數位「考證」：人文學者的新素養

林富士（中央研究院歷史語言研究所特聘研究員）

目次

壹、引言：從「史料學派」講起
貳、「數位化資料」的陷阱
　一、不忠的「複本」
　二、選擇性的記憶
　三、魅惑的關鍵詞
參、「原生數位資料」的問題
　一、流動的文本
　二、抄襲的創作
　三、偽造的事實
　四、量化的危險
肆、是生機？還是危機？
　一、誰主浮沉？
　二、注意「眾人協作」
　三、小心機器人
　四、想想「人工智慧」
伍、結語：是素養，不是技能

摘要

本文主要是從「考證」著眼，探討數位時代人文研究的「典範轉移」（paradigm shift）問題。本文認為，所有的人文研究，無論是傳統時代還是數位時代，都需要立足於「資料」（data），也需要「考證」。但是，資料的形式和運用資料所會遭遇的問題，以及「考證」的方法和工具，則是變動不居的。就此而言，數位時代的人文研究「典範」必須也必然會「轉移」。

因此，在「資料為王」、「機器人」（robot）稱霸的數位時代，人文學者的確需要兼習數位科技與人文知識，兼具人文涵養與「運算思維」（computational thinking）。但是，這樣並不保證我們就能在數位世界騰雲駕霧、暢行無阻。畢竟，這只是素養（literacy），不是技能（skill）。而這種新素養，或許可以稱之為「數位考證」（digital textual criticism）。

壹、引言：從「史料學派」講起

1928 年，史語所的創辦人傅斯年（1896-1950）在《中央研究院歷史語言研究所集刊》的發刊詞中，針對〈史語所工作旨趣〉，提出三個「凡」和三個「宗旨」。三「凡」是指：「凡能直接研究材料，便進步」、「凡一種學問能擴張他所研究的材料，便進步」、「凡一種學問能擴充它研究時應用的工具的，則進步」；三「宗旨」是指：「保持亭林百詩的遺訓」、「擴張研究的材料」、「擴張研究的工具」。

這些主張，簡單的說，就是要「直接研究材料」、擴張研究材料與研究工具。最後，他還歸結指出：「我們不是讀書的人，我們只是上窮碧落下黃泉，動手動腳找東西」。這句話意在打破以「書本」為尊的傳統學問，而所要找的「東西」並不侷限於文字，而是含括了器物、圖像、口語、聲音的材料。

這在二十世紀初的中國，是一種「革命」性的主張，也引領了一批中國文史研究者走出一條新的道路。由於旗幟鮮明，有人便將史語所所代表的學風稱之為「史料學派」（Data School）。這個標籤有褒有貶，褒的是質樸、堅實、寬闊，貶的是偏枯、乏味、老舊。

然而，到了 20 世紀末至 21 世紀初，當各先進國家致力於墾拓「數位人文學」（Digital Humanities）時，我們赫然發現這個年逾 90 的老機構竟然也很早就踏入這一個新而時髦的領域（1984）。這看似奇怪，其實並非偶然。

傅斯年當年的主張，用數位人文學的話語來說，就是要掌握「海量資料」（Big Data），並充分運用各種分析大數據的工具與方法。而史語所對於「史料」的高度重視，和數位人文學完全以「資料」（data）（源於各種形式、材質與載體的數位化資料及原生數位資料）為主體的做法，兩者同一陣線。此外，「動手動腳找東西」類似於資料檢索、資料倉儲與資料探勘。而且，在數位人文學的世界中，搜尋、解讀資料時通常更仰賴「機讀」而非「人讀」，同時，「書」也不斷被解構、消解成各種「文本」，漸漸的，「讀書人」也在消逝中。兩相對照，傅斯年的名言，儼然成為隱喻式的預言！也為史語所在舊有的殿堂中接納、營造數位人文學留下了空間。

但是，在進入數位時代之後，數位人文學不斷攻城掠地，人文研究的「典範轉移」（paradigm shift）似乎呼之欲出，因此，我們不禁要問：以史語所為代表的那種「典範」，會不會舊了？會不會被替代？畢竟，傅斯年所要「保持」的「亭林百詩的遺訓」，其實是清初顧亭林（1613-1682）和閻百詩（1636-1704）所開創的「考證學」（樸學）傳統，距今已經有數百年之久，而且，那是以文字為主、以書本為尊的時代。不過，這個問題太大，必須經由時間檢證。因此，我們不妨先從「考證」著眼，想想數位時代的人文研究還需不需要「考證」，舊有的「考證學」又會遭遇怎樣的挑戰，並藉此思考「典範轉移」的問題。

