

# 信任與陷阱：以隨機抽樣檢證結果分析臺灣歷史類史料資料庫之正確性

林佳蓉\*

## 摘要

電子資源的普及使得資訊的來源更為多元，以往傳統上只能呈現於紙本的傳統知識，可藉由科技技術的協助，以多形式多媒體表現出來，且不受時空限制，具有容易取得、多人共享、容易檢索等特性。過去圖書館重視實體文化保存，隨著時代潮流的影響，也逐漸轉變為數位的文化保存行為，數位館藏成為各機構發展重點之一，由於各國實際的支持，世界各地如火如荼的展開文化遺產數位化的工程。

文化遺產數位化提供許多珍貴的數位化文獻、文物，對歷史學研究者而言，數位典藏技術也帶來新的研究題材與新的研究方式。收集史料，並加以詮釋，是歷史學研究中重要的過程，對於史料的詮釋認識，攸關於歷史研究的正確性。故經過整理後而呈現的數位典藏資源，在數位化過程中的詮釋內容其正確性以及方向性，是數位典藏能否提供研究者更好研究基礎的重要關鍵。有鑑於此，本文以臺灣歷史類數位典藏資料庫為範圍，以臺灣重要歷史事件為關鍵字，實際操作方式檢證上述資料庫內容，並期作為日後發展臺灣近代史數位典藏相關資源時之參考應用。

關鍵字：數位典藏、史料詮釋

\* 國立臺灣歷史博物館展示組研究助理

來稿日期：2012年1月31日；通過刊登：2012年4月20日。

## 一、前言

臺灣歷史類資料之數位典藏工程發展，已有十餘年歷程，資料庫的數量逐步成長，使得以往許多不易取得及使用的史料，可在彈指之間呈現於電腦螢幕上。但是就像人類歷史上文字的出現及印刷術的問世一樣，雖然帶來了人類文化進程的革命性變化，但因為可能生產出錯誤的資料，同時也就引發「無錯不成書」的疑慮，除了造成使用上的不便，甚至也更引發研究上的扭曲。擁有更便利取得性的數位典藏資料庫的同時，是否也有同樣的問題呢？

歷史學研究除了是一門重要的人文學科外，對於一個國家或地區的文化傳承與集體知識的建構亦甚重要。歷史學常被譽為時間之學，透過對過往文獻、文物的詮釋、分析，進而重建、探討社會脈絡的變遷，並與現今社會搭起一座對話的橋樑。「歷史是過去與現在永無止盡的對話」（Edward. H. Carr 著、江政寬譯，2009：126）可說是為這個學門立下了最好的註腳。作為一門時間之學，歷史學的特性之一便是大量使用一手文獻、原物件來從事研究建構。根據 Duff, W., Craig, B. 與 Cherry, J. 在2004年所發表的研究，高達95%的歷史學者其研究資料來源為原本、原物件（Duff, W., Craig, B., Cherry, J., 2004:17）。這是因為歷史學門要求不能憑空想像，必須以「有一分證據說一分話」的嚴謹態度來從事研究，因此為了追尋一手史料，必須「上窮碧落下黃泉，動手動腳找材料」，研究成果總是要由所掌握證據的質與量，及了解該證據意義來完成。

然而，以文字、圖像、原件所積累的浩瀚史料，往往因散落各處，讓歷史學者需耗費大量時間、精力搜尋；而且大量的史料，又佔用了相當大的實體空間，攜帶亦有不便，因而壓縮了研究品質。史學研究者以往須以卡片紀錄的方式整理和建立自己研究領域的相關資料，並以人力方式排列研究材料，也必須地毯式地蒐集相關資料，並找出資料的關聯性。因此在從事研究之前，歷史研究者必須投入大量時間、精力作事前準備工作，方能進一步展開研究。

但數位典藏（Digital Preservation）技術的出現，使得快速檢索與壓縮儲存空間得以實現。數位典藏為以網際網路為媒介、以珍貴自然文化資產為內涵、以資訊技術（information technology）為工具，所打造的虛擬資訊庫（整合博物館、圖書館、與檔案館）；期望利用不斷演進的資訊處理技術，以數位化的方式保存、處理、與匯集

資訊，並進一步加值整合，以創造更大的資料價值與應用層面，達成知識管理與發展（knowledge management and discovery）的目標。若以產生的過程來看，「數位典藏」指的是以數位方式處理（拍攝、全文輸入、掃描等），並加上詮釋資料（Metadata）的描述，以數位檔案的形式儲存。透過這樣的數位化過程，使得歷史文獻、文物資料具有便於檢索與查閱的特性，可讓更多的研究者更容易取得第一手資料，也是歷史學者第一次可以如此快速掌握龐大的資料（項潔、洪筱盈，2006：2-18）。

數位典藏已將資料掃描建檔，所以研究者利用古籍善本、檔案、手稿等資料時，不再需要像過去一一拜訪典藏單位閱覽典藏。快速、免費取用之特質，對研究者而言，可減少搜尋及取用的成本，擺脫必須到特定地點才能做研究的限制。數位典藏穩定性的特質，可使研究者對數位典藏資源反覆且持續性的使用，並進一步作不同的分析（洪淑芬、羅雅如，2006：65-99）。現今已有眾多研究探討相關資料庫的建置過程及其帶來的改變，並多持正面肯定的態度。<sup>1</sup>

隨著數位典藏資料庫的大量出現，使用者的大幅增加，可看出其便利性的成形。但在方便之際，這些數位典藏資訊的正確性是否值得信賴，又是否會帶來研究錯誤？都是不可忽略的問題。但現今關於臺灣歷史類數位典藏的探討，多半著重於其所帶來的方便性，或搜尋技術的更新，而少觸及最根本的建置內容之正確性與否。本文嘗試彙以隨機的方式，對數個重要的臺灣近現代史史料資料庫進行檢索與查找，以期回應上述的問題，並作為日後深化研究時之參考。

<sup>1</sup> 隨著近年資料庫數量的擴增，相關研究的積累亦快速增加，可參見本文最後所羅列的參考資料。

## 二、資料庫之驗證—以隨機抽樣檢索方式

現有的各式臺灣歷史史料資料庫，就資料庫建置、設計及使用端所檢索頁面的差異，大致分為兩類：1、全文類型資料庫：資料庫採用全文打字建檔，如「臺灣文獻叢刊資料庫」<sup>2</sup>，「淡新檔案資料庫」、「臺灣歷史數位圖書館」、「智慧型全臺詩資料庫」等，部分並附有原文影像及後設資料。2、全宗類型資料庫：提供影像檔，輔以後設資料以供檢索及分類，如「清代宮中檔奏摺及軍機處檔摺件」、「臺灣日日新報」、「漢文臺灣日日新報」、「日治時期圖書全文影像系統」、「日治時期期刊全文影像系統」等資料庫。

歷史研究者使用資料庫的方式，最常為輸入研究所需的關鍵字詞，於資料庫內進行檢索與查找，後根據檢索所得進行分析。故資料庫中的全文內容或是後設資料，包括對於古籍的全文打字，及後設資料的選取建置標準之正確性便甚為重要，一旦有所錯誤，便可能造成檢索的不正確性。

