

數位資源對於 歷史學研究的 意義： 一個數位人文 研究者的觀察

項潔 國立臺灣大學數位人文研究中心主任及資訊工程學系
特聘教授、中央研究院資訊研究所研究員

一、歷史學家想要的是什麼？

我是一個資訊工程學者，不過這些年來卻和歷史學家有許多互動。其淵源要追溯到近二十年前，當時我的研究興趣發生了很大的轉變，從抽象的數理邏輯和自動推論領域轉入了數位化研究的新天地，亦開啟了與歷史學及圖書館界學者的合作之旅。其後，從臺大數位圖書館計畫開始，以文建會（現文化部）的國家文化資料庫和國科會（現科技部）的數位典藏與學習國家型計畫為拓展，再以臺大數位人文研究中心為平台，我和人文領域學者及機構的合作亦日漸增多。不過，要論關係最為密切的，無疑還是歷史學領域，規模龐大的臺灣研究史料總匯——臺灣歷史數位圖書館（THDL），（註1）便是一個例證。

在與歷史學家的接觸中，我頭腦中始終縈繞着一個問題，那就是「歷史學家到底如何想問題？」這個問題與其說是出於對其思想世界的人類學式的興趣，毋寧說是出於功利性和實用性的目的，更直接的說，我想要了解的，就是「歷史學家想要從數位資源中得到的是什麼？」因為我想要建置的是符合歷史研究者需求的數位資料庫，如何克服將龐大蕪雜的歷史資料數位化的種種困難，無疑首先是我們要面對和解決的，但接下來要面對的更大問題，是歷史學家如何能夠有效地運用它來做研究。

經過反覆接觸觀察，我發現多數時候，

歷史學家其實是懷着明確的問題意識走進資料庫的，他們很清楚要在其中捕捉什麼。也就是說，他們的問題，還是「歷史學」的問題，然而我想要知道的，是他們的研究工作與數位資料庫這個新史料載體之間到底有何關聯。有位歷史學家的回答讓我眼前一亮，他說：「我想要的是一個積極的研究助理。」在他看來，這個數位資料庫不應只是被動應對檢索的關鍵詞，它還應該積極地揣測他的需要，並提供相關的文本或分析，以便對於其研究有所助益。

這個答案之所以打動我，不止是因為他巧妙運用修辭將數位資源的功能凸顯出來，更重要的是與我多年來從事數位人文的追求不謀而合。多年來我們試圖建置的研究者資料庫，就是不想要再拘泥於傳統數位化資料庫常見的單一關鍵詞檢索，而是要跳出窠臼，透過數位化資料庫優越於紙本的靈活性和穿透力，為研究者提供多種觀察、比較、分析、類比文本的視野，以達到更為強大的、輔助研究工作的功能。

那麼，數位資源發展至今，它是否能夠以及如何扮演「積極的研究助理」的角色？它又能否超越研究助理的階段，成為創造新問題和解決新問題的途徑？在這篇簡單的文章中，我便試圖從一個數位人文研究者的角度，提出自己的一些初步思考和構想，以期與歷史學界的諸位方家進行討論。

二、數位資源可以做什麼？

我所談的數位資料庫雖然內容以人文研究所需要的史料為主，但是隱藏在資料庫內部的資料組織和運算邏輯，無疑和人文研究是相距十萬八千里的。在我主持的臺大數位人文中心，資訊科學背景的工程師和歷史學背景的研究人員之間，便有一場「曠日持久」的爭論。曾經建置了 THDL 系統的資訊科學研究員，試圖再重新設計一個以歷史學者個人研究興趣為取向的資料庫，在其中，歷史學家可以上載各種不同類型的史料，並運用時間、人名、地名等元素，進行史料之間的交叉比較，並進行相應的統計分析。然而，歷史學者對這一做法似乎尚不滿意，他們希望資料庫能夠有辦法找到彼此之間「真正」存在關聯的史料，而不是割裂成諸多關聯性可能有限的板塊。

我必須承認，當資訊工程的邏輯運算遇到千變萬化的史學史料時，其局限性是顯而易見的。想要找到「有關聯」的史料，並非完全沒有辦法，譬如我們可以一些計算方法找出文件中有意義的辭彙，再找出出現率或共現率（co-occurrence）最高的辭彙或文件。但是一方面，它所呈現的關聯未必符合歷史學者期待的「真正存在的關聯」的要求；另一方面，歷史學家自己也承認，「各種史料都有一套自己的慣用語。如果歷史學家注意它們，就得記下那個時代的常用語，它們的每一次出現都有不同的含義。」而使

事情變得更複雜的是，「事物發生了變化，其名稱卻未必發生相應的改變。因為所有的語言都具有因襲傳統的特性，而且大多數人是缺乏創造性的。」（註2）在這種情況下，考察詞彙本身變化的歷史過程，已經是歷史學的研究課題，試圖讓根據數學原則運算的數位資料庫達到這一境界，近乎一種苛求。

