

多重脈絡—— 數位檔案之問題與挑戰

項潔*、翁稷安**

摘要

數位化的檔案保存已經成為時代的趨勢，卻也面對質疑的聲音。其中最大的不安，在於原本紙本檔案保存的脈絡，是否會在數位化過程中遭到破壞。為了回應這樣的質問，本文回顧了傳統檔案學的保存原則，即依機構、依時間、依事件的整理，其目的便在於保存所謂檔案的既有脈絡，並指出這些原則本身所隱藏著的矛盾，如依事件的整理本身就會打破檔案依機構和時間所產生的原生脈絡，編纂者的後見主觀和選擇必然會影響檔案。更何況從研究的角度，人文研究本身便是要從史料既有脈絡中，重新找出新的解釋脈絡的過程。檔案數位化保存後，並不一定會破壞原本脈絡，恰恰相反，數位化後的檔案反而較紙本更具調整和變動的可能，產生紙本檔案所無法想像的多元脈絡。要實現這樣的可能，必須建立起一個功能強大的系統，此為檔案數位化的關鍵，數位檔案無法離開系統而單獨存在，系統的好壞將決定檔案的可用度。檔案檢索系統應盡量提供文件的各種脈絡及觀察脈絡的環境。

本文不只是被動的回應質疑，也試圖提出對數位化檔案保存的積極看法，即多重脈絡的挖掘。系統的建構不應再只是檔案數位化過程中的附屬物，而是依據檔案特性所設計的重要環節。一個好的系統所能提供的各種脈絡可能，分別為：（一）原始脈絡的保存；（二）重組原始脈絡產生的多重脈絡；（三）縱觀的脈絡；（四）鳥瞰型的脈絡；（五）文件間統計型的脈絡；（六）文件間隱藏的新脈絡。總結上述，經由檔案的數位化，檔案單一脈絡的限制被打破，檔案整理與檔案研究之間的矛盾隨之減輕；而透過檔案檢索系統的設計，檔案的豐富且多重的脈絡可被發現、被呈現、被觀察。因此，檔案檢索系統是未來在思考檔案數位化時不可或缺的關鍵方法。

* 國立臺灣大學資訊工程學系特聘教授，國立臺灣大學數位典藏研究發展中心主任。

** 國立臺灣大學數位典藏研究發展中心碩士後研究員。

Multiple-contextualization: Problems and Challenges on Digital Archives

Jieh Hsiang^{*}, Chi-an Weng^{**}

Abstract

Although preserving archives through digitization has become the trend of the era, it has also faced criticisms. The primary concern is that the original context of archive might be compromised during the digitization process. In response, this paper reviews the traditional principles of archiving: by institution, by timeline, and by event. Regardless of which one is adopted, its purpose is to preserve the original context of the historical documents. We point out that these principles might in fact be contradictory. For instance, re-organizing an archive according to events is bound to compromise the timeline structure or organization according to agencies. An archivist's personal view and preferences also inadvertently disrupts the original context. Furthermore, humanities research, historical studies in particular, largely involves the discovery of previously un-observed context from among primary documents. This paper argues that digitization will not only preserve the original context of the archive (if there is any), but also yield an enormous possibility of re-contextualization. To realize this possibility, however, it is imperative to design a powerful search system. This is the key to digital archives, since a digital archive can only be used and utilized through a system, not the naked eye. In addition to search and retrieve, such a system should also provide a environment that allows the forming of multiple contexts and the facility for their observation.

Therefore, the software system for a digital archive should not be an after-thought of the digitization procedure. Rather, it should be a key component designed to fully utilize the nature and characteristics of the documents of the archive. The kinds of contextualization that the system should provide include: (a) the original context; (b) re-ordering of the original context according to other metadata attributes; (c) horizontal context; (d) holistic context; (e) statistical context among the documents; (f) hidden (semantic-driven) context. In summary,

* Distinguished Professor, Department of Computer Science and Information Engineering (CSIE), National Taiwan University. Director of Research Center for Digital Humanities, National Taiwan University.

** Research Associate, Research Center for Digital Humanities, National Taiwan University.

through digitization, a digital archive breaks away from the limitation of the single context of the original (paper) archive and reconciles the tension between the archival requirement and the contextualization need of the researchers. Through a carefully designed system, the rich and multiple contexts implicitly existed in an archive can be discovered, displayed, observed, and explored.

