

搭起數位與人文的橋

祝平次*

「數位人文研究人才培育計畫」(MOST 103-2420-H-007-005-MY3) 是「數位人文推動計畫」其中的一個子計畫；總計畫由臺大數位人文研究中心項潔教授主持，另一子計畫則為中央研究院歷史語言研究所林富士教授主持的「數位人文學白皮書研討計畫」。培育人才子計畫（以下簡稱本計畫），從 2014 年 3 月 1 日到 2017 年 6 月 30 日執行完畢，一共三年四個月的時間。是筆者從事的第一個推動計畫。比較起來，推動計畫和一般的研究計畫，不管是在過程、還是成果，有著很不一樣的經驗和要求。一般的研究計畫，著重在研究本身，成果往往是發表的論文。而推動計畫過程和較為靜態的實驗室研究、電腦前或圖書館裡檢索閱讀詮解資料非常不同，主要是在各地舉辦工作坊，在過程中與參與者接觸、互動，了解參與者的動機與下一次活動可能改進的地方。計畫的成果則是推動的成效，而本計畫的推動目標則在於使人文社會學科領域的學者能更加熟悉將數位工具應用到自己的研究或教學上的方法。計畫期間，總共主辦與協辦了 45 場的活動，參與的人次共有 1,300 多人次，邀請了二十多位學者擔任講師或講員。

工作坊的目標，在於讓參與的學員在一天或半天的工作坊結束後，對於所介紹的資料庫或數位工具能夠操作或使用，所以大部分都會安排實作。在三年多的時間裡，本計畫和不同的單位合作，舉辦了不同類型的工作坊。這些類型包括資料庫、文本分析、文本標記、地理資訊系統、社會網絡分析。雖然，這些處理資料的方式對於某些學術領域可能是慣常使用的方法，在人文學科裡卻仍屬稀少。這些方法的使用，使得一位人文學者能夠在巨觀資料的整體掌握與微觀的閱讀詮釋中，更自由快速的移動聚焦，以便在更堅實的資料基礎上，進行詮釋。而這些方法既然與其它學科使用的方法一樣，也使得人文學科更容易與其它學科溝通與在知識的範式上相互刺激。

* 國立清華大學中國文學系副教授

就資料庫而言，本計畫介紹了幾種線上資料庫的建置與「中國歷代人物傳記資料庫」(China Biographical Database, 簡稱 CBDB)。

臺灣大學數位人文研究中心建置了許多線上的資料庫，這些資料庫都呈現在臺灣歷史數位圖書館(Taiwan History Digital Library, THDL)裡。¹經由項潔老師團隊在工作坊的介紹，學員了解線上資料庫的建置過程與各種功能。其中，有不少是團隊為政府機構所建置的查詢系統，因而透過這些工作坊的介紹，參與的學員也能夠讓自己更具備理解政府機構所收藏文獻的情況，如臺灣省議會史料總庫、中華民國地方議會議事錄總庫、國史館檔案史料文物查詢系統等資料庫。

CBDB 則是目前聞名全球漢學界的一個資料庫，由美國哈佛大學費正清中國研究中心、臺灣中央研究院歷史語言研究所、及北京大學中國古代史研究中心三個單位合作開發建置。²目前收有 35 萬多筆人物基本資料，大多數人物來自於唐到清五個朝代。除了人物資料，也記錄了與這些人物相關的仕宦資料、籍貫、親屬關係與社會關係。有了這麼多龐大的資料，就可以向資料庫提問集體傳記學的問題，例如以資料庫的資料為合理樣本，在資料庫查詢統計「宋代進士平均中舉年齡為幾歲」、「明代進士的平均年齡為幾歲」。³也可以配合下面要介紹的「社會網絡分析」、「地理資訊系統」，向資料庫提取北宋時代王安石的政治網絡，以及他的政敵司馬光的政治網絡，觀察兩個人的政治網絡型態、組成分子有何異同；或是探勘宋代有名十世為官的呂祖謙家族在地理上如何從北宋遷移到南宋。這些查詢都可以配合其它軟體，做成統計圖表、社會關係圖，以及地圖，不但可以快速擷取匯聚整理資料，還可以從比較直觀的視覺呈現來觀察資料。這種轉換資料呈現的方式，不只對於研究很有幫助，同樣也可以作為更有效表達的教學工具。臺灣師範大學的李宗翰老師、目前就讀哈佛大學的毛元亨同學都曾經擔任本計畫的講師，教學員實際操作。

地理資訊系統(Geographical Information System, 簡稱 GIS)是地理學家、景觀學家、城市規劃專家乃至於軍事專家研究地形、地物不可或缺的工具，過去由於地圖資料、軟體系統昂貴、技術門嵌高，所以一直無法普及；但現在由

¹ 臺灣大學數位人文研究中心，<http://www.digital.ntu.edu.tw/> (2017-12-01)。

² Harvard University, Academia Sinica, and Peking University, China Biographical Database (December 1, 2017), <https://projects.iq.harvard.edu/cbdb>.

