

# 「數位人文學」發展概述

翁稷安

## 一、前言：台灣數位人文的起點

「數位人文學」，顧名思義，即是希望能將資訊工程領域的相關數位技術，運用到以文史為主的傳統人文研究中，兩者結合，發揮相輔相成的效果，提供人文學研究新的視野和發現。然而，這樣常識的理解，並無法完全含括數位人文試圖作為一獨立領域所面臨的挑戰和難處。

台灣數位人文的發展，通常會回溯到上世紀以國家之力推動的數位典藏，首先是1980年代所建立的大型資料庫，如中國圖書館學會與國立中央圖書館推動的「全國圖書館自動化作業系統計畫」和中央研究院歷史語言研究所的「史籍自動化計畫」，以此兩個計畫為起點，始有之後一系列將文本、物質等各式典藏進行數位化的開展，於21世紀開花結果。先是由當時行政院國家科學委員會於2001年啟動「數位典藏國家型科技計畫」、2002年開展「數位學習國家型科技計畫」，並且在2008年整併為「數位典藏與數位學習國家型科技計畫」，在前後計畫總主持人中研院劉翠溶院士、王汎森院士的推動下，串聯了上百所不同單位，建立了數百萬筆的典藏資料。

如同王汎森在訪談中所提及的，該計畫替「人文計算時代」奠定了基礎，建立了以「關鍵字搜尋」快速檢索典籍的資料收集方式，大大減輕人文研究者在資料求索上的時

間和精神成本，可以更全面、不用擔心遺漏地對待所收集的資料，改變了人文史料的屬性，以及研究者和資料之間的關係。然而，他也意識到這樣快速所帶來的風險，他指出「由資訊化成知識、再昇華為智慧，宛如一趟必須經由個人親自體驗方能冷暖自知的旅行。」甚至可以說「困惑、追尋與挫折，是非常重要的資產。」對人文學者而言，「眾裡尋他千百度」的過程是必要的。數位典藏所帶來的便利，是否壓縮了這困惑與追尋的旅程，仍值得有志者省思和警醒。<sup>1</sup>

之所以不厭其詳，回顧這段新舊時期交接的數位典藏發展始末，是希望點出在台灣數位人文學的基本性格中，最為根本的基因。以典藏為出發，以資料的搜尋和運用為思考核心，並對於數位技術所帶來的便利保持警惕，無論日後數位人文學如何蓬勃發展，百花齊放，這最根本的特性和所帶來的正反影響，始終留存在台灣的數位人文學之中，也是投入這個領域必需要正視和反省的課題。

## 二、數位人文學的出現與定義

以大量的典藏為出發，在被動的典藏和保存之外，研究者是否能更進一步地主動運

1 〈當人文遇見科技·數位典藏與數位學習國家型科技計畫王汎森總主持人專訪〉，「數位典藏與學習電子報」10卷1期（來源：<https://reurl.cc/yeO1ra>）。

用數位科技，結合人文學的問題意識，對保存的資料進行觀察、探勘、分析和呈現，就成為數位人文學作為學門的願景。需要特別強調的是，數位和人文不應是主從的隸屬關係，而是相輔相成、互相進行的有機合作。如台灣大學數位人文研究中心所定義的：數位人文「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。」<sup>2</sup>這當然是作為學門創設的最高標準和理想，希望替當時仍屬草創期的新興領域，預留各式各樣的可能。

隨著這十幾年下來的發展，這個學門有著更明確的學科邊界（discipline）和應用範式，由中央研究院數位文化中心所出版的《「數位人文學」白皮書》即是試圖呈現該領域作為一門獨立學科應有的面貌。在林富士所撰寫的〈「數位人文學」概論〉中就指出，在大數據（巨量資料）（Big Data）、數位匯流（Digital Convergence）、虛擬世界、多語情境、以網路為連結的全球化的前提下，外國學界於2004年催生了數位人文學，並逐漸產生出相對應的成果和學術機構。在定義

上「主要旨趣是利用數位科技探索或解決人文領域的問題，或是從人文的角度探索、反思數位科技如何形塑人文世界。」實務上則是應用數位資源（digital resources），包括「數位化材料」（digitized material）即是指將原本非數位化的物件轉為數位格式的資料；「原生數位材料」（born-digital material）則是指直接利用數位工具所產生的資料，進行工具的開發、資料的觀察，以及研究結果的呈現。<sup>3</sup>

### 三、數位人文學的發展動向

近年來台灣以數位人文學為名的研究或資料庫設置，如雨後春筍般出現，顯示著數位和人文的結合已成為不容忽視的主流，台灣數位人文學會於2016年成立，2018年發行《數位典藏與數位人文》期刊。由台灣大學於2009年所發起的「數位典藏與數位人文國際研討會」，也正式邁入第11屆，並由投入數位人文研究的不同機構輪流主辦，在在說明了數位人文學的成熟。整體發展的動向，不出前述的兩種定義：一是尋找跨領域結合的新火花與新可能，另一則是在數位資源的應用上，試圖開拓出新議題與新成果。可以將大致的發展情形，歸結為下列三個主要方面，分述如下：

#### （一）機構典藏或個人研究的資料庫開發

前已述及，台灣數位人文是以機構典藏作為出發，資料庫的設立扮演著關鍵的角色，資料庫的設立絕不只是數位化後逕行上網而已，而是利用數位工具提供使用者在資料庫集合的史料中，進行「多重脈絡」的探