「考證」的基本工作在於「辨別真偽」，先是辨識材料的真偽，從而判斷事實的真假。就這一點來說，只要是「實證研究」，考證永遠有其必要。而且，無論是何形式和載體，材料通常會有一些元素可供檢驗，如創造者、創造時間、創

造地點、儲藏地點、傳輸方式等。但是，數位時代或數位情境中所要處理的材料，涵蓋了「數位化資料」（digitalized data）和「原生數位資料」（born-digital data），因此，傳統的考證方式和技能是否能夠適用，確實必須重新檢視、省思。

貳、「數位化資料」的陷阱

首先，我們發現由傳統文獻轉化為數位格式的「數位化資料」，在運用於研究時，或是說用來「重建事實」、「「發現真相」時，必須留意以下陷阱。

一、不忠的「複本」

版本學和校勘學一直是傳統考證學不可或缺的一部分。因為，在「文字時代」，無論是手寫傳抄，還是刻版印刷、排版印製，始終有抄本、刻本、印本「不忠」於「原本」的情形。造成這種「複本」背離原典的原因，有時是基於商業或技術的考量，有時是政治或宗教力量的介入，有時是後人基於不同的認知或意識形態而做的「編輯」和「整理」，有時則純粹是複製過程中大意、疏忽所造成。

這樣的問題也出現在傳統文獻轉化為數位資料的過程中。以史語所的「漢籍全文資料庫」來說，其前身是 1984 年開始推動的「史籍自動化計畫」（1984-1990），主要的工作是將中國史籍的全文轉製為電腦可以呈現、閱讀、檢索的數位文本，第一年先完成宋代以前的「正史」〈食貨志〉，其後則展開「二十五史」的全文輸入工作。那時候（1985），我正在台灣大學歷史學研究所碩士班就讀，因緣際會，以兼任助理的身份參與了這個計畫，負責「史記」、「漢書」的校對工作，並定期和資訊科技領域的專家、工程師討論資料庫的建置工作。那時，史語所人文領域的學者堅持「電子版」的史籍必須「忠於原典」。這個要求看似合理而簡單，但如今回想，在那個時候，其實是個災難，也是不可能的任務，因為，「中文電腦化」（包括建立中文字的標準交換碼，中文字型的數位檔以及中文輸入法）在 1980 年前後才初步完成，交換碼和數位中文字的數量遠遠不足以「忠實」的呈現史籍所用到的字，因此造成了大量的「缺字」現象，而且，「史記」中份量不輕的「表」（十卷），當時的技術也無法呈現。換句話說，當時的電子版「史記」其實和紙本的「史記」有相當大的差異，並非「史記」的原貌。

缺字和表格呈現的問題，後來雖然逐漸獲得解決，但仍然不能算是完善。而在「史籍資料庫」完成之後（1990），史語所隨即展開二十五史之外的各種經、史、子、集的傳統文獻全文資料庫建置工作，並在 1993 年將資料庫更名為「漢籍全文資料庫」，但是，這樣的擴張也帶來了新的挑戰，其中最不易克服的仍然是「忠實」問題，例如，「正統道藏」中有不少獨立或是和文字夾雜的符、圖、像、表，如何完整而真確的在電腦上「再現」，並可以被檢索、近用，仍然有待改善。

此外，為了節省時間和人力成本，一些漢籍的全文資料庫在建置過程中，都

會採用光學字元辨識（Optical Character Recognition，OCR）技術，這項技術雖然已經發展多年，而且技術成熟，但是，辨識率即使高達九成或九成五以上，對於一個提供專業研究的資料庫，其錯誤率仍然令人難以接受。校勘依然是尊貴的行業！

至於色彩、聲音、影像的數位化，我相信一樣會有「失真」的情形，但這不是我的專業，在此不談。無論如何，如果欠缺良好的「數位保存」（digital preservation）策略、技術和管理，任何一種數位資料，都可能產生資料缺損、變改、錯亂，甚至「消失」或無法被「近用」的情形，這也算是一種原典的破損或毀滅。

二、選擇性的記憶

記憶是歷史最核心的部位，長期以來，許多人都相信有所謂的「集體記憶」或「共同記憶」（collective memory），而且相信這和族群、社會、文化、國家認同息息相關。因此，不少國家、社群都曾經進行過形形色色的建構「共同記憶」（歷史）的文化工程。以傳統中國來說，書寫「正史」、國史、地方志、族譜，編纂佛教和道教的「藏經」、醫學領域的「醫方」（如：太平聖惠方、聖濟總錄、普濟方）、大型的「類書」（如：藝文類聚、太平御覽、冊府元龜、古今圖書集成）和叢書（如：四庫全書）等，都是這一類的文化工程。而台灣政府從 2000 年起陸續啟動的「國家典藏數位化計畫」、「數位典藏國家型計畫」、「國家文化記憶庫計畫」，也都可以算是這一類的工作。