³ 可以參考 CBDB 的使用手冊：傅君勳 (Michael Fuller) (2017)，〈中國歷代人物傳記資料庫用戶指南〉[EB/OL]. [2017-12-01]. <https://projects.iq.harvard.edu/cbdb>.

於開放軟體、開放資料、以及便利的使用者介面，使得一般使用者也都有能力使用 GIS 來進行研究。⁴ 時空兩大因素，是文史學科無法迴避的對象，因此本計畫也和其它單位共同籌辦了好幾場的 GIS 工作坊。彰化師範大學的李宗信、東華大學郭俊麟兩位老師帶領學員實作了和商業軟體可以匹敵的 QGIS (Quantum GIS)。臺灣大學鄧志松老師則帶大家實作他自己利用微軟試算表軟體 Excel 寫的巨集程式 Excel2Earth：一方面利用 Excel 的強大計算能力來進行一些地理計算，一方面則利用流行面較廣的 Google Earth 來進行地理繪圖。⁵ 2016 年鄧老師在美國大學訪問、筆者在歐洲萊頓大學也都介紹了 Excel2Earth 給國外的學者。中央研究院 GIS 中心張智傑、廖汝銘兩位老師則推介他們開發的 SinicaView，還有和日本學者一起合作的 HuTime。⁶ SinicaView 是和 Google Earth 類似的 GIS 平臺，但 SinicaView 則還加上了 GIS 中心所製作的圖資，所以很容易就可以使用各式各樣的歷史地圖，非常方便。由於學者和政府一起合作，這十幾年來，不管是原住民族傳統領域的釐測、社區重建再造、環境看守，GIS 都成了不同社群看顧自己團體的利器。如果再加上時間的向度，就更容易聯繫起過去與未來的空間關係。

社會網絡分析 (Social Network Analysis, 簡稱 SNA) 是從社會學發展出來，認為人的社會網絡關係對社會中的人影響甚大，所以以網絡關係為研究主題。其中，有名的小世界理論，即認為透過六個人，即可以將全世界的人連在同一個關係網絡中。目前在現實的應用上，可以利用來分析學童彼此間的關係，以便了解被霸凌學童的偵測；企業組織的弱點。目前在網路上有各式各樣的分享軟體可以將網絡關係視覺化加以觀察。政治大學熊瑞梅老師是臺灣社會網絡研究的資深學者、同系鄭力軒老師、龍華大學林芳苓老師、中山大學葉高華、臺灣大學鄧志松都為本計畫開辦工作坊，分別使用了下列的工具：Pajek，UCINET 和 NodeXL。⁷ 就文史學科的應用方面，最常被利用來分析歷史人物彼此之間的

⁴ 改寫自李宗信上課簡報檔。

⁵ 鄧志松，Excel2Earth (2017-12-01) <http://excel2earth.ccs.ntu.edu.tw/spatial/download.html>

⁶ 中央研究院地理資訊科學研究專題中心，SinicaView (2017-12-01)，<http://3dgis.rchss.sinica.edu.tw/>。Tatsuki Sekino, HuTime (2017-12-01), <http://www.hutime.org/>。

⁷ Pajek 可以免費自網路上下載，運算速度快，適合進行大型網路的繪製：Pajek (2017-12-01) <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>。UCINET 使用 60 天後需付費，是歷史悠久的一款軟體：UCINET (2017-12-01), <https://sites.google.com/site/ucinetsoftware/downloads>。NodeXL 是配合微軟公司 Excel 的附加範本，可以很快的幫使用者建立網路資料，但進階版需付費：NodeXL (2017-12-01), <https://nodexl.codeplex.com/>。

關聯以及個別在政治社會的影響力、小說裡的角色關係，以及概念和概念之間親和性的關係。

最後，人文學科當然脫離不了文本分析。政治大學的劉昭麟老師與學員分享他所開發的 TaiwanDH 程式。⁸ 程式雖然小，卻能很快地進行關鍵詞的脈絡檢索、找出共現詞、以及計算詞頻，並且能夠進行不同層次的累積計算，例如，要找的詞組分別在一些書與這些書裡的章節出現以及分布的情況。這樣的功能，使得程式使用者可以很快地對大量的文本有一整體性的掌握，很能滿足人文學者的需求。臺灣大學闕河嘉老師則與學員分享她的團隊開發的庫博語料分析軟體 (Corpro)。⁹ 這個軟體能夠自動將中文斷詞再進行分析。也設計成能夠編輯辭典、停用詞以及同類詞清單以便進行不同層次的累計。對於分析當代漢語的語料，庫博功能強大。雖然網路上也有類似的英語軟體，但庫博可以斷詞，介面又是中文，所以更適合國內的使用者。