若要核對正確性，需比較數位資料與史料原件（或掃描影像），現今所累積的數位資料甚多，已超越個人力所能掌握者。故筆者以下即嘗試對於數個臺灣史研究者較常使用，且資料性豐富的資料庫進行隨機驗證，並且將分析標的限縮於幾個關鍵字詞的精確度上，這樣的考量基本上是從資料庫建置流程圖（圖1）<sup>3</sup>的檢討而來，在整個資料庫建置流程中資料檢核的環節雖然是一項「是」、「否」的流程篩檢，然而對文獻資料而言，這個環節卻是最有可能出現錯誤態樣，在此試著以臺灣近代歷史上數個重要事件的關鍵字，來檢證幾個常用資料庫的內容。

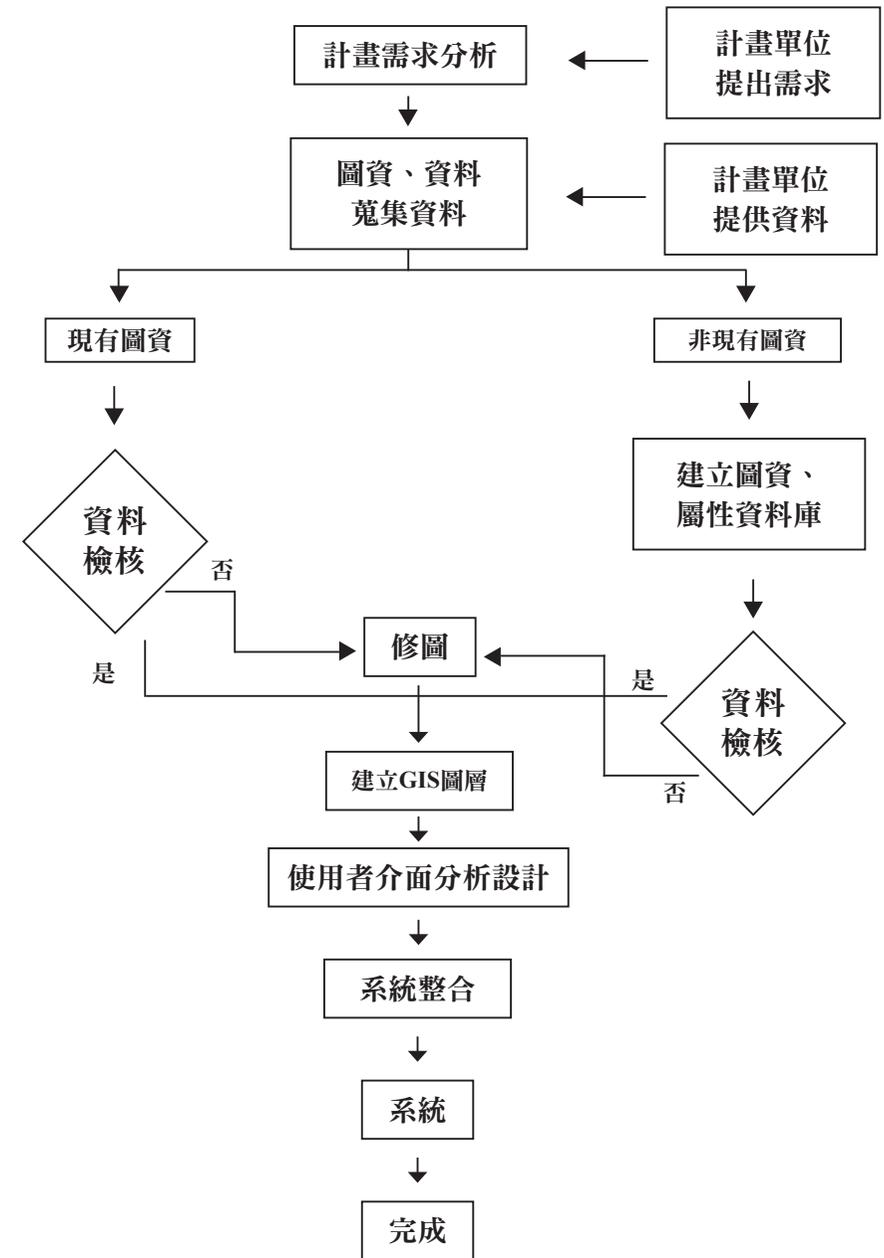


圖1 數位典藏空間資訊系統開發流程圖

<sup>2</sup> 本文中指涉資料庫者，以「」標示，若指涉紙本出版資料，則以《》表示。如「臺灣日日新報」，係指數位資料庫，《臺灣日日新報》則為臺灣日日新報紙本。

<sup>3</sup> 中央研究院計算中心，《計算中心通訊》第18卷11期，2002年05月06日。

(一) 錯誤態樣舉隅：缺漏字的疑義

(1) 案例一

檢索條件	由於臺灣文獻叢刊 <sup>4</sup> 內有眾多古字、俗字及異體字，如何顯示這些漢字，成為重要且必須解決的問題。已有數十年發展歷史的此資料，若以缺字「□」（系統顯示為□）進行檢索，為最直接之檢視方式，故以下即以此設定進行檢索。
檢索條列	<p>四 臺海使槎錄/卷三、赤崁筆談/物產</p> <p>- 53 -</p> <p>稻有___（稭）、糯。___（稭）稻種於水田者曰早占、曰晚占，色白；種於園者曰埔占，色赤。糯稻種於水田者曰赤秫，穀色赤；種於園者曰禾秫，粒大，色白。麥有大麥、小麥，而小麥最佳。稷之屬有細米，黃、白二種，蕭 ___（攏）、麻豆諸社有之。</p> <p>二一八 臺灣南部碑文集成/甲、記(下)/安平第一橋碑記</p> <p>- 350 -</p> <p>安平第一橋，大清同治十三年十二月除夕日吉立。</p> <p>由郡城出大西門至安平鎮，路經北勢街尾有港焉，阻往來者。郭氏渡以筏，人索錢三，遂專其利。同治壬申夏四月抵任，觀海於安平，道由是出。春夏之交，商人、舟子，漁戶、農民，赴城市購物者，日且千數百</p>

<sup>4</sup> 「臺灣文獻叢刊資料庫」是以臺灣銀行經濟研究室所出版的臺灣文獻叢刊309種子目為底本所建置的全文資料庫。該叢書係由原臺灣大學法學院院長周憲文，號召相關學者專家，蒐集包含清代臺灣方志、明鄭史料、清代官府檔案、私家著述、私人文集等資料，時代橫跨自唐、宋、元、明、清以下迄至20世紀前期，並旁及荷蘭、美國、英國、日本有關史料。可謂匯集當時後可見之臺灣相關歷史、地理、風俗、民情、政治、經濟、社會、文化、法制等之重要文獻集。其內容全套書計三百餘種、近六百冊、四千八百多萬字，選材經典，規模宏大，向為研究臺灣歷史最重要之文獻。1990年，中研院臺灣史研究所的前身臺灣史田野研究室開始規劃「臺灣史籍自動化」的長期工作，經史語所同意，並獲得計算中心的大力協助，比照二十五史資料庫，將製作臺灣方志資料庫列為第一階段的目標（詹素娟，1995），是第一個全文化的臺灣歷史史料資料庫，有相當之意義。故筆者選取此一代表性資料庫進行觀察。

