所累積之經驗與動能，進一步開啟數位人文之新頁⁴，並由於數位人文架構下各式技術工具的發展，如數位文本標引的標準化與格式化、詮釋資料觀念、地理資訊系統（Geographic Information System, GIS）方法等，人類文明世界進入嶄新的大數據（Big Data）時代⁵。

檔案文獻數位化效益，黃寬重在〈數位時代人文研究的衝擊與蛻變〉指出，將具備以下四項新生功能：「首先，典藏原件能夠在不影響公眾利用下，獲得更好的保存，甚至永續度藏。其次，經數位化後，文物典藏圖像資訊可

* 薛化元，國立政治大學臺灣史研究所教授；江子揚，國立政治大學俄羅斯研究所碩士生；蔡銘峰，國立政治大學資訊科學系助理教授；劉吉軒，國立政治大學資訊科學系教授。

¹ 項潔、涂豐恩（2011）。〈導論——什麼是數位人文〉，項潔主編：《從保存到創造：開啟數位人文研究》，臺北：臺灣大學出版中心，頁 14。

² 同註 1，頁 22。

³ 鄭文惠（2014）。〈從人文到數位人文：知識微縮革命與人文研究範式的轉向〉，《人文與社會科學簡訊》，15 卷 4 期，頁 170。

⁴ 項潔、翁稷安（2011）。〈導論——關於數位人文的思考：理論與方法〉，項潔主編：《數位人文研究的新視野：基礎與想像》，臺北：臺灣大學出版中心，頁 12。

⁵ 同註 3，頁 169。

透過便捷的網際網絡，不限次數送至任何地點，有效突破地域限制。再者，數位資料可透過建置後設資料與聯合目錄，將各類資料彙整分類，便於利用。其四，也是最重要的一點，數位資料能極快速搜尋檢索」⁶。檔案數位化工程，除可破除傳統文本檔案的單一脈絡現制，以及消弭檔案整理與研究間矛盾外，更得以有效建構一套以研究作為主要思考取向的資訊系統，不同檔案所底蘊的豐富且多重脈絡，將可從中獲得發現、呈現與被觀察⁷。數位化檔案系統所能提供的檔案脈絡，項潔歸結為六大要項：(1) 檔案原始脈絡；(2) 重組原始脈絡所產生的多重脈絡；(3) 鳥瞰型脈絡；(4) 子文件集的縱觀脈絡；(5) 文件間統計型脈絡；(6) 文件間隱藏的新脈絡⁸。

然而，就在檔案數位化工程進展的同時，數位人文研究仍須留意以下情況：首先，研究者如何管理駕馭豐富龐雜的資訊，即所謂「富裕難題」(problem of abundance)之困境；二者，數位資源可能對人文研究產生衝擊之處，首先是數位資料庫的品質，資料庫品質的差別將直接影響到研究成果；三者，則在於研究者對數位資訊的詮釋能力，亦即若研究者缺乏專業訓練與先備知識，將影響推論結果，進而損害研究品質；四者，以數位技術承載知識，則免不了出現新的知識利用門檻，亦即研究者是否具備數位資源的掌握能力，將決定資訊時代下其所面臨之衝擊與命運⁹；五者，在科技整合過程，如何有效弭平數位落差的情況，即人文學者對於資訊專家及其研發技術工具的過度依賴。¹⁰

置於數位人文脈絡下的史學研究歷程，有兩個重要的面向：一、計量史學(Cliometrics)的統計方法嘗試；二、數位典藏以及資料庫的建構¹¹。而數位人文在歷史研究的面向，研究者可藉由資訊技術的應用，從大量繁瑣的史