這一類的文化工程，主導者和主要參與者基本上都是少數的菁英，甚至是統治階層的人士。他們的工作其實是一種「記憶篩選」。他們將符合「政治正確」（political correct）及其信仰、利益、價值、品味的「材料」納入「共同記憶」中，不符合的則被忽視、捨棄、刪除。因此，我們可以看到宗教經典有所謂的「疑經」、「偽經」，有的「入藏」（編入佛藏、道藏），有的「出藏」（排出佛藏、道藏）。我們也可以看到「四庫全書」有所謂的「著錄書」（全書收入）、「存目書」（只保存書目而不收入全書）「未收書」（無書目也不收入全書）和「奏毀書」（不僅不保留書目，全書還必須銷毀，禁止刊印、典藏）。而近代中國和台灣的「禁書」令，依然是部分人士的親身經驗和「共同」記憶。

因此，我們的「共同記憶」，其實一直都是少數人「整理」出來的，只是手段有時粗暴，有時精巧。我們的歷史也一直是掌握權力者所書寫、所廣播的聲音，只是內容有時動聽，有時難聽。

因此，如何讓那些被遺忘、被刪除的記憶恢復，如何讓那些被噤聲、被消音的話語喧嘩，如何掌握所有的材料，重建真正的全體的歷史，依然是數位時代嚴苛的挑戰。因為，「數位化資料」無論是否忠於「原典」，到目前為止，依然是一種高度選擇性的記憶。

三、魅惑的關鍵詞

自從有了數位化資料庫和簡便的檢索系統之後，有些人文學者覺得如獲至寶，認為只要掌握幾個「關鍵詞」（keywords），就能找到自己所要的資料，並逐漸拋下過去仰賴的一些工具書（如引得、辭典），也很少再做閱讀筆記或協助索引的卡片。而且，隨著數位化資料數量的快速增長，人文學者不得不經常進行泛覽、翻閱與檢索，花在精讀與熟讀文獻的時間愈來愈少。這種工作習慣的改變，看似微小，但在一、二個世代之後，恐怕會對人文研究帶來革命性的轉變。而轉變的關鍵就在於充滿魅惑力的「關鍵詞」。

「關鍵詞」確實非常好用，可以節省學者大量的時間和精力，讓研究者可以在短時間之內，掌握一批陌生的資料或一些研究課題，甚至進入一個全新的領域。但是，快速也正是危險所在。過度仰賴「關鍵詞」檢索資料，有時會漏失重要資料，造成誤判。有時則會撈捕到大量的垃圾資料，反而耗時。

造成這種缺憾的原因，主要是中國傳統文獻有相當多的異體字和假借字，而且常會因為不同時代的「避諱」而出現缺筆字或替代字，如果不先進行研究並設定「同義字詞表」或權威檔，只憑簡單的「關鍵詞」進行檢索，便會有資料缺漏的情形。其次，中國文字，一字（一詞）多義、一物（一地、一人）多名和一名多物（多地、多人），相當常見，往往必須根據字詞所在的文本類型和文章脈絡，才能判斷其意涵，因此，如果逕自將「關鍵詞」檢索的結果納入「詞頻」分析、關係網絡分析、視覺化呈現，其準確度便會大打折扣。

參、「原生數位資料」的問題

至於「原生數位資料」，或多或少也會有上述的陷阱，而且還會有一些更複雜的問題。

一、流動的文本

在古代中國，訊息的傳布、知識的創作和情感的表達，曾經歷過從口語到文字的漫長轉換過程，早期的「文本」是以篇、章、卷的形式和單位在書寫、誦讀和流傳，而傳誦者通常可以自由的增刪、改造，因此，「祖本」與傳本之間的差異有時會相當驚人，從春秋時期（西元前 8 世紀），一直到漢代末期（西元 3 世紀），透過不斷的整理和編輯，一本一本較具體系性的「書」才出現，「文本」才被「固定」下來。因此，現代學者考證先秦古籍時常會說：「此非一時一人之作」。換句話說，古代的書，確切的作者和確切的創作年代，幾乎無法考證。即使是在漢代之後，由於散逸、傳抄、編輯、注疏、偽託等因素，書的作者和年代考證，