<p>人，筏限以一，每次載六、七人，兩岸候而爭渡者相鬬也。民大不便，乃購木將為橋焉。有止之者曰：『橋成，誰與為守者？恐□訟滋發，而□□徒擾也』。余忍(?)焉。甲戌冬，觀察筱濤夏方伯築營於岸之北，鎮</p> <p>- 351 -</p> <p>海營駐焉。余以建橋之議商。夏方伯曰：『善』！歸而料匠未集。□□欽差沈大臣南巡琅（嶠），亟請於前。□□凌公司其事，鳩工庀材，不十日而成。橋長七丈八尺二寸，翼以欄，如其長之數而兩之；廣一丈四尺三寸。凡金工、木工、石工、土工之數，灰、磚、杉、□之料，糜白金百八十兩。時北軍有市驛者，厥駒良，購以酬之。</p> <p>候巡道臺灣府知府績溪周懋琦記并書</p>
<p>七五 恒春縣志/卷九 物產(鹽法)/飲食之屬</p> <p>(略...)</p> <p>- 188 -</p> <p>〈山魯〉岫鹽：海潮落後，水漬石上，括淋而攤晒之。今禁。</p> <p>(秫)米酒：白色。</p>

根據檢證結果，系統顯示共有92段，僅以上述3例來說明其中的錯誤態樣。

漢字屬於表意體系的文字，但現行電腦中的編碼（Encoding）設定，所有的字都只用 1byte 來儲存，1byte 包含 8bits，每個 bits 都只能表示 on/off，也就是 1byte 只能表示 0000 0000 到 1111 1111 的編碼範圍，只有 256 個編碼空間，這對中文而言是相當不足的。故這個問題不只是在此資料庫，由於古籍包含眾多手寫古字、異體字，如何將這些文字轉化為數位文字，一直是資訊相關研究者困擾的問題。

採原字呈顯的方式優點是可使讀者瞭解史料原有樣貌，但若為現代文字編碼系統中沒有的字型，呈現便必須透過特殊編碼。當時在臺灣最通行的編碼為「大五碼（Big5）」，若該字無在此編碼系統內，資料庫建置者另可造字，如臺灣文獻叢刊資料

庫現有提供造字檔安裝，但版本僅提供Windows 3.1至Windows XP間的版本，且部分電腦也非能完全安裝成功。故該資料庫雖已有十餘年的發展歷程，但相關問題仍無法解決，共出現92筆資料無法顯示，而如同前述三個個案，第一例為記錄清代臺灣農作，且各史料中常引用的物產資料，第二例為地名之例，第三例則為地方特色物產的顯示，都有相當的史料價值，卻皆無法顯示。因中文字偶有一字多型的情況，對歷史研究者而言，史料所提供的「事實」或許較書寫者在字型的選擇意義上更為重要，但字體若是無法呈現，即將帶來：1、影響檢索的成果，使得部分字詞無法被搜尋到。2、使用者對於地名，及對於涉及的特定內容物（農產品）進行判斷，導致研究者仍須查閱紙本，影響研究者的判讀及耽擱研究時程。

(2) 案例二

檢索條件	以「□」為檢索條件，檢證《臺灣日日新報》資料庫 <sup>5</sup> 中漢文版的資料。
檢索條列	防疫組合（承前） 日新街組合組合長葉連亭副組合長劉定坤委員七名 區域□九間仔街九間□後街日新街南市、□西市、同東市同北市、井仔頭街新店尾街長興街 稻新街組合組合長山口義章副組合長林金瑞委員十六名（內本島人十二名）

<sup>5</sup> 《臺灣日日新報》是臺灣日治時期發行時間最久，銷售量最大的報紙。其前身為1896年創刊的《臺灣新報》與次年創刊的《臺灣日報》，1898年在官方的主導下合併為《臺灣日日新報》，1944年再與臺灣其他五報合併為《臺灣新報》。戰後該報於1945年10月改隸臺灣省行政長官公署，改名《臺灣新生報》，其發行年代幾乎與臺灣總督府共始終。該報向以政府言論機關及「模範的殖民地報紙」自許，帶有御用報紙色彩，而報紙的深度、廣度皆有限制。但該報內每天數百條的新聞、廣告、專論，提供了豐富的當時臺灣社會之信息而且該報紙也登載漢文、日文的創作，從資訊量來說是其他史料難以望其項背者。它除了是全面性瞭解日治時期臺灣的重要史料，對於晚清的臺灣實況也有不少回溯性報導，對於戰後初期的歷史亦能提供一定的延續性線索（許雪姬，2004：1082-1083；蔡錦堂，2004：214）。國內現有兩大版本，分別漢珍版及大鐸版，分別由漢珍數位圖書股份有限公司及大鐸資訊股份有限公司建置。

區域建成街一、二、三丁目建和街建成後街南興街稻新街法主公街新興街城隍廟後街得勝街得勝外街
太平街組合組合長王國棟副組合長廖定國委員七名
區域太平街太平橫街平和街建興街建興後街怡興街維新街朝東街朝東後街
建昌街組合組合長管沼寬藏、副組合長陳粟委員十名（內本島人四名）
區域建昌街一、二丁目港邊街六館街一、二丁目千□街六館後街建昌後街城隍廟前街鴨寮街蘆竹腳街
媽祖宮口街組合組合長程文榛副組合長林再新委員六名
區域南街中街朝陽街永和街長樂街鐘厝街媽祖宮口街媽祖宮後街媽祖宮口後街
埤仔墘街組合組合長陳江流副組合長李祿星委員三名
區域李厝街港邊後街普願街杜厝街大有街埤仔墘街
國興街組合組合長郭三永副組合長陳錦秀委員七名
區域益保祐街枋寮街國興街枋寮後街依仁里街頂牛磨牛車街下磨車街大橋頭街
怡和巷街組合組合長陳采臣副組合長張家坤委員五名
區域中北街□和街怡和巷街珪瑜粹街飲和街獅館巷街隆街記後大龍峒街百斯篤豫防組合
大龍峒街組合組合長王慶忠副組合長陳培梁委員□名
區域大龍峒街
舊街組合組合長□澤八十次郎副組合長富士村格兵衛委員十

構造畜舍。內開溝道。以通污物。該處派出所警官。充當約束人。每月定巡視兩回以上。如不遵規定者。處以裁制其畜舍構造之方法完全列左。

- 一、畜舍須距離飲料水井五間以外。
- 二、構造畜舍之床盤。須用石板或洋灰煉瓦及灰土。後面左右側宜配通溝道。以排泄糞尿。
- 三、排泄溝須以石或灰土構造之。糞尿貯積所。內外施以藥水。須用甕或其他不滲透質之器具。
- 四、畜舍外圍。須用牆壁或以竹周圍之。畜舍構造完全。尤當時時洗滌糞尿貯積所。務使一切清潔。其能隨時排泄於公共溝渠更佳。即不然。亦宜並草收貯於污物之容器內亦可。