同時，資料庫最為擅長的統計運算和分析，亦是一把雙刃劍。它的好處有目共睹，對於工程師和歷史學者而言，統計都是最容易運用的方法，可以直觀地比較不同要素之間的對比關係。然而，統計在使用上有一個先決條件，就是資料要有足夠的代表性。這個要求通常是指資料要不就足夠全面、要不就反應所有的面相（也就是統計裡所謂樣本的隨機性）。對於史料而言，兩者任一被滿足的機會都微乎其微，若太過依靠統計結果進行分析和解釋，便有可能陷入另外一種倒果為因的境地。更何況歷史學者都很明白，最常出現的詞彙並不代表是最有意義的詞彙，許多思想的湧現、延續和斷裂，中間有更為複雜的歷史過程和背景需要分析。

再進一步而言，數位資料庫背後的思維邏輯，與歷史學家的思維邏輯之間亦存在很大差異。就我個人的經驗而言，近二十年前我的學術轉向，便與當時臺大及臺灣學術界鼓勵跨學科合作的氛圍有關。經過多年的實踐，我感覺學科之間的整合確非易事。這並非僅僅是因為學科建置、學科規範和技術問題，也涉及到行業內部一些只可意會、不

可言傳的「個人知識」。我所受到的純數學領域學術訓練及資訊工程領域的技藝，常常是人文學者視若畏途、避之唯恐不及的；而歷史學者在探討歷史問題時，其問題意識的敏銳、論證的有力和文字敘述的美感，也常常是令我望塵莫及、望洋興嘆。其最大的差別，我認為在於數學領域裡講的是同一種抽象的語言，而在歷史學領域，誇張一點形容，每個歷史學家的寫作和思考方式都存在差異，可以說每位學者都有自己的一套語言，正如布勞岱爾（Fernand Braudel）所言的，「這是個人風格和氣質問題。」（註3）

然而，歷史學家的問題並非沒有合法性，資訊工程學者亦不能只停留在安全領域，而迴避來自歷史學家的更高要求。我相信，數位人文未來的發展亦會不斷朝向這個方向努力。問題是在目前階段，我要如何消除歷史學者的疑慮，並試圖說明目前的數位化資料庫有助於其研究工作呢？

借用德國社會學家韋伯（Max Weber）的詞彙，「除魅」（disenchantment）也許是一個重要的起點。因為人類技術的發展趨勢，歷史學者未來必須要在新的數位環境下作研究，已經是不可逆轉的趨勢。然而這並不意味着他要從頭學習資訊工程的程式語言，這既不實際，也沒有必要。我們想要打造的資料庫，不止是假定史料之間存在關聯，並且這些關聯可以被比較和分析；我們更強調的是，這個資料庫中關聯建立、比較與分析的過程和方法，應該都是可以被歷史

學者觀察和理解的，例如《明實錄》、《清實錄》和《朝鮮實錄》之間的關係，就可以透過時間的軸線建立。之所以這樣做，一方面是希望捨棄太複雜的數學公式，將用來比較文本的元素透明化，以免把歷史學家嚇跑，另外一方面則希望換位思考，理解歷史學者在閱讀史料時思維方式的運作，對其而言，時間、空間、人和事件等要素是其閱讀史料時比較的第一步，這也是為什麼我們極為注重這些元素的原因。

歷史資料之間的關聯性，當然不僅僅建立在時間、空間等這些電腦好處理的因素之上，亦可能建立在只有史家敏銳察覺到的、隱形的關聯之上。資料庫便應主動提供一些跳脫人時地等簡單關聯的相關性文件，供歷史學者進行選擇。目前市面上的數位資料庫並非完全沒有這一功能，例如中國大陸的電子期刊全文數據庫（CNKI），便運用共同引文的關聯，為查閱某一篇文章的讀者提供相關的其它文章，這便是一個相關性文件的例證。不過對於歷史學者而言，這可能還是不夠的，我們曾經在臺灣歷史數位圖書館收藏的大量臺灣古契書基礎上，建立了不同土地契約之間的上下手契關係和土地移轉圖，其方法和基礎，結合了空間（地點）、時間和買賣雙方的親屬關係等，這是一個進階的例證。（註4）

我當然明白，要量身打造符合歷史學家需求的使用者資料庫，我們還有很長的路要走。不過也要強調，數位資料庫的可能性早

已遠遠超越了關鍵詞檢索，它已經在更深的意義上，滲透並改變着歷史學家的研究環境，甚至改變其問題意識的產生方式。

三、數位資源可能做什麼？

有一次一個歷史學者便和我開玩笑，「當我在看資料庫中的資料時，頭腦就全面發動，管它到底有幾個脈絡在運作。」我明白他的玩笑中隱含着批評，因為我曾經寫過一篇文章，論述從資料庫不同的呈現方式中，可以觀察到隱藏在文本之間的數種不同脈絡，如鳥瞰的脈絡、縱觀的脈絡等。這句話背後也透露着自信，即歷史學家是問題意識的主體，為了解決其感興趣的議題，他會運用全部的史學智識和技巧，在史料之海中披荊斬棘，淘沙撿金，而不必回頭審視方法和路徑是什麼。而其對數位資料庫最常見的批評，也是它並沒有辦法幫忙解決其既有的問題。