有時依然是一大難題，因此，文獻考訂、目錄學、版本學一直是傳統中國文史研究不可或缺的學問。

「原生數位資料」看似不會有這類的問題，因為，一篇文章、一張圖片、一首音樂、一段影片，在產出過程中，產出設備大多會自動記錄、產出「後設資料」（metadata），提供基本的資訊（如時間、地點、規格等訊息），編修的記錄也可以自動保存。但是，當「原生數位資料」是採取「眾人協作」（crowdsourcing）的方式生產，或是將資料公佈於網際網路、社群媒體，那麼，即使不是出於惡意，全部或部分內容，便很容易被擷取並進行改作、拼貼、傳播，原本「固定」的文本，會像病毒一樣，快速的複製、演化、流傳，成為一種「流動」的文本。除非有很好的控管機制和技術、明確的使用規範，並且仰賴閱聽者的高度自制與遵從，否則，不久之後，許多「原生數位資料」仍然會被說是：「此非一時一人之作」。

二、抄襲的創作

中國傳統經典有很長遠的注疏傳統，在經典「原本」之上，層層堆疊、交錯著不同時代、不同學者的註解。至於註解的形式，有些「以經解經」（在「原本」的脈絡中，利用經文解釋經文），有些旁引其他文獻，有些直抒己見。雖然注疏家大多會交代其所引述的材料來源，但是，經過後人的「整理」、傳抄或是刻板、排版印刷之後，原本層次分明的經文、注文、疏文，有時候便會出現混亂，讓我們無法判斷某段文字的原創者。不過，這終究是無意於抄襲或剽竊。

類似的情形也可見於中國傳統醫學文獻。「醫經」（如「黃帝內經」）的情形和儒家經典、諸子百家的經典一樣，也有注疏傳統。但是，除此之外，「醫方」（藥方）和「本草」類的文獻還有很特殊的「輾轉相抄」現象，不交待材料來源以及添加己見的情況也更加普遍。而且，常有醫學類書的編纂。因此，「非一時一人之作」的醫學文獻也相當多。宗教經典也是如此，尤其是「道藏」，一些科儀、方術、法術類的文獻，大多是師徒之間歷經多代的「秘傳」，有時還夾雜口訣、秘本，或是宣稱是神人、仙真傳授。這類文獻的後續增刪者或編輯者，基本上也是無意於抄襲或剽竊，但是，這類的做法往往讓人很難考證真正的作者和時代。

除此之外，中國傳統文獻還有類書（類似西方的百科全書）與筆記這兩種文類，常會隱沒最原始或真正的作者和時代。大部頭的官修類書，由於編者的文化水平較高，而且有官方資助，對於資料來源的交待通常比較清楚而正確，但有些私修的類書，或是出於商業考量，或是限於知識，或是為了炫耀才華，有時就會不註明原始出處，有意或無意的抄襲或剽竊前人的創作。這在一些以「炫奇」、「炫博」的志怪、小說、筆記中，更是常見。

不幸的是，這種「無意的抄襲」或是「抄襲的創作」在「原生數位資料」中還是會存在。在網際網路和社群媒體的世界裡，隨意的「複製」他人所創作的內容（部分或全部），然後不加一語或附加摘述、評論，按一下「分享」，其實已經

產製了另一個文本。而一些引起爭議的「貼文」，留言區的討論文字和意見，其數量與豐富度往往超過「原貼」，若將這些「討論」以及被「分享」出去所引發的「討論」集結在一起，視為一個「文本」，那麼，原創者與抄襲者將會難以區分，也不是「非一時一人之作」一語就可以形容。至於「維基百科」這類的文本，如何釐清其內容中的原創與抄襲，正確與錯誤，則又是另一種難題。

三、偽造的事實

從晚清到民國初年，中國學術界瀰漫著一股「疑古」的思潮，只要是「古代」的，動輒被質疑為「偽造」的，古書、古董、古人、古史都被人一一重新檢視，而在一連串的「打假」行動之後，逐漸有人相信原本的「信史」有不少內容只是「神話」，甚至是刻意的「虛構」。這股「疑古」思潮強勢的時間並不長，但卻有持續的「解構」力量，三皇五帝的歷史鎖鏈崩解之後，中國古史的「重建」工作也一直延續到現在。而歷史材料和「歷史事實」（historical facts）會被偽造，也成為常識，大家似乎變得更聰明，更有警覺心，更不容易受騙上當了。

然而，「原生數位資料」中的「偽造」風氣仍然相當盛行，而且更難甄別。其中，最常見的就是網路媒體的「新聞」。有些只是將「廣告」偽裝成「新聞」報導，有些則是利用聳動、羶腥的「假標題」，誘導讀者點閱一則和標題毫不相干的新聞。更為可議的是，基於政治、宗教、商業、仇恨等因素，還有人會刻意的製作「假新聞」，並加以傳播。此外，社群媒體中的「假消息」，在社群「同溫層」的催化之下，更容易在短時間之內成為多人所相信的「事實」。