檢證結果說明  
標題錯置，該事件發生的地點是在臺南、而非屏東，顯示建檔者對於該報的結構並不瞭解，以致張冠李戴。

<p>檢證結果說明</p>	<p>原文有標記、分段、空格，以便於讀者閱讀判斷，但建檔時卻予以忽略，使得部分字詞連著，易造成使用者難以判讀。</p>

(二) 錯誤態樣舉隅：後設資料位置錯置

(1) 案例一

<p>檢索條件</p>	<p>依據關鍵字「阿緱」，驗證漢文版《臺灣日日新報》中的內容。</p>
<p>日期</p>	<p>標題：阿緱片信／規定構造畜舍 規定構造畜舍 臺南市人民飼養羊豚者。大都不構造畜舍。以致有時市內清潔。往往不能周至。在警察官吏。雖常致意於此飼養羊豚者。或告發其放出。或科罰其積糞。尚有未十分完善者。此次為豫防鼠疫起見。於衛生上研究種種方法。幾乎無微不至。當警務課規定。自來年一月起。凡有飼養羊豚者。須</p>

(2) 案例二

<p>檢索條件</p>	<p>以關鍵字「蓬萊米」分別檢索大鐸版及漢珍版「臺灣日日新報資料庫」<sup>6</sup></p>
<p>檢索條列</p>	

<sup>6</sup> 《臺灣日日新報》是臺灣日治時期發行時間最久，銷售量最大的報紙。其前身為1896年創刊的《臺灣新報》與次年創刊的《臺灣日報》，1898年在官方的主導下合併為《臺灣日日新報》，1944年再與臺灣其他五報合併為《臺灣新報》。戰後該報於1945年10月改隸臺灣省行政長官公署，改名《臺灣新生報》，其發行年代幾乎與臺灣總督府共始終。該報向以政府言論機關及「模範的殖民地報紙」自許，帶有御用報紙色彩，而報紙的深度、廣度皆有限制。但該報內每天數百條的新聞、廣告、專論，提供了豐富的當時臺灣社會之信息而且該報紙也登載漢文、日文的創作，從資訊量來說是其他史料難以望其項背者。它除了是全面性瞭解日治時期臺灣的重要史料，對於晚清的臺灣實況也有不少回溯性報導，對於戰後初期的歷史亦能提供一定的延續性線索（許雪姬，2004：1082-1083；蔡錦堂，2004：214）。國內現有兩大版本，分別漢珍/まに版及大鐸版，分別由漢珍數位圖書股份有限公司及大鐸資訊股份有限公司建置。

<p>檢證結果說明</p>	<p>以蓬萊米為關鍵字，大鐸版出現3,323筆資料，漢珍版有3,337筆。而蓬萊米係1926年4月由伊澤多喜男所命名，故不可能早於此月。但大鐸版中，有35筆出現在1926年4月之前，漢珍版有3筆。另觀察其他新聞，並未事先公布蓬萊米，而這38筆資料，也非敘述蓬萊米的命名之事。如漢珍版的3筆，分別為「早季蓬萊米□移出漸增」、「昨年晚糯移出」、「懸案中である□日月潭の問題は□遅くも五月頃迄には□何とか眼鼻が附かう」，而點閱資料庫提供的原件影像，第一篇並無此一報導，第二篇報導內容僅記載「內地種米」，第三篇內容無論及蓬萊米，恐是將「蓬萊丸」誤判讀為蓬萊米。故可知資料庫在建置時建檔人員誤判。</p>
---------------	---

(3) 案例三

「日治時期圖書全文影像系統」<sup>7</sup> 為以後設欄位所提供的文字供讀者檢索，並提供讀者該檢索所得的圖書之數位影像檔（PDF檔）。但卻也同時將應該是圖書的《臺灣史料稿本》之目錄內容打成文字檔匯入目前的資料庫中，因此就成了圖書內容也成了後設資料之資料庫的奇怪現象。

<sup>7</sup> 中央圖書館臺灣分館藏有為數可觀的日治時期臺灣相關圖書，現共約有16萬餘冊，原係日治時期臺灣總督府圖書館及南方資料館所藏，內容橫跨產業、政治、經濟、社會、醫學、歷史、宗教等方面。但多數圖書因歷時久遠，不堪翻閱，因此自1995年起，臺灣分館陸續將其中較重要者攝製為微捲，提供讀者利用。2007年後，更進而將這些館藏的日治時期圖書建立數位資料庫，使讀者能不受時空的限制，透過網際網路使用這些珍貴的典藏。《日治時期圖書全文影像系統》收錄許多重要圖書，目前約有3,700本圖書可線上檢索，雖名為「全文」，但現僅提供原件掃描之PDF檔影像，而非全文打字檔。

(4) 案例四

檢索條件	採用關鍵字「磯永吉」，檢證「日治時期期刊全文影像系統」 <sup>8</sup> 的內容。
檢索條列	 <ol style="list-style-type: none"> <li>1. 辭令欄（臺灣教育第三百三十八號）作者不詳 無版權頁 n338 內容</li> <li>2. 歐米に於ける遺傳研究の概況 磯永吉 無版權頁 內容</li> <li>3. 滿洲の農産 磯永吉 無版權頁 內容</li> <li>4. 臺灣の小麥 磯永吉 無版權頁 內容</li> <li>5. 目次(第一編 臺灣の小麥作事情) 磯永吉 無版權頁 內容</li> <li>6. 蓬萊米の品質特に胴割、食味色澤に關する研究 磯永吉 無版權頁 內容</li> <li>7. 水稻内地種 臺灣總督府中央研究所技師 磯永吉 作者不詳 無版權頁 內容</li> <li>8. 臺灣農事報 磯永吉 1912-12-20（篇名應為「稻の品種改良就て」）</li> </ol>

<sup>8</sup> 「日治時期期刊全文影像系統」的建置源由與「日治時期圖書全文影像系統」類似，主要是針對期刊內容。目前線上可查詢瀏覽的有《臺灣教育會雜誌》（後易名《臺灣教育》）、《臺灣建築會誌》、《臺灣警察協會雜誌》（後易名《臺灣警察時報》）、《臺灣通信協會雜誌》（後易名《臺灣通信協會雜誌》）、《臺灣產業雜誌》、《臺灣 水利》、《臺灣醫事雜誌》、《實業之臺灣》、《臺灣山林會報》（後易名《臺灣 山林》）、《臺灣婦人界》、《臺灣文藝》、《木瓜》、《蕃情研究會誌》、《臺灣經濟雜誌》、《臺灣農友會會報》（後易名為《臺灣農事報》）、《蕃界》、《南瀛佛教》（後易名為《臺灣佛教》）、《臺灣博物學會會報》、《專賣通信》……等79種。亦僅提供原件掃描之PDF檔影像，非全文打字檔。

12. 臺灣農事報 磯永吉 1913-12-20（篇名應為「作物育種と淘汰法」）
13. 論說報文 磯永吉 1914-09-30 n017
23. 調查研究 磯永吉 1938-10 n383
24. 調查研究 磯永吉 1938-11 n384
26. 論說研究 磯永吉 1942/9/30 n009



檢證結果說明

以第8筆「臺灣農事報 磯永吉 1912-12-20」為例，該篇文章題名為〈稻の品種改良に就て〉，但資料庫卻顯示為「臺灣農事報」；而其他數篇亦有以「論說報文」、「調查研究」、「調查研究」、「論說研究」等為標題。顯然建檔者係以在該篇文章中所出現的第一行作為標題，因而造成極大的錯誤，將嚴重誤導使用者。此外，有數筆顯示無版權頁者，其實可以相當容易在論文內找到出版年代。如〈滿洲の農産〉一文，其出版時間為1934年1月。該資料庫中類似的錯誤甚多，應是在公開前未進行對後設欄位作一定的檢核，如此將使得研究者必須再費時間判斷其所檢索到的資料。