⁶ 黃寬重(2012)。<〈數位時代人文研究的衝擊與蛻變〉，《漢學研究通訊》，31卷1期，頁2。

⁷ 項潔、翁稷安(2012)。<〈多重脈絡——數位檔案之問題與挑戰〉，項潔主編：《數位人文要義：尋找類型與軌跡》，臺北：臺灣大學出版中心，頁56。

⁸ 同註7，頁41。

⁹ 同註6，頁3。

¹⁰ 陳詩沛、涂豐恩、何浩洋(2013)。<〈數位人文趨勢與「21世紀的史家技藝」工作坊〉，《國史研究通訊》4期，頁171。

¹¹ 項潔、翁稷安(2011)。<〈數位人文和歷史研究〉，項潔主編：《數位人文在歷史學研究的應用》，臺北：臺灣大學出版中心，頁13。



料檔案中採擷相關所需資料，而對多元、豐富的資訊流進行保藏與儲存，同時更奠基於此，得以克服時間與空間的局限，發掘史料背後的隱含意義及其脈絡軌跡，進而透過多重向度對歷史發展進行分析與詮釋¹²。再者，藉由數位人文學方法，除可以重複驗證統計數據，加深前人研究成果的正當性外，亦能補充傳統研究法之不足。同時，透過資訊科技的應用，可使研究者能迅速地勾勒出觀念發展圖像，並藉圖像的觀察而挖掘出「歷史殊異點」，而使得研究者能進入研究的核心焦點，進而在此基礎上進行歷史解釋¹³。對思想史研究而言，與過去歷史研究的計量方法相較之下，數位人文有一個重大的突破：全文的檢索與驗證，這在傳統思想史的內容分析上，幾乎是一個不可能的任務。

筆者在 1990 年代中期參與中研院「臺灣研究網路化」的主題計畫兩期，開始接觸「數位典藏」計畫正式展開前的資料庫建構與研究工作。同一時間，則以傳統的歷史研究取徑，配合簡單的計量分析，進行 1950 年代臺灣史上自由主義代表刊物《自由中國》的研究工作。而於 1996 年出版專書《〈自由中國〉與民主憲政》¹⁴，同時在此前後持續進行《自由中國》與其主要負責人雷震的相關研究。主要有〈臺灣自由主義思想發展的歷史考察（1949-60）：以反對黨問題為中心〉¹⁵、〈《自由中國》「反對黨」主張的歷史考察〉¹⁶、〈雷震的國家「統治機構」改革主張——對臺灣自由主義的一個考察〉及〈戰後臺灣自由主義與民族主義互動的一個考察——以雷震及《自由中國》的國家定位為中心〉¹⁷等。同時，在《自由中國與民主憲政》的基礎上，編輯了工具書《自由中國全 23 卷總目錄暨索引》一書。

¹² Peter K. Bol, "Historical Research in a Digital Environment", 《第三屆數位典藏與數位人文研討會論文集》，2011 年 12 月 1-2 日，國立臺灣大學，頁 1。

¹³ 金觀濤、邱偉雲、劉昭麟，〈「共現」詞頻分析及其運用——以「華人」觀念起源為例〉，《第三屆數位典藏與數位人文研討會論文集》，2011 年 12 月 1-2 日，國立臺灣大學，頁 214。