這些行為，大部分看似只是資料的偽造，但是，基於這些假資料所建構的歷史事實，其實就等於是「偽造事實」，而當未來的出版形勢更加多元、自由而快速時，目前學界內部的「同儕審查」能否繼續執行「事實確認」的任務，不無疑問。

四、量化的危險

傳統中國時期，歷代政府留下了不少的官方統計資料，但不論是最重要的人口（戶口）數量、土地面積，還是次要的災害損失、物價數字等，學者在運用時都會非常小心，因為，從調查到登錄，從統計到呈報，每個環節都可能因為技術、能力、法規的限制，或是主事者個人的利害考量，而無法呈現真實的數字。因此，「量化史學」在傳統中國史研究中一直不太受人重視。

未來，這個情勢似乎要逆轉了。因為，「量化」與「計算」正是數位時代的基石，每一筆數位資料的「後設資料」（如時間長度、解析度、字數、尺寸、面積、體積等），基本上都可以被準確的「度量」和描述，其內容也可以被切割成更小的單位或是被標記，進而和其他的資料連結，或是進行統計、分析。

事實上，目前許多「數位人文」（digital humanities）研究都是奠基於這樣的

「量化」工作上，如詞頻、文字雲、社會關係網絡、視覺化呈現等。但是，稍有不慎，就很危險。例如，我們一般認為「笑」是正向的情緒詞彙，因此，很容易利用詞頻統計，找出情緒較為「正向」的文本、作者、時代，但是，在傳統中國醫學的知識體系中，笑卻是疾病的病徵之一，甚至就是疾病的名稱（笑疾），倘若不加甄別，其研究結論就容易淪為笑柄。

其次，大黃是傳統中醫的常用藥物，就像多數藥物一樣，從被認識到被使用，往往有盛衰起伏的過程，因此，藥物史研究的核心課題之一就是找出這樣的過程，而以「大黃」為關鍵詞，利用大型的跨代的漢籍全文資料庫，進行詞頻統計，似乎是最為便捷的方法之一。但是，偏偏大黃不僅是藥名，有時還是地名、旗名或其他物名，而且還有個俗稱叫「將軍」，一旦離開醫藥文獻，要自動而準確的找到「大黃」這味藥，並不容易。同樣的，想使用關鍵詞和詞頻統計來研究官職「將軍」，也必須小心。

再舉個例子，「祝由」一詞首見於「黃帝內經」，是傳統中國醫學的一種咒術療法，起源很早，而且，大約從西元 6-15 世紀，還被列為官方的醫學分科之一（先後被稱之為咒禁科、書禁科、祝由科），而即使是在官方廢除這個科目之後，一直到 20 世紀，中國社會都有祝由科醫者活躍的身影。而當我們以「祝由」為關鍵詞，查找資料量相當大的數位化的「大正新修大藏經」和「正統道藏」時，會發現「大正新修大藏經」中完全沒有這個詞彙，「正統道藏」則只有六筆資料，其中四筆還是因為「正統道藏」收錄了「黃帝內經」的關係。假如我們根據這樣的量化統計進行推論，認為中國的佛教和道教都沒有或是忽視「祝由」，那就大錯特錯了。因為，這兩個宗教的文獻較少或不使用「祝由」這個詞彙，但其經典中卻不乏咒術療法，尤其是道教，其看重咒術療法的程度絕不亞於官方醫學。

上述這三個例子，都是「數位化資料」以及「語意」問題可能產生的量化誤差，還不涉及人為的操控或有意的扭曲。假如我們將眼光轉向目前在網路世界最流行的幾個詞彙，如：點閱率、瀏覽人次、粉絲人數、網路聲量、按讚數、熱搜榜、暢銷榜等，那麼，稍稍有點警覺的人應該都會懷疑：那些統計數字是如何產生的？有沒有造假或人為干預？然而，當我們研究當代的新思潮、流行文化、消費習慣、經濟榮枯、社會恐慌、群眾心理、網路行為、公眾人物的社會形象與社會影響力等問題，我們卻又不得不仰賴這樣的量化數據。因此，量化即使危險，還是必須冒險。

肆、是生機？還是危機？

上述這些數位資料所存在的陷阱和問題，有時候會單獨存在，有時候則會交互作用而產生。有些是前數位時代的積累和遺毒，有些則是數位時代的新毛病。有些可用傳統的考證手法解決，但有一些則必須仰賴新的數位科技與思維。然而，使用任何新工具，都必須付出代價。