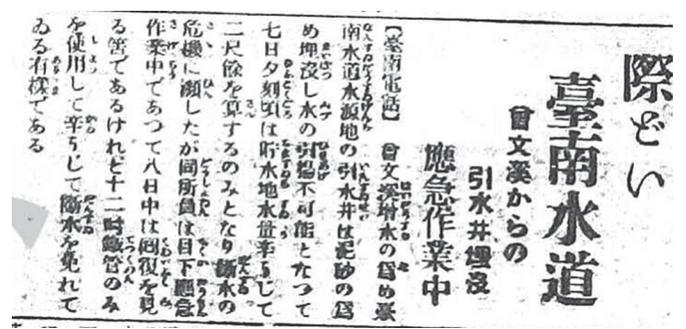
(三)、錯誤態樣舉隅：內容判讀錯誤

(1) 案例一

檢索條件	以「□」為檢索條件
檢索條列	<p>全島書房統計</p> <p>欲使新教育普及。將如何處置舊來書房乎。今尚為教育上一大問題。聞最近之統計。其數如左。並將同期中公學校生徒數。記載在下層。以供彼此對照之便。</p> <p>廳別/房數/男生/女生/合計/公學校生徒</p> <p>臺北/九四/二、二八四/二一/二、五〇五/三、一三三</p> <p>基隆/五七/九五/一/二五/九七六/八三八</p> <p>宜蘭/六二/一、六九一/八/一、六九九/八二二</p> <p>深坑/二七/四二一/一八/四三九/六八〇</p> <p>桃園/七五/二、一九一/二/二、一九三/一、五七九</p> <p>新竹/一三五/二、七五三/五八/二、八一/一、四三三</p> <p>苗栗/一三五/二、二三八/一七/二、二五五/一、四三八</p> <p>臺中/五七/一、二八〇/□/一、二八〇/一、七八六</p> <p>彰化/一九/三六六/二/三六八/二、六九二</p> <p>南投/一七/二九八/三/三〇一/八一七</p> <p>斗六/九六/一、三三三/三六/一、三六九/五六五</p> <p>嘉義/四二/八四四/六/八五〇/一、五五七</p> <p>鹽水港/九八/一、七八六/八/一、七九四/一、〇四八</p> <p>臺南/六七/一、〇六七/二一/一、〇八八/一、〇二〇</p> <p>蕃薯寮/七/一三六/□/一三六/三三五</p> <p>鳳山/一九/二八二/六/二八八/九六三</p> <p>阿緱/二九/五三七/□/五三七/一、三二六</p> <p>恒春/□/□/□/三〇七</p> <p>臺東/□/□/□/□</p> <p>澎湖/四六/九六八/四/九七二/八三九</p> <p>計/一、〇八〇/二一、四二六/二三五/二一、六六一/二三、一七八</p>

	<p>即其生徒數。公學校僅較多千餘人。而所要注意者。為與文明之普及。共得所在受教育悉加多。乃書房生徒。年年不減少。轉使從來幽處深閨之女子。出而受書房教師之教者。有二百三十五人。就令公學校極隆盛。而書房自然不能斷絕。觀今日之形勢。其可顯然證教育普及者乎。</p>
檢證結果說明	<p>若以□來檢索，檢索結果共 22,773 筆，約佔該資料庫全部筆數的0.07%。但從一般的史料處理原則來推斷，□可能包含多種情況，例如缺字（包括無法辨識等）。比對原件：這個例子裡的□，其實是「一」，即無統計資料，而非缺漏。類似這個例子將無統計之料逕行以□表示，變容易造成研究者的誤判。</p>

(2) 案例二

檢索條件	關鍵字：曾文溪
檢索條列	<p>標題：際どい 臺南水道 曾文溪からの 引水井埋没 廳急作業中</p> 
檢證結果說明	<p>原件顯示為「應急作業中」</p>

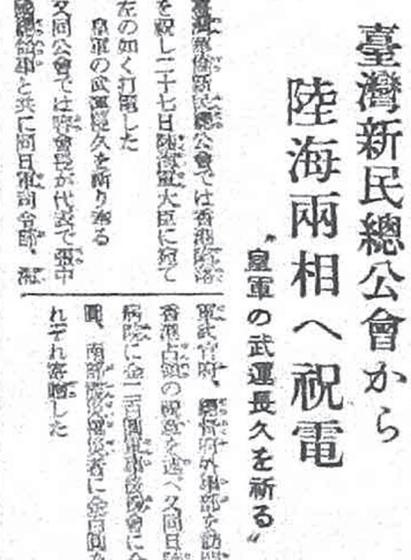
(2) 案例三

檢索條件	關鍵字：台中
檢索條列	標題：人事 磯永吉氏（臺中國技師） 
檢證結果 說明	原件顯示為「臺中廳技師」

(4) 案例四

檢索條件	關鍵字：台中
檢索條列	標題：臺中國優勝 全島小公學校職員 陸上競技大會終る 
檢證結果 說明	原件顯示為「臺中國優勝」。

(5) 案例五

檢索條件	關鍵字：台中
檢索條列	標題：臺灣新民總公會から 陸海兩相へ祝電 “皇軍の武運長久を祈る” 系統顯示關鍵詞：臺灣華僑新民總公會；臺中國總領事 
檢證結果 說明	原件為「張中國總領事」。

(6) 案例六

檢索條件	關鍵字：磯永吉
檢索條列	要題名：觀光の臺灣 作者(姓名)：金高佐平 語文：日文 總頁數：27 出版年(西元)：1937年[昭和12年7月28日] 出版單位：合資會社臺北活版社

	<p>出版地: 臺北市</p> <p>索書號: 0748 116</p> <p>發行者: 金高佐平</p> <p>關鍵詞: 臺灣、觀光、鐵道、臺灣宣傳事業社</p> <p>目次: 目次</p> <p>序……鐵腕生 p.1</p> <p>臺灣と丙地（正確應為內地）との交通 p.1</p> <p>主要題名: 臺灣の米</p> <p>作者(姓名): 西田龍八</p> <p>語文: 日文</p> <p>總頁數: 189</p> <p>出版年(西元): 1934年[昭和9年]</p> <p>出版單位: 東京米穀商品取引所検査課</p> <p>出版地: 無版權頁</p> <p>索書號: 0792 158</p> <p>發行者: 無版權頁</p> <p>服務單位: 東京米穀商品取引所検査課</p> <p>職稱: 検査課長、技師</p> <p>第二節 風土と稻の禮（正確應為種）別の關係 p.41</p> <p>關鍵詞: 臺灣米、蓬萊米、東京米穀商品取引所検査課長西田龍八</p> <p>目次: 封面</p> <p>第二章 生産</p> <p>第一節 耕種の概要 p.17</p> <p>第二節 風土と稻の禮（正確應為種）別の關係 p.41</p>
<p>檢證結果 說明</p>	<p>在該資料庫中，有眾多錯字，主要集中在目次內，將影響檢索率及使用者的判斷。</p>