¹⁴ 薛化元（1996）。《〈自由中國〉與民主憲政——1950 年代臺灣思想史的一個考察》，臺北：稻鄉出版社。

¹⁵ 薛化元（1996）。〈臺灣自由主義思想發展的歷史考察（1949-60）：以反對黨問題為中心〉，《思與言》，34 卷 3 期，頁 241-286。

¹⁶ 薛化元（1995）。〈《自由中國》「反對黨」主張的歷史考察〉，《臺灣風物》，45 卷 4 期，頁 9-45。

¹⁷ 薛化元（1995）。〈戰後臺灣自由主義與民族主義互動的一個考察——以雷震及《自由中國》的國家定位為中心〉，《當代》141 期，頁 32-45。

在此基礎之上，當 2013 年國科會（今科技部）推動數位人文計畫時，筆者和本文的共同作者參與「文本史料數位實驗室之打造與創新人文研究之探索」整合型計畫，擔任計畫共同主持人（總主持人劉吉軒教授），並主持子計畫「數位人文脈絡下的史學應用研究——《自由中國》內涵之檢證與詮釋」。總計畫下另有兩個子計畫：「文本史料資訊檢索與探勘工具之開發與實踐」（主持人蔡銘峰助理教授）、「文本史料社群網絡分析工具之開發與實用」（主持人劉吉軒教授）。這兩個子計畫，主要負責進行資訊分析工具的研發，而筆者則利用歷史訓練的背景，以人文學者角度出發，透過質性視角對資訊工具所分析的量化結果，進行人文的檢視與研判，並將結果回饋給資訊專家，經過此一連續不斷的雙向修正歷程，促使結果指標更具準確性與詮釋力。

而在執行上，則透過人工勘誤與校正，先建備《自由中國》全文內容的數據資料庫，並以此為統計樣體，作為計畫後續數位人文研究的基礎。其次，則應用作者／時間／文章三個變數，對《自由中國》進行統計並觀察整體時間的作者文章發表排序、個別年間的作者文章發表排序、個別作者在不同年間的文章發表排序等不同情況，並分析其發展趨勢。三者，藉由資訊工具的自動篩選計算特性，分別求算出（1）扣除掉「文藝」類的《自由中國》文章分類，以及（2）《自由中國》的社論文章分類。透過資訊工具計算結果，再與過去研究的人工判讀結果，進行比對與內容分析，並進一步呈現二者在不同年度各分類文章的分布趨勢。此舉除補足過往由人力處理資訊的局限，更通過自動化統計而得以通盤掌握整體文本的內容。

從文章總發表量觀察，若僅就作者本名進行統計，則排序前五分別為：雷震、龍平甫、徐訏、殷海光、蔣勻田，但在納入如筆名與社論等因素後，前五名的排序即發生變動，分別為：雷震、殷海光、夏道平、龍平甫、傅正。而從此間差異可以獲得若干資訊，如除龍平甫外，其他四名作者皆有為數頗多的文章散布在社論文章及筆名發表文章中，其中，差別顯著的如夏道平原本的排序為十四，調整後提升至第三；傅正原本排序在第十八位，調整後提升至第五位，而夏道平與傅正在社論及筆名發表文章的加總，皆多於以本名發表的數量。其中筆名和社論的執筆者，主要透過人工比對雷震、殷海光、夏道平、傅正四人的全集、文集或是選集，才能找到筆名或是執筆的社論。

而《自由中國》雖以標舉自由主義著稱，但是全文檢索的計算結果，也出乎意料之外。相對於與自由主義對立的其他主義，特別是共產主義與社會



表一 主義在《自由中國》全卷文章（排除文藝類）與社論文章（超過 100）的詞頻統計

詞彙	《自由中國》全卷文章（排除文藝類） 詞頻數	社論文章詞頻數
共產主義	2,742	170
社會主義	2,065	52
自由主義	742	54
民族主義	497	55
三民主義	364	52
民主主義	225	9
馬列主義	236	12
馬克思主義	136	11

主義各接近自由主義詞頻的 4 倍和 3 倍，自由主義詞頻明顯偏低。在某種意義上，呈現《自由中國》主要以批判對立的主義，來論述自由、民主的主張（詳見表一）。

此外，筆者在計畫展開之後，固然觀照 1990 年到 2010 年《自由中國》的相關研究成果，不過，「數位研究」主要的比較對象則是筆者過去對《自由中國》內容的整體性觀察與初步統計的研究成果。

筆者的團隊以蔡銘峰團隊運用電腦對《自由中國》文本進行全文自動斷詞篩選，將結果進行詞頻統計，作為基礎。再進行人工汰選，去除單字詞、形容詞等較無實質意涵的字詞等，最後，利用汰選過的自動斷詞配合人工關鍵詞，進一步篩選出有關的「核心詞語」，而後再透過「核心詞語」進行《自由中國》文本的分析。本計畫透過「核心詞語」自動篩選，《自由中國》文章主題分類（扣除「文藝」類文章）之計量與排序，與過去的研究成果對照，主要發表在〈1950 年代台湾における自由主義的な政論の主張——『自由中国』雑誌のデータ分析〉一文¹⁸，囿於篇幅，不再詳述。