一、誰主浮沉？

在網際網路剛興起的時候，有句半開玩笑的話說：「沒有 IP（internet protocol address; 網絡互連協議位址）就沒有 ID（Identity; 身份）」。目前，大家還不會太認真看待這句話，但是，隨著數位化和網路化的進展，漸漸的，我們會發現，凡是在網際網路上無法被搜尋到的，幾乎就不存在，或是會被忽視，個人、團體、事物、歷史無不如此。因此，要剝奪一個人的權利，最簡單的方法就是不給予任何 IP 位址，因為，失去 IP 位址，在網路世界，將會失去身份、訊息、工作、朋友，甚至連日常生活（如就醫、消費）都難以維持。

同樣的，未來，一個人文學者，倘若離開了網際網路（或其他類型的通訊系統、雲端系統）或是數位情境，就會失去研究資料和研究工具，就會像沒了水的魚一樣，無法存活。然而，我們必須要問：是誰決定怎樣的資料能夠或應該被數位化？是誰決定怎樣的資料能夠或應該在網路世界存在並被「近用」(access)？是誰決定哪些資料應該從網路世界消失？是誰決定哪些人有「近用」資料的權利？是誰決定其他人應該如何「近用」資料？

過去，掌握資料生死與浮沉者，主要是少數的統治者、知識精英、宗教權威和大商人，現在，表面上看起來，也是如此，只是大家會多強調一點所謂的「市場機制」和「群眾智慧」（wisdom of crowds），也就是相信多數人的品味、價值取向、行為模式和實際行動也可以左右數位世界的動向。

但是，真的是如此嗎？好像不是這麼簡單。

二、注意「眾人協作」

過去的人文學者，從搜集、整理、分析資料，到撰寫、發表研究成果，大多數都採取單打獨鬥的「獨立」（或孤立）研究。目前，得利於全球化的網路系統和平台，許多工作都可以透過「眾人協作」（crowdsourcing）進行。這種方式，可以讓身處不同國家、地區的學者，可以群策群力的工作，不僅可以用來協助研究者搜集、整理資料，也可以用來分析資料、討論議題。至於各種協作平台則可以用來進行集體的協同寫作，線上出版通常也可以採取這種模式。有些資料的擁有者（包括圖書館、博物館、檔案館、美術館等），甚至會開放其數位藏品及數位空間，讓研究者和閱聽者自由「近用」資料，進行研究或線上策展，並上傳、補充自己的藏品（資料），這也讓知識的建構和傳播更具「民主」色彩和群眾性格。

就此而言，未來，對於人文學者來說，似乎就是個「美麗新世界」。然而，只要是涉及「集體」或「群眾」，就會有良莠不齊、牛驥同皂的問題。就人文研究而言，成於眾人之手的東西，或許會有數量和代表性的優勢，但很難避免流於品質低劣、觀點平庸。其次，人性貪婪，在共創共享的工作團隊之中，想要坐享其成者多，願意犧牲奉獻者少，很難讓所有成員都竭盡其能。此外，為了「容眾」，讓更多人參與協作，必然有許多作為必須「民主」、必須妥協，必須講求標準化

與一致性，如此，其作品雖然可能可以吸引更多的「受眾」，但是，終極而言，傑出人文學者的獨特觀點與價值判斷，以及學術社群的多樣性必然會被斲傷。

更執得注意的是，和我們「協作」的還不一定是「人」！

三、小心「機器人」

自從 IBM 的超級電腦「深藍」（Deep Blue）打敗世界西洋棋的棋王（1997），Google DeepMind 的「阿法狗」（AlphaGo）擊敗諸多頂尖圍棋高手，成為世界第一之後（2017），許多人突然發現原來「人」已經被超越了，而且是被自己所創造的「物」超越了。這種物，或許可以概稱為「機器人」（Robot）。雖然不是所有機器人都比人聰明，但是，目前機器人已經可以代替人類戰鬥、破壞、欺騙、犯罪，或是救難、勞動、生產、社交，也可以提供人類醫療、娛樂、資訊、諮商、學習等方面的服務。

事實上，每一部電腦、每一支智慧手機，也可以說是一個機器人，而各種裝置和網路之中的軟體、應用程式，也都算是機器人。因此，機器人自然也是人文學者從事研究工作的利器、夥伴。目前，大多數的人文學者，在知識的生產、儲存、管理、呈現和傳播過程中，幾乎每個環節都無法排除機器人的參與。而參與就會干預。我們必須學會它們的語言，才能發號施令、讀取訊息。我們必須理解、順應它們的功能和行為模式，才能「頤指氣使」、當家作主，而且，還很難質疑或不接受它們的工作成果。它們一旦罷工或生病，都會是我們的災難！換句話說，在數位情境中，人文學者和機器人將互為主僕、互為君臣！