(四)、錯誤態樣舉隅：未設權威檔 (authority file)<sup>9</sup>

權威檔又稱為規範文檔，它是由權威記錄 (authority record) 組成的電腦檔。由於權威標目的種類，包括書目記錄的檢索點 (編著者名稱)、統一題名、叢書名、主題標目等。因此，權威檔可分為名稱權威檔、主題權威檔、叢書權威檔3種。其中，名稱權威檔包括個人名稱、國家名稱、機關團體名稱、地理名稱和統一題名等。權威檔作用在於實行權威控制，保證機讀目錄中文檔標目的唯一性和一致性，以便有效實現對機讀目錄的統一管理，下列案例可能是因未設有「權威檔」，所導致的錯誤態樣。

案例：

<p>檢索條件</p>	<p>關鍵字：磯永吉、磯博士</p>
<p>檢索條列</p>	<p>版本：漢珍版 磯永吉：47筆 磯博士：59筆 版本：大鐸版 磯永吉：18筆 磯博士：30筆</p>
<p>說明</p>	<p>未做權威檔，使研究者需以自身的專業知識判斷，如上述例子，需鍵入不同關鍵字分別檢索。</p>

<sup>9</sup> 權威檔可控制標目能有唯一性和一致性，包括下列三層含義：1.標目的唯一性：對於同實異稱之標目，只允許選其一為標準形式；對於異實同稱之標目，則加註識別形式。2.標目的一致性：對於多人同時作業或同一人異時作業，其所選取之標目，為一致的、一貫的。3.標目間的關係：包括同實異稱標目間之「見」參照，機關名稱間之改名、改制、改之「參見」參照和異實同稱標目間之「說明語」形式。使用於歷史類資料庫內，將可對於同實異稱之標目進行連結（國家圖書館編目組，2006）。

### 三、臺灣歷史史料資料庫發展的議論

各個學科的成員，常有獨特的研究特性。根據吳明德於2006年研究整理指出人文學者研究的特性，主要可歸納為數項：1、常反覆閱讀古老的文本。2、需不斷回溯過往的研究，再融合新的研究。歷史學者使用資料的資訊年限遠超過其他學科。3、多為獨立工作，較少其他學科合作。4、搜尋文獻時較少委託他人，包括在使用圖書館時亦少使用參考諮詢，而是使用特定的書單搜尋二次資料，或利用特定的期刊等。5、最常使用的資料類型為圖書等（吳明德等，2006）。一般多認為，包括歷史學在內的人文學者對於資料庫的使用比較其他學科為低（Bates, M. J., 1996:514-523），臺灣亦有類似的看法。之所以會造成這樣的現象，主要的原因應該是史料的數位化進程較慢所致。

#### （一）臺灣歷史類史料資料庫發展的進程

由於國家文物典藏的數位化可以提升知識的累積、傳承與運用，成為知識經濟的基礎，先進國家陸續積極展開數位化工程。概略回顧臺灣歷史類數位資料庫的發展，可從1984年中央研究院的歷史語言研究所和計算中心合作，開始開發「廿五史食貨志全文資料庫」為始點。至1990年，中研院臺灣史研究所的前身臺灣史田野研究室開始規劃「臺灣史籍自動化」的長期工作，經史語所同意，並獲得計算中心的大力協助，比照二十五史資料庫，將製作臺灣方志資料庫列為第一階段的目標（詹素娟，1995）後陸續完成「臺灣方志資料庫」、「臺灣文獻叢刊」等，臺灣歷史類資料庫才開始較有系統的發展。

1990年代中葉以降，隨著公部門投注大量人力、物力進行數位典藏計畫，臺灣歷史類資料庫得以有著飛躍的成長。除中央研究院，包括國史館、國立故宮博物院、臺灣大學圖書館、國家圖書館、中央圖書館臺灣分館、國立臺中圖書館、臺灣省文獻委員會（現為國史館臺灣文獻館）、國立臺灣歷史博物館，以及等眾多學校、地方單位，陸續建置完成多種數位典藏資料庫。另並有民間企業或自行建置，或與公部門單位合作，陸續建置多個資料庫，如「臺灣時報」、「臺灣日日新報」、「漢文臺灣日日新報」等，相關資料庫可謂如雨後春筍般出現。

如何將現有的資訊，轉換成為得以檢索、進行關連解釋的應用領域，一直是資訊工程中重要的課題，當數位資訊技術被導入人文學科時亦是如此。當電子資源快速普及，使得資訊的來源更為多元，且可突破時間空間限制的限制、可快速、全面性的檢索，及可進行關連性觀察等優勢，隨著前述各項資料數位化工程的進展，進入了臺灣歷史史料領域，在研究的歷史上，臺灣史研究者可以透過各種資料庫，運用巨量的史料，這必然對於臺灣史研究有著全面性的影響。資訊學者將資訊學研究當中發展出來的科學方法，用在解決歷史學者的研究問題時，應該是不會一拍即合的（陳詩沛，2011：160）。而學者間又如何看待這波影響？

#### （二）樂觀的期待

1997年臺大公布「臺大電子圖書館與博物館」，就時間點而言，是臺灣歷史數位建置逐漸進入高峰期的階段。對此，三位來自不同領域的學者分別發表其願景。臺大計算機系統研究中心主任項潔指出，整個計畫結合臺大資訊工程系、歷史系、人類學系及圖書館系等不同的學門，將原住民史料、臺灣開拓史、以及臺灣文化遺產等重要史料的數位化資料與研究成果，公開給所有相關領域的研究學者，也提供一般民眾難得的歷史教材。透過有系統、大量且正確的提供管道，歷史文物得以呈顯真正的意義，科技在其中扮演的，則是傳送和永久保存的角色。負責史料整理的臺大歷史系教授吳密察則指出，研究歷史的過程中，需費時找尋資料，然許多珍貴脆弱的文件需要妥善保存，無法反覆提閱，學者只能以有限的資訊判讀史事，常常無法有適當的結論。電子圖書館與博物館計畫所做出來的資料庫，可以節省研究者查找資料的時間，而淡新檔案、岸裡大社文書及伊能文庫的珍貴資料首度全面性地公開，更提供給學者重要的研究資源。臺大人類學系教授胡家瑜則指出，完成平埔族和臺灣開拓史的資料數位化，接著再完成各個族群的原住民長期發展史，隨後再加上挖掘出土的古物，將有助於重建臺灣未開發前的狀態，以推敲古代人民的生活。<sup>10</sup>

<sup>10</sup> 彭慧明，1997/10/8。世界第一品牌 網路唱出臺灣之歌，聯合報，第45版。

當時他們三人的談話，表達了學者及建置者對於數位典藏的樂觀期望。未來臺灣的史學研究者不再需要浪費大量時間蒐尋資料，有更多時間投注在研究資料的解讀及查證考據等史學研究的核心工作，而且由於大家都能掌握同樣多元豐富的資訊，如何提出獨到正確的見解，便是在考驗學者的治學功夫，在良性競爭下，臺灣的史學研究會有更進一步的發展。<sup>11</sup>的確，以搜尋出史料即算完成研究的時代已經逐漸遠去，未來的文章若停留在資料豐富，顯然是不夠的。項潔並指出，歷史學家一如廚師，在資源匱乏的時代，找到好的食材就已領先他人一截；告別匱乏年代後，廚師要證明自己的本領，就必須以同樣素材，做出更加精緻而具有創意的菜色。未來，當資料取得越來越方便，駕馭史料、發掘問題的能力和眼光，將比以往任何時代都更形重要。