提到文本分析，當然就不能不提 TEI (Text Encoding Initiative) 標記標準。¹⁰ TEI 是已經有 30 年的國際組織，旨在發展人文學數位資料的標記標準。有了這樣的標準，標記後的資料就可以被不同的群體、軟硬體使用，而不會有相容性的問題。法鼓山聞名世界的中華電子佛典 (CBETA) 就是利用這套標準進行標記。¹¹ 本計畫也請到十年來主持該計畫的洪振洲老師帶領學員實作 TEI，以體會實際的標記工作怎麼進行、以及標記好的文本有什麼樣的功能。TEI 也符合 XML (eXtensible Markup Language) 的規範，因此標記好的語言，也可以被視為是一種半結構資料庫。因而，TEI 對於人文學者來講特別重要，因為它可以把原始文獻藉由標記的手段而變成資料庫。但手動標記是一項非常煩細的工作，也容易出錯。雖然，社會學科常用它來整理訪談資料，但若要大量批次處理就是個問題。因而，本計畫也趁萊頓大學魏希德 (Hilde De Weerd) 教授受邀在臺進行研究時舉辦工作坊，引介實作她的團隊所開發的線上標記平臺 MARKUS。¹²

⁸ 劉昭麟，TaiwanDH (2017-12-01)，<https://sites.google.com/site/taiwandigitalhumanities/>。

⁹ 闕河嘉、陳光華，「Corpro 庫博中文語料庫—分析工具」(2017-12-01)，<http://dh.lis.ntu.edu.tw/resource.html>。

¹⁰ <http://www.tei-c.org/index.xml>。

¹¹ 中華電子佛典協會，CBETA(2017-12-01)，<http://www.cbeta.org/>。

¹² Ho, Hou Ieong Brent & Hilde De Weerd. MARKUS. Text Analysis and Reading Platform. 2014- <http://dh.chinese-empires.eu/beta/> (2017-12-01) Funded by the European Research Council and the Digging into Data Challenge.

經由她這幾年的推介，MARKUS 現在也是國際漢學界知名的工具，未來也將整合不同的語言，使得它的功能更為提升。

綜上所述，計畫安排的工作坊包括了各式各樣的研究工具，這也是數位人文研究必須要有的認識：亦即沒有單一工具就能幫我們完成所有的工作。如何對於不同的工具有不同的認識，並且能因為研究標的需要，能夠結合數種工具來進行研究，是數位人文精神之所在。畢竟人文學的精神就在於從不同的面向對於人的世界進行理解，而人的世界原本就是多樣複雜的。這也和目前注重實作與模組學習的教育理念相符合。

由於籌備工作坊需要相當的時間，所以通常在一個學年裡我們只選定三到四種工具，由特定的講師在北中南東輪流進行，以節省安排工作坊的溝通成本。2017 年 3 月也在金門舉辦過兩天的工作坊。在三年多的時間裡，參加的學員包括大學部同學、博碩士班同學、現任的教師、退休的教師，也有業界的朋友來參加。南部的廖平喜老師、北部故宮張志光先生、師範大學邵軒磊老師更是從陌生到熟識，邵老師現在更全力在自己的領域推進數位人文研究，自學多種分析工具，更是令人佩服，可以說是計畫成功的案例之一。

在每次的工作坊，我們都會提醒學員，與其將該工作坊視為僅有的學習機會，不如把它視為是一個引發興趣的契機。數位人文的發展就像很多線上的學習群組一樣，可以沒有前提地加入，互相幫助協作。在工作坊進行時，我們也要求每個學員與鄰座的學員互相幫忙一起保持學習的進度。雖然，每次實作課程都會安排助教，但因為人數的關係，不一定都能夠照應到每個需要幫忙的學員，依靠著學員彼此之間的幫忙，能大大增進學習的效果，也比較能夠保持講師的進度。另外，為了增進學習效果，本計畫總共錄製了四種螢幕教學：CBDB、Excel2Earth、GIS、SNA。這四種錄影帶可在 YouTube 尋得，提供學員在課後複習使用，也讓沒有時間來參加工作坊的人可以在家自學。¹³

在執行本計畫這三年多的時間裡，由總計畫推動的臺灣數位人文學會於 2016 年 4 月成立，科技部的學術研習營也於 2016 年 11 月設置了跨領域的數位人文課程，教育部則在一年左右的先導計畫後也即將於 2018 年開始推動四年期的數位人文創新培育課程。雖然本計畫成果牽涉到對於人的能力的評估，所以很難量化，但的確在這個過程裡，我們看到了部分參與的學員也開始讓數位人

¹³ 可以在下列網址找到與本計畫相關的影片：

<https://www.youtube.com/channel/UChyyjTUusnyryNDVUuOB6zQ>。

文的方法進入她們的研究，並且能在臺灣每年一度的數位人文國際會議發表論文，雖然數量不多，但卻是在這個萌芽階段的重要種子。希望將來能夠憑藉這些種子，使得數位人文可以成為人文學界熟悉易用而普遍的一種研究方法。