而與過去筆者自己的研究成果對照，也有明顯的落差。在《〈自由中國〉與民主憲政》中，參考數量分析上常用的 Pareto principle¹⁹（巴略多法則）進行篩選，並以《自由中國》所涉及臺灣的 1,297 篇民主憲政主張的文章為統計母

¹⁸ 薛化元，〈1950 年代台湾における自由主義的な政論の主張——『自由中国』雑誌のデータ分析〉，「東アジアの近代化と台湾社会の変遷」研討會，關西大學，2014 年 8 月。

¹⁹ 有關 Pareto principle 的說明，參見楊必立、劉水深（主編）（1988）。《生產管理辭典》，臺北：華泰書局，頁 229-230。

表二 《〈自由中國〉與民主憲政》有關七項文章主題分類的統計結果與本計畫相對應的研究成果

(各分類篇數 / 分析總篇數)	自由民主的基本理念	表現自由 / 《出版法》問題	反對黨	地方選舉問題	教育 / 救國團	外交 / 聯合國問題 / 「中國問題」	總統三連任問題、臨時條款 / 國大修憲
《〈自由中國〉與民主憲政》	114/1,297	151/1,297	97/1,297	71/1,297	182/1,297	140/1,297	70/1,297
本計畫	1,065/3,052 ²¹	175/3,052	177/3,052	25/3,052	633/3,052	679/3,052	547/3,052

體，人工判讀統計後占總篇數 5% 以上的項目²⁰，包括：自由民主的基本理念（占總篇數 8.79%）、表現自由／《出版法》問題（占總篇數 11.64%）、反對黨（占總篇數 7.48%）、地方選舉問題（占總篇數 5.47%）、教育／救國團（占總篇數 14.03%）、外交／聯合國問題／「中國問題」（占總篇數 10.79%）、總統三連任問題、臨時條款／國大修憲（占總篇數 5.40%）等七項。而本計畫則是扣除藝文類的文章，以 3,052 篇作為母體，但是仍可看出電腦全文檢索與人工判讀之間的明顯差異（詳見表二）。

本計畫由於是採用電腦搜尋計算，因此，只要文章中提及相關「核心詞語」，便可以納入計算，是以篇數較人工判讀時多出許多。其中，比較特殊的是地方選舉的部分，由於討論地方選舉未必使用「地方選舉」這個詞彙，因此，電腦搜尋的篇數較人工判讀少。針對社論文章進行分析時也有類似的情況。

根據目前的研究經驗，檢討前兩年的研究，參考既有的研究成果，有幾個問題值得再思考。首先，可能數據過於單薄而不具備代表性，以及相關數據在紀錄時的界定與標準不甚明確，若僅依數據資訊作詮釋，將可能忽視當時的整體社會與政治結構²²。其次，過度依賴關鍵詞所取得之訊息，影響對文本間關聯性與底蘊的掌握與分析²³。特別是，過度仰賴計量，存在相對忽視傳

²⁰ 除此七主要子項外，其他的各子項統計結果參見薛化元（1996），《〈自由中國〉與民主憲政——1950 年代臺灣思想史的一個考察》，頁 412-415。

²¹ 3,052 篇的數值，是以《自由中國》的總篇數 3,530 篇扣除藝文主題相關文章 478 篇後所得出。

²² 同註 11，頁 15。

²³ 同註 11，頁 17。



統歷史研究著重的歷史時空脈絡及其解釋的問題。因此，在應用數位人文途徑進行歷史研究時，資訊工具主要在於保存與處理資料，而整體研究必須注意原本歷史研究的人文判斷與思維投入²⁴，使數位人文的研究，在有效應用數位工具之時，避免忽視原本歷史研究的人文特色，才能有效整合，發揮更大的研究效益。

²⁴ 同註 4，頁 13。