因為規模龐大的歷史數位資料，在資訊科技的介入下，幫助形成歷史研究的新手段（陳詩沛，2011）。而且，藉助電腦科技，也將處理以往的研究者無法處理的研究課題，以臺大「臺灣歷史數位圖書館」（THDL）中的土地契約為例，由於數位資料庫的建立，研究者可在彈指之間，將檢索所得的契約資料以年代、出處、作者、分類等範疇，加以排列組合。例如，該數位圖書館中約35,000件的契約中，透過電腦比對，總共發掘出了2,409對的上下手契，878群的鬮分契。換言之，透過程式設計，電腦找出了土地在不同人之間轉手、流動的過程。若再進一步搭配鬮分契與買賣契，幾個家族的地方開發史甚至隱然從中浮現（項潔、涂豐恩，2009。黃于鳴，2009）。面對這樣的結果，故也參與數位化計畫的中研院歷史語言研究所副所長林富士便稱：「有一場大革命正在悄悄進行！」，「因為數位科技將造成人類記錄經驗、傳遞訊息、溝通情感、學習知識、創造文化等方式的大變革，帶來煥然一新的文明。」<sup>12</sup>

### （三）「關鍵字」的疑慮

然而，在從事數位典藏工程的資訊學者樂觀的提出未來的願景時，人文學者對於數位典藏仍有相當程度的疑慮。根據吳明德等人的研究，人文學者雖多肯定資料

庫帶來檢索的方便性確有助於研究，但大部分學者卻擔憂資料庫內容的正確性，因一字之差，便有可能帶來研究的錯誤（吳明德，2006）。故部分資料庫在建置規範中，便將錯誤率目標設定為萬分之一，以圖控制品質。且眾多的古字、俗字及異體字等如何呈現，並是一直困擾中文資料庫建置者、且尚未能完全解決的問題，更有所謂「缺字必定層出不窮，加上造字管理不易」之嘆（王雅萍、謝筱琳，2011：27、30-33）。顯見就專業研究者的角度來觀察，資料庫的建置除了在系統、大量之外，如何保持正確性，更是重要的問題，即便前述的持樂觀態度的資訊學者項潔，也強調正確的基礎性。

另一份疑慮則在於一旦史料數位化後，會不會造成部分使用者及一些新進研究者，在對於背景未能掌握前，即以檢索取代閱讀，產生以管窺天的誤讀？陳弱水、劉增貴、王汎森、莊樹華、龐桂芬，甚至於前述的吳密察等學者，在肯定檢索所帶來的便利性之際，卻同時擔心這種檢索將成為一把雙面刃，因關鍵字詞搜索過於便利，使得研究者有可能不再細究關鍵字詞出現的脈絡，檔案文件之間的關聯性，及古籍中各式標題、段落設定之意義。且標點後的資料，對於紙本原有的版式、字體、序跋、批註等內容，常無法以原貌呈現，對於瞭解關鍵字詞在文章中的脈絡，可能帶來破壞。去脈絡化及過於便利的取得，讓歷史學者深感憂慮（吳明德、黃文琪、陳世娟，2006：7-10。吳密察，2009。項潔、翁稷安，2011）。

使用者會使用相關資料庫，其前提必在於對於這些資料庫的「信任」，相信資料庫內容的正確性，並且可提供與原件類似的脈絡。故若完全信任資料庫，一旦有因建檔中各式錯誤的情況產生，則可能必須承擔錯誤研究、甚至危及學術生涯的風險，將使研究者在信任資料庫之餘，反而像是走入一個充滿陷阱的地方而不自覺。

### 四、結語

在過往，由於典藏分散與提閱限制等因素，史料取得與閱讀較為不易，相對使研究者往往必須耗費大量時間、精力於收集資料上。但隨著數位典藏技術及公部門資料的導入，以中研院臺灣史籍自動化計畫為起點，超過二十年的時間內，臺灣歷史類資料庫蓬勃發展。後隨著更多民間參與者的加入，更多資料庫再被建置。各資料庫形式上或有差異，但眾多臺灣歷史資料在數位化、電腦化之後，可跨語言跨時空檢索，取

<sup>11</sup> 彭慧明，1997/11/2。文物上網路 研究更具臨場感，聯合報，第19版。

<sup>12</sup> 曹銘宗，2005/6/14。文化傳承 數位狂潮 e 把抓，聯合報，C6版。

得想要資料變得更加方便，且資訊科技導入人文研究後，以數位化方式陳列能讓使用者有效自由運用與廣泛的結合。無論是從事歷史研究的專業人士，甚至於對於歷史有興趣的一般民眾，皆可透過眾多的資料庫來方便的擷取所需的資料，並透過電腦快速的運算能力，建構史料間的關係，進而形成歷史研究的新手段。

然而，在享受使用方便，這些資料庫所提供正確性便是下一個階段所需驗證的問題。所謂「先求有、再求好」，現今的歷史研究者，固然不必煩惱缺乏資料庫的問題，卻必須開始正視資料庫正確性的問題。

要核對正確性，需比較數位資料與原件史料，現今所累積的數位資料甚多，已超越個別人力所能掌握，即便資訊技術已進步到可以利用程式系統自動判別詞句、進行分類歸納，然而對於資料庫內容的正確率檢視卻尚未發展出標準規範。故本篇僅能運用個人專業判斷，以臺灣近代歷史上數個重要事件的關鍵字，對數個資料庫進行隨機檢索，結果發現不少錯誤的現象。綜合以上檢索結果，可將這些錯誤大致歸為四類，包括：1. 缺漏字：如漏字未打及電腦轉碼出現無法顯示的現象等，影響研究使用。2. 後設資料位置錯置：如建檔人員在未瞭解史料結構下，錯放資料內容，或將作者張冠李戴，以及年代標示錯誤等種種問題。3. 內容判讀錯誤：多為錯別字。4. 未設權威檔。

對於處於科技時代的歷史學者來說，數位資源最大的前景不是它取代了使用原始材料的需求，而是它能在察覺定位檔案材料當下，即刻提供支援給研究者以利用（Duff, W., Craig, B., Cherry, J., 2004:22）。資料庫的建置及網路的應用使搜尋資料變得容易，但在方便之際，這些數位典藏資訊的正確性是否值得信賴，是否有需注意之處，是不可忽略的問題。若內容當中包含相當比例的錯誤，反而需要逐一檢視，造成研究的阻礙。且若面對當下檢索得來的資料，卻無法信任，仍須舟車勞頓往返資料原件所藏地逐一檢對，那不啻是耗費更多的時間與精力。科技的便利，仍須奠基於「正確」，否則便可能帶來前述的反效果。

前述四大類問題，看似僅是個別史料的小錯誤，然在單一資料庫內如果數量累積過多，使用者需花費較使用紙本更多的時間來比對，反造成不便，失去建置資料庫可帶來研究便利的本意。且就研究而言，在轉化的過程中應不可有任何錯誤，否則對研究者可能造成影響研究成果的正確性，甚至於學術生涯的進程。對於一般民眾而言，

錯誤的知識透過網路快速傳播與複製的功能，將造成歷史知識的混淆。在現今對數位典藏多半抱持正面肯定的聲浪之際，前述所舉的問題，對資料庫的使用而言則是一大警訊，若因這些錯誤造成研究者對數位資源的懷疑與不信任，則耗費鉅資投入建置的資料庫反成負擔。