我們的生活中，已經到處都有機器人的身影。有些是可以觸碰的、具體可見的，甚至是人模人樣。有些則是隱形的，匿蹤、存活於網路世界。未來，我們可能不再問「機器人能做什麼？」而是問「機器人不能做什麼？」「人機共和國」將逐漸成形！因此，未來，人文學者所要研究的「社會」，將必須同時注意人和機器人的行為與活動。未來的「數位文明」必將是「人機共創」！至於會不會演變成只剩「機器人文明」，暫時也許不必那麼悲觀，但究竟會如何，也許要問問具有「人工智慧」（Artificial Intelligence; AI）的機器人。

四、想想「人工智慧」

在「人工智慧」這個詞彙剛出現的時候，有些人偶而會戲謔的稱之為「工人智慧」，但是，逐漸的，各行各業都在談 AI。有智慧和沒智慧的，有錢和沒錢的，有權和沒權的，資訊人和非資訊人，不少人開始意識到 AI 快要等同於智慧、金錢、權勢和資訊。前述的「阿法狗」（AlphaGo）就是「人工智慧」的代表作，它可以說是一個具有擁有感知、學習、思考、推理、判斷和創造能力的機器人。

有一天，具有「人工智慧」的「物」，甚至可能成為人類崇拜、仰賴的神！至少，會逐漸宰制、影響許多人的行為、心靈和命運。目前，「人工智慧」已經

開始應用於「大數據」分析，用來「預測」疾病、氣象、生產、政治、經濟與市場的趨勢與變化，而有些人還會根據這樣的預測採取因應措施，甚至交給機器人代為決策與行動（如：股票交易）。

那麼，擁有這樣一種「神」一樣的武器好嗎？以人文研究來說，無論好壞，「人工智慧」和「大數據」帶來的將會是天翻地覆的改變。未來，我們必須處理的確實是「海量資料」，因此，必須仰賴具有高超智慧和運算能力的機器人，然而，我們用來分析、處理資料的方法，觀察和描述人類行為與社會現象的角度，以及呈現探索結果的工具和形式，都將因此產生極大的變化。

我們可以大膽預測，未來的人文研究，探索現象之間的「關聯」性將會取代「因果」關係的分析，「預測未來」將重於「了解過去」，「鳥瞰」（distant reading）資料將取代「細讀」（close reading）資料。文字敘述的魅力會逐漸消退，詮釋意義和解釋事實的工作會逐漸式微，取而代之的是可以簡單的、一目了然的「視覺化呈現」的數字、圖表與空間分布圖，以及近乎詩、格言與廣告詞的短語。而且，智慧的機器人所分析的資料，大半都還是它們自己所製造或偽造的！

這樣的世界，究竟是天堂還是地獄？讓我們再想想！

伍、結語：是素養，不是技能

中國古人尊崇「易經」，並說「易一名而含三義」：「一易簡」、「二變易」、「三不易」。翻譯成現代語言，意思是說：「易」這個字有三個含義，一是簡單二是變動，三是不變。倘若借用這個道理或是這樣的語言，來回答本文所提出的問題，那麼，我們的結論將會是這樣的「簡單」：一、所有的「典範」都會「轉移」，這是不變的、單一的原則。數位時代的人文研究「典範」必須也必然會「轉移」。二、所有的人文研究，無論是傳統時代還是數位時代，都需要立足於「資料」，也需要「考證」。但是，資料的形式和運用資料時會遭遇的問題，「考證」的方法和工具，則是變動不居的，這是不變的、單一的原則。

因此，在「資料為王」、機器人稱霸的數位時代，人文學者的確需要兼習數位科技與人文知識，兼具人文涵養與「運算思維」（computational thinking）。但是，這樣並不保證我們就能在數位世界騰雲駕霧、暢行無阻。畢竟，這只是素養（literacy），不是技能（skill）。而這種新素養，或許可以稱之為「數位考證」（digital textual criticism）。

2019/11/23 完稿於南港中央研究院「學人宿舍」

参考書目

Aiden, E. and Michel, Jean-Baptiste. *Uncharted: Big Data as a Lens on Human Culture*. New York: Riverhead Books, 2013.

Bartscherer, T. and Coover, R. eds. *Switching Codes: Thinking through Digital Technology in the Humanities and the Arts*. Chicago and London: University of Chicago Press, 2011.