一、前言：在傳統與創新之間

隨著資訊科技的進步與發展，數位化已成為影響各領域的時代趨勢，人文研究亦無法置身其外，數位與人文兩個研究領域的結合隱隱然成為蓄勢待發的學術範疇。數位人文所涉及的範圍十分廣闊，非三言兩語所能盡述。¹在人文研究和數位技術結合的過程中，檔案數位化的典藏保存可視為最基本的第一步；以此為基礎，數位和人文兩者才能擁有共同對話的基準點。事實上，單從檔案保存的角度，將檔案進行數位化整理，其必要已十分明顯，此外，透過網際網路的連接，也促進了檔案取得和使用上的便利。也因此，近年來，無論是國家政府機關或各學術研究單位，檔案資料的數位化工程都成為備受重視的首要之務。

然而，在這股熱潮中，也開始出現質疑的聲音，其中以數位化的形式保存檔案，在收藏上雖有著種種優勢，但是否會割裂切斷檔案既有的脈絡？換句話說，從研究的角度，檔案的數位化在帶來方便的同時，是否反而對史料的原始脈絡造成更深層的傷害？質疑者認為數位後的檔案資料，在性質上必定和紙本保存方式不同，重新數位化便等於徹底的重新編輯，當書本裡一頁頁的資料變成一個個的圖檔或文字檔，原本貫穿書本的編輯邏輯勢必亦將瓦解，失去了原本紙本編排的邏輯後，史料原本的意義脈絡也將面臨了崩壞的危機。

這些質問是檔案數位化所無法規避的問題，它們反映了兩個面向：一是反映了檔案學長期發展後所形成的研究傳統，一種對保障檔案真實原貌的堅持和使命；任何有可能對檔案原始狀態造成傷害的方法，都應該謹慎。在這樣的思路中，數位化很可能是包裹著糖衣的毒藥，保存或搜索的便利很容易讓研究者走入見樹不見林的陷阱裡。另一方面，這些懷疑也反映對數位化最根本的不安全感，數位化不同於紙本，並不是使用者能親手觸摸的實際存在，一旦紙本編成的史料被沒有實物的資訊技術取代，過去的閱讀習慣必將有所改變，這會不會只是造成數位化的「斷爛朝報」？尤有甚者，以紙本方式存在的斷爛朝報人們還有手動整理的可能，而數位化似乎只能依靠資訊技術人員一途了。

在本文接下來的討論中，將試圖證明這樣的恐懼是出於對數位化技術的表面認知，以及面對紙本閱讀習慣改變時的多慮。檔案的數位化並不會違背檔案學長久以來形成的學術規範，正好相反，通過數位化的重新整理，反而可以糾正紙本編輯所造成的「編輯者的盲點」，還原史料原本的脈絡，貼近使用者的研究習慣和需要。此外，檔案數位化不但不會造成斷爛朝報的現象，透過資訊技術和工具的研發，反而可以透過建立數位資料的觀察系統，讓研究者可以更自在的去觀察史料，更自由

1 相關討論可參見項潔、涂豐恩，2011，〈什麼是數位人文〉，《從保存到創造：數位人文研究的發端》，頁9-28。

的進行分析與論述；換句話說，數位化並建立一套可供使用者有效觀察的系統，可使檔案的潛在價值得到完全的解放。數位技術和人文研究應是相輔相成而非相互抵觸的概念，這也才是檔案數位化工作真正的意義。

創新的技術必然會面臨傳統的質疑，但並不表示創新和傳統兩者之間為不可跨越的對立，事實上，所有的創新都一定要去回應傳統，試著和其對話，不斷尋找共識的建立，如此才能讓創新的技術有更加深化的可能，而這將是本文的主要目的。本文將分成兩部分，首先對傳統的檔案文獻保存方式進行概括式的描述，指出其保存的邏輯和重點所在；然後進一步討論傳統檔案保存和檔案研究之間的關係。第二部分，則由數位時代的保存方式出發，指出它和過往的不同之處，以及這樣的特殊性對檔案研究所能帶來的新衝擊與貢獻，尤其在打造一個適合於研究的系統後，將會替數位時代的檔案保存勾勒出新脈絡的可能。