本文用意並非全盤否定資料庫，而是提醒使用者小心謹慎，仍要反覆求證。資料浩瀚，在轉製的過程中不免出現錯誤，只是要如何解決與避免，將程度降到最低。對於已完成並開放使用的數位典藏資料庫，現可進行的補救措施，筆者之建議包括：1. 定期聘請學者進行檢視：聘請相關人文學者，對資料庫內容進行隨機式抽樣，並以及所檢閱出的範例，進行大規模搜尋比對。2. 標記或組合缺字：遇異體字便造字補充的方式，並非每一位使用者之電腦皆可所判讀，採取註記標示今正體字之樣式、純文字表示之組字，及構字方式組合標示其形，可使讀者瞭解該缺字的意涵。3. 回饋機制：利用使用者回饋機制進行修正，並定期公布修正成果。

使用者或機關單位建購資料庫時，也可就相關面向評估：1. 建置單位是否為研究信譽良好之單位。2. 資料建置過程是否有審查機制。3. 完成後的資料庫能否提供暢通的錯誤回報管道。但最終若能針對資料庫內容正確性建立一套完善的檢測機制，在求真與求便之間取得平衡，應是未來發展歷史資訊學時需進一步探討的課題。

## 參考資料

1. 中央研究院日治時期檔案整理小組，2002。臺灣總督府數位化檔案整理紀要。臺北：中央研究院臺灣史研究所籌備處。
2. 王梅玲、蔡明月、陳志銘、柯雲娥、蔡佳榮、陳勇汀、林怡甄，2008。臺灣圖書館史數位圖書館建構之研究，圖書館學與資訊科學，34卷1期：15-38。
3. 王雅萍、謝筱琳，2011。漢籍全文數位化工作流程指南。臺北：數位典藏與數位學習國家型科技計畫。
4. 吳明德、黃文琪、陳世娟，2006。人文學者使用中文古籍全文資料庫之研究，圖書資訊學刊，4卷1/2期：1-15。
5. 吳密察，1989。臺灣對話論。臺北：自立晚報。
6. 吳密察，2009。歷史文獻資料庫的建置與使用。發表於臺大數位典藏研究發展中心主辦，「數位典藏與數位人文國際研討會」。臺北：臺大數位典藏研究發展中心。
7. 李玉瑾編輯，2008。臺灣研究相關資料庫調查（97年度）。中和：國立中央圖書館臺灣分館。
8. 李光真，1994。歷史學者怎麼說？，光華雜誌，1994年1月期：101。
9. 李婉羽，2008。臺灣棒球文物數位化流程與數位典藏系統之規劃設計。淡水：淡江大學資訊與圖書館學系碩士班碩士論文。
10. 拉迪裏 (Le Roy Ladurie, Emmanuel)，楊豫，舒小昀，李霄翔譯，2002。歷史學家的思想和方法。上海：上海人民出版社。
11. 林宛頻，2006。傳統布袋戲數位典藏詮釋資料之研究。臺北：中國文化大學資訊管理研究所碩士論文。
12. 林信成、鄭國祥，2008。運動文化資產之數位典藏與增值利用：以臺灣棒球為例，圖書與資訊學刊，66期：14-34。
13. 林淑芬，2005。數位時代圖書館的報紙典藏—兼論國家圖書館報紙數位化，國家圖書館館刊，94年第1期：189-214。
14. 林淑慧，2008。十九世紀臺灣在地文人資料庫的建構與應用，臺灣圖書館管理季刊，4卷4期：46-63。
15. 林富士，2006。數位革命與歷史研究，國史館館刊，復刊41期：25-34。
16. 邱子恆、蔣以仁，2008。美國檔案知識庫發展經驗之研究，檔案季刊，7卷3期：22-43。
17. 學圖書資訊學研究所碩士論文。
18. 張文熙，2008。我國檔案知識庫建置之檢討，檔案季刊，7卷3期：44-57。
19. 張夢禪，2009。紅樓夢結構化之研究—以情愛事件為例。臺北：世新大學資訊傳播學系碩士論文。
20. 陳詩沛，2011。資訊技術與歷史文獻分析。臺北：國立臺灣大學資訊工程學研究所博士論文。
21. 陳麗玲，2008。臺灣地區地方文獻典藏數位化及運用--以國家圖書館為例，研考雙月刊，32卷3期：26-34。
22. 黃于鳴，2009。臺灣古地契關係自動重建之研究。臺北：國立臺灣大學資訊工程學研究所碩士論文。
23. 項潔、洪筱盈，2006。大學圖書館數位保存與館藏發展策略：以臺大圖書館為例，大學圖書館，10(2)：2-18。
24. 項潔、涂豐恩，2009/5/15。數位歷史如何可能—簡介臺大數位典藏研究發展中心之工作旨趣。「數位資料庫於臺灣文史研究中的應用」研討會，臺灣師範大學臺灣文化及語言文學研究所，國立臺灣師範大學綜合大樓503會議室。
25. 愛德華·卡耳(Edward. H. Carr)著，江政寬譯，2009。何謂歷史？。臺北：博雅書屋。
26. 詹素娟，1995。「臺灣方志資料庫」簡介，中央研究院計算中心通訊，11(6)。
27. 鄒擅銘，2008。國史館臺灣文獻館數位典藏之運用：一個實證的觀點。台中：東海大學公共事務碩士在職專班碩士論文。
28. 盧家慶，2008。臺灣古契書自動分類與依分類定義契書角色。臺北：國立臺灣大學資訊工程學研究所碩士論文。
29. 蘇倫伸，2008。日治時期日文臺灣文獻數位典藏計畫概述，臺灣圖書館管理季刊，4卷4期：75-81。
30. Dalton, M.S., Charnigo, L. 2004. Historians and their information sources. College and Research, Libraries, 65(5): 400-425.
31. Duff, W., Craig, B., Cherry, J., 2004. Historians' use of archival sources: promises and pitfalls of the digital age, Public Historian, 26(2): 7-22.
32. Jenn Riley, Michelle Dalmau, 2002. The IN Harmony project: Developing a flexible metadata model for the description and discovery of sheet music, The Electronic Library, 25(2): 132-147.

# Trust and Trap: Analysis on the Digital Archive for History of Taiwan

Lin, Chia-rong \*

## Abstract

The generalization of e-resource multiplied the source of information. Knowledge, which only used to be demonstrated on paper, can be expressed by multi-media with the assistance of technology nowadays. It possesses the features of unrestricted by time and space, high accessibility, easy to share and search. In the past, libraries emphasized on the preservation of substantial culture yet it gradually switched to the preservation of digital culture in accordance with the trend of age. Digital archive had become one of the key developments for each institute. Hence, the digitalization of cultural heritage has been put into practice like wildfire to every corner of the world with the supports from the governments.

The digitalization of cultural heritage provided abundant valuable digital literature and relics. For historians, digital archive also aroused new research topics and methods. Collecting and interpreting historical materials are the essences of historical study and the interpretation of historical materials determine the quality of study. The digital archive demonstrated after being organized as well as the accuracy and direction of interpretation during the process of digitalization play a crucial role on whether solid research foundation is provided for the researchers. Therefore, the study examined the quality of the digital archive for history of Taiwan through the actual operation methods and the results can be the reference for the future development of digital resource for history of Taiwan.

Keyword: digital archive, historical interpretation

---

\* The Research Assistant of the National Museum of Taiwan History.