2 項潔、涂豐恩，〈導論——什麼是數位人文〉，項潔編，《從保存到創造：開啟數位人文研究》（台北：台灣大學出版中心，2011.11），頁11。

3 林富士，〈「數位人文學」概論〉，林富士主編，《「數位人文學」白皮書》（台北：中央研究院數位文化中心，2017.12），頁2-22。

勘，台灣大學研究構置的「台灣歷史數位圖書館（THDL）」，可以說是最好的例子。<sup>4</sup> 這樣機構資料庫的應用，也催化了許多個人式數位平台的誕生，可以想像成將原本資料庫中的檔案移去，由使用者自行上傳資料，應用工具中所附的功能，如標記、統計、呈現等，自建個人的資料庫。目前較熱門者有台灣大學建置的「DocuSky 數位人文學術研究平台」以及「數位人文研究平台」，不少學者以這兩大平台提供的功能為基礎，陸續交出許多研究成果。

然而不可諱言，數位人文學最大的強項還是在資料的「量」，越多越廣的資料集合，越能夠有「概念上跳躍」的可能，提供宏觀的觀察。再加上相較於個人，機構還是生產「數位化材料」的重鎮，並擁有在數位技術上更多的資源調度。許多單位如國立台灣文學館近年仍積極投入資料庫的建置，又或者對既有資料庫進行橫向串聯，如文化部推動的「國家文化記憶庫」。最終，機構或個人不同規模的資料庫開發，可以視為殊途同歸、不可偏廢的兩端，個人對資料庫的運用，深化了對於數位人文運用的思索；機構則不斷藉由資料的數位化、研發新的數位技術，提供整個學門長期的累積和堅實的基礎，共同構築數位人文的正向循環。

## （二）文本關聯性的資料探勘和呈現

此為數位人文研究目前在實作上的主要範式，大體而言，是藉由數位工具對一定範疇內的文字資料，對資料內部特定名詞進行關聯性的探勘和連結，然後再將結果以靜

態或動態的數位圖像加以呈現，進而給予人文學的解釋和說明。綜觀歷屆「數位典藏與數位人文國際研討會」，以及《數位典藏與數位人文》所收錄的論文，多數皆屬此類，可以說是台灣數位人文學的「標準作業程序」，已近乎某種學門的研究典範。

可以再細分成下列幾種不同類別：

1. 「關鍵字」的數量統計：顧名思義，將特定的關鍵字進行量化的統計，以數量上的增減，觀察某些觀念或物品在特定時段的流行與否，並解釋其意義。在此類別中，最核心的要素在於「關鍵字」的選擇和鑑別，「選擇」考驗著研究者處理問題的功力，「鑑別」則是考驗數位技術是否精準，能去除不必要的雜音。

2. 「關鍵字」之間的關聯：主要集中在人際網路的呈現上，藉由數位技術的搜索、統計和繪圖，將文字僅能概述的人際關係，給予量化和圖像的傳達。這部分最困難的，除了人名在不同別名、字號間權威檔的建立，更重要的，如何用明確的數量和圖示，去呈現出幽微的「關係」，而不至於簡化。尤其當關係可以無限延伸的情況下，在包山包海的連結中，截取出具解釋意義的網絡圖。

3. 跳脫文字脈絡的重現：這部分或可概括為廣義的「比對」，諸如不同文本之間內容的比對，以細究文本之間的傳抄引用。又或者將事件置於時間軸進行，除了和時間軸相對應外，也可以與同時間點的其他事件進行合觀。最後，則是目前數位人文在應用上，最受期待的——利用地理資訊系統（Geographic Information System, GIS）將文字的紀錄轉換成空間的理解和判讀，將距離和地景帶入對於文字史料的解讀之中。此類別

4 關於「多重脈絡」的討論，可參見項潔、翁櫻安，〈多重脈絡——數位檔案之問題與挑戰〉，項潔編，《數位人文要義：尋找類型與軌跡》（台北：台灣大學出版中心，2012.11），頁25-59。

的挑戰在於，是否能找到合適的比對點，以及比對之後是否能有合理的意義和解釋。

上述三者，終究還是考驗著人文學者是否有明確、清晰的問題意識，能貫徹於不同數位工具所能帶來的統計或呈現的結果，並進而做出有效的判讀，提供和傳統學界成果對話的可能。在三者已成為某種「標準」研究的同時，如何一方面加以深化，另一方面再另闢新的可能，將是數位人文學目前發展上最直接的目標。

### （三）多元的展示與應用

這部分大多數人會將之劃為機構所擁有的長項，特別是博物館等典藏機構，利用如AR、VR等不同的新科技，將既有的展品，以更沉浸式的方式展現給參觀者。隨著數位技術入門門檻不斷的降低和開放，以及「教育部數位人文社會科學教學創新計畫」的推動下，目前許多數位技術已經在大學第一線的教學現場獲得應用。前述「國家文化記憶庫」，更是鼓勵一般民眾利用所收錄資料，進行線上的策展，開拓出新的應用可能。凡此，都說明了展示和應用大有可為，不再專屬於特定規模的機構，呼應著數位時代的開放精神。但如是開放，同時也考驗著不同背景的應用者，在這些數位的「聲光效果」輔助下，所欲傳達的核心究竟為何。

## 四、結語

綜合前述，還是回歸王汎森在上世紀的觀察，「基礎建設」即更完整而開放的數位化材料的持續建置，是數位人文甚至是一般人文研究念茲在茲的根本，唯有更多不同性質的資料庫，彼此串聯，才能提供人文學者更便利的觀察。不只是資料檢索上的快

速，數位工具在統計、關聯、比對和展示之間提供的強大能力，將大幅降低研究者的時間和精神成本。然而，當數位人文技術提供了那麼多的方便，人文學者能否投注更多心力在叩問和解釋過程中的「困惑、追尋與挫折」，才是數位人文學門能否更上層樓的關鍵。