我的問題則恰恰相反，如果我們把資料庫看作歷史史料的總匯，那麼歷史學家有無可能在一個新的基礎上，重新思索其預設問題的位置呢？再次借用布勞岱爾（Fernand Braudel）的話來說，「歷史學家第一眼看到的就是人生競爭中獲勝者的那一堆事件。但是，這些事件在一個具有多種並相互矛盾的可能性的框架裡再次安置和排列自己，生活最終在那些可能性中做出抉擇。一種可能性實現了，十種、百種、千種可能性就消失

了，甚至還有無數中可能性。在我們看來並不是謙遜、隱蔽的不至於一下子對歷史學施加影響。」（註5）

雖然今天的歷史學家已經不像 19 世紀末 20 世紀初的學者一樣，信奉史料和追求客觀真實的歷史，他們更願意相信歷史是今天賦予過去意義的方式，甚至是對人類命運悲喜劇文學化敘述的方式，但是如何在史料最大化的基礎上對歷史問題進行思索，一向是歷史學家孜孜以求的。而一旦數據庫的規模夠龐大，便能夠在更堅固的基礎上，接近已經消逝的歷史時空樣態。因此運用數位化資料庫的意義之一，便是消解單一的獲勝者歷史，而把各種可能的力量，帶回其衍生、發展和消亡的歷史時空，從而明白我們如何走到今天。

更進一步而言，歷史學家是否可以摒棄預設，從資料庫中發現全新的文本脈絡，並衍生出全新的問題意識呢？這是因為資料庫資源的擴展和整合功能，以及其靈活排列組合、重新對比分析的特性，讓學者早已超越了在汗牛充棟的紙本史料中搜索的階段，而思考如何以新的眼光看待和處理史料的問題。目前我們在進行的一個計畫，便是對中國古代的大型「類書」，如《太平御覽》、《古今圖書集成》等進行對比研究，觀察古人知識和思維世界的整體變化。這是因為動輒幾百萬字的類書，古人往往作為工具書使用，前代歷史學者也多根據感興趣的問題再進行查閱，而很少從整體上進行比較

分析，借助於數位資料庫的功能，這一工作才第一次有可能進行。

最後，我想回到歷史學家的觀點，與我的觀察做一互見和比較。八十餘年前，歷史語言研究所所長傅斯年，曾經提出了對後世影響很大的說法，那就是學者要借用新工具、擴張新材料、發掘新問題，這也是歷史語言研究所的旨趣所在。（註6）在我看來，數位資料庫除了作為新工具的功能之外，亦是一種新方法和新問題衍生方式，歷史學家可以從中發現迥異於傳統史學、或傳統史學無法解決的新問題，在這個意義上，它便超越了僅僅作為歷史研究中一個工具的作用。

歷史學家未必贊同我的觀點，但是同樣深刻感受到新環境之下研究工作的轉變。在傅斯年之文發表了四分之三個世紀後，歷史學家王汎森重新思索了史語所的研究旨趣，便指出在當今時代，史料和史家的關係早已經改變，尤其是電子文獻的迅猛發展，史家面臨了全新的環境和挑戰。他透過自己的思想史研究也發現，「在追溯一個重大思想觀念的變化時，電子資料庫方便我們在極短的時間內得到一個長遠的、縱深的輪廓，也可以對共時性的歷史現象得到廣泛的理解，因而可以進行更有把握的推論，同時也使得另一種層次的歷史解釋與歷史意義的追索成為可能。」（註7）

我想要指出的是，在重新定位學者和史料的關係上，資訊工程學者並不是完全隱形和透明的。或許在不久的將來，新問題透過

新工具衍生的時代，數位化資料庫扮演的將不僅僅是一個「積極的研究助理」，而更是歷史學家進行研究時不可或缺的合作夥伴角色。

【註釋】

1. <http://thdl.ntu.edu.tw>
2. 馬克·布洛赫 (Marc Bloch) 著，張和聲、程鬱譯，《歷史學家的技藝》(上海：上海社會科學院，1992年)，頁115。
3. 費爾南·布勞岱爾 (Fernand Braudel) 著，劉北成譯，《論歷史》(臺北：五南圖書公司，1988年)。
4. S.P. Chen, Y.M. Huang, J. Hsiang, H.C. Tu, H.I. Ho, "Discovering Land Transaction Relations from Land Deeds of Taiwan", *Literary and Linguistic Computing: Oxford Journals* 28 (2), pp 257-270, 2013.
5. 費爾南·布勞岱爾 (Fernand Braudel) 著，劉北成譯，《論歷史》。
6. 傅斯年，〈歷史語言研究所工作之旨趣〉，《中央研究院歷史語言研究所集刊》，第一本第一分，1928年。
7. 王汎森，〈歷史研究的新視野：重讀「歷史語言研究所工作之旨趣」〉，《古今論衡》，第11期(2004年)，頁7。