Bauerlein, M. *The Digital Divide: Arguments for and Against Facebook, Google, Texting, and the Age of Social Networking*. New York: Jeremy P. Tarcher, 2011.

Berry, D. M. *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012. [online]

Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Mass.: The MIT Press, 2015.

Cohen, D. J. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press, 2005.

Davidházi, Péter ed. *New Publication Cultures in the Humanities: Exploring the Paradigm Shift*. Amsterdam: Amsterdam University Press, 2014.

Deegan, M. and McCarty, W. *Collaborative Research in the Digital Humanities*. Farnham: Ashgate, 2012.

Deegan, M. *Text Editing, Print and the Digital World*. Farnham: Ashgate, 2009.

Deyrup, Marta Mestrovic ed. *Digital Scholarship*. New York: Routledge, 2009.

Eve, M. P. *Open Access and the Humanities: Contexts, Controversies and the Future*. Cambridge: Cambridge University Press, 2014.

Gold, M. ed. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2012. [online]

Hayles, N. K. and Pressman, J. eds. *Comparative Textual Media: Transforming the Humanities in the Postprint Era*. Minneapolis: University of Minnesota Press, 2013.

Inman, J. A., Reed, C. and Sands, P. eds. *Electronic Collaboration in the Humanities: Issues and Options*. Mahwah, N. J.: Lawrence Erlbaum Associates, 2004.

Jockers, M. L. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press, 2013.

Leadbeater, C. *We-Think: Mass Innovation, not Mass Production*. London: Profile Books, 2009.

Lupton, Deborah. *Digital Sociology*. London and New York: Routledge, 2015.

Mandell, L. *Breaking the Book: Print Humanities in the Digital Age*. Chichester, West Sussex, UK: Wiley Blackwell, 2015.

Mayer-Schönberger, V. and Cukier, K. *Big Data: A Revolution That Will Transform how we Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.

McCarty, Willard. *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*. Cambridge: Open Book Publishers, 2010.

McGann, J. *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*. Cambridge, Massachusetts: Harvard University Press, 2014.

Meyer, Eric T. and Schroeder, R. *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, Massachusetts: The MIT Press, 2015.

Moretti, F. *Distant Reading*. London: Verso, 2013.

Ragnedda, Massimo and Muschert, Glenn W. eds. *The Digital Divide: The Internet and Social Inequality in International Perspective*. New York: Routledge, 2013.

Ramsay, S. *Reading Machines: Toward an Algorithmic Criticism*. Champaign: University of Illinois Press, 2011.

Ridge, M. *Crowdsourcing our Cultural Heritage*. Farnham: Ashgate, 2014.

Rosenzweig, R. and Grafton, A. *Clio Wired: The Future of the Past in the Digital Age*. New York: Columbia University Press, 2011.

Sabharwal, A. *Digital Curation in the Digital Humanities: Preserving and Promoting Archival and Special Collections*. Oxford: Chandos Publishing, 2015.

Schor, N., Weed, E., and Rooney E. eds. *In the Shadows of the Digital Humanities*. Durham, NC: Duke University Press, 2014. [A special issue of *Differences*, 25:1(2014)]

Schreibman, S., Siemens, R., and Unsworth, J. *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. [online]

Schreibman, S., Siemens, R., and Unsworth, J. *A New Companion to Digital Humanities*. Chichester, West Sussex, UK: John Wiley & Sons Inc., 2016. [online]

Shirky, Clay. *Here Comes Everybody: The Power of Organizing without Organizations*. New York: Penguin Press, 2008.

Underberg, Natalie M. and Zorn, E. *Digital Ethnography: Anthropology, Narrative, and New Media*. Austin: University of Texas Press, 2013.

Vandendorpe, C. *From Papyrus to Hypertext: Toward the Universal Digital Library*. Champaign: University of Illinois Press, 2009.

Terras, M., J. Nyhan and E. Vanhoutte, eds., *Defining Digital Humanities: A Reader*. Farnham: Ashgate, 2013.

Wang, Fan-sen. *Fu Ssu-nien: A Life in Chinese History and Politics*. Cambridge; New York: Cambridge University Press, 2000.

Weller, M. *The Digital Scholar: How Technology is Changing Scholarly Practice*. London: Bloomsbury Academic, 2011.

Weller, T. *History in the Digital Age*. Abingdon: Routledge, 2012.

林富士主編，《「數位人文學」白皮書》。臺北市：中央研究院數位文化中心，2017。

曹銘宗撰稿，林富士總策畫，《遠見與承擔：中央研究院數位人文發展史(1984-2015)》。臺北市：中央研究院數位文化中心，2017。