二、傳統檔案保存的脈絡和研究方式

（一）傳統檔案保存的方式及其原則

讓我們試著藉由下面兩位學者的討論為例，來呈現華人世界對檔案保存的看法。首先是大陸前輩學者秦國經，他於1980年代起任職中國第一歷史檔案館，長期鑽研明清檔案學，撰寫過許多相關文章，並於2004年出版了厚達近900頁的鉅著《明清檔案學》，其內容仔細介紹了明清所留下的相關檔案，從檔案的來源、分類、收藏、管理和史料價值等等都有著詳盡的介紹。在該書關於檔案整理的篇章中，他指出明清檔案應按照下面步驟進行整理：首先是區分全宗，即按行政機構歸屬進行約略的區隔。其次則是就全宗內的檔案加以分類，可以依照文件名稱（例如將內閣檔案分詔書、題本、奏啟等類）、機構（如將宗人府檔案分為經歷司、左司、右司、黃檔房、銀庫、玉牒館）、所涉及問題（如將外務部檔案按所涉國家區分）和年代（如會議政務處檔案從光緒27年依時序分至宣統3年）等四種方式加以細分。最後，則是加以組卷和編目的工作。行文中作者不斷強調「文件之間具有緊密的聯繫，不能任意加以分割」，所以要努力保持機關檔案的聯繫和完整；即如按類劃分細目時，亦特別強調「類目的設置應以保持文件的歷史聯繫和便於利用為原則」，其理想是希望「類目要如實反映出檔案的內容」。²凡此，皆可見得其對「如實保存」，尤其是按歷史原貌的「如實保存」的堅持。

這樣的堅持也成為中國第一檔案館等相關機構，在處理檔案文件的基本流程。事實上，秦國經所服務的中國一檔館作為一個積極的保存者，除了依文件出處和時

2 參見秦國經，2005，《明清檔案學》，頁761-769。

間的整理外，更一直在進行依事件分類編輯、出版檔案的工作，諸如《明清宮藏中西商貿檔案》、《明清宮藏臺灣檔案集編》、《中琉歷史關係檔案》等，都是近年該單位出版的重要史料集成。

另一位學者則是政治大學圖書資訊與檔案學研究所的薛理桂，薛教授從事相關教研工作多年，對檔案管理的發展，以及國內各機關的檔案收藏了解極為深入；並出版了《檔案學理論》、《檔案學導論》等書，可說是臺灣檔案學領域中，影響最大的參考書籍。他指出，檔案管理人員較喜歡用「描述」(description)與「編排」(arrangement)，取代圖書館界慣用的「分類」(classification)，此乃因為檔案的數量和內容，都無法和一般書籍等量／等質齊觀，通常以「文件系列」(series)為建檔的基本單位，即指在全宗之下的構成單位，由同一業務所產生的一組文件。這組文件具有類似的形式、排列順序或主題性質。文件系列可依其性質再細分為若干個「副系列」(subseries)。換句話說，是根據全宗原則的整理。他將對檔案的處理歸納為三點原則：一是尊重檔案的來源原則，按檔案產生的出處，維持其完整，其預設為「檔案產生係隨著機構與個人的業務、生長與活動，呈有機的成長」，不當輕易加以割裂，這樣的處置將會使史學研究者更加了解檔案的來源與目的。第二個原則是尊重檔案的全宗原則，即是來自於同一個行政單位、公司或家庭視為一個全宗，應當組合在一起。第三個原則是尊重檔案的原始順序，此原則強調經由檔案的原始順序，可反應原有產生與使用情況，意即當檔案為現行文書時，有其出於業務需要而生成的特質，如果強迫使用人為的方式，會對檔案的歷史證據價值帶來傷害，所以原始順序對檔案的處理至關重要，不當任意予以改變。³

薛理桂的著作充分闡釋了19世紀至20世紀以來檔案學發展的風貌，更可說替檔案學作為一專業學門劃下了明確的學科界線；而他綜合西方各家研究所歸納出的三原則，具體而微地反映了檔案學對文獻保存的理念與態度。秦國經則從長期實作中，順應著傳統既存的檔案結構，歸納出檔案保存應有的準則，依名稱、機構、年代和所涉問題的分類，概括檔案整理多數的需要。兩者可謂殊途同歸。綜合兩位學者的論述，我們可以將檔案學的檔案脈絡大致整理成下列三點：一是以檔案出處為中心的保存脈絡，這可說是一種最安全而妥當的保存方式，作為一份文書資料在仍是現行文書時，本身便有其產生的需要與邏輯，顯示了該機構運作的各個面向，因此檔案的生成機構可說構成了該檔案的天然疆界，以此為範疇，完好保留其運作的規律，提供了日後研究者在觀察和切入時的重要參考座標。第二種整理脈絡則是以時間的順序作為檔案典藏的脈絡，這是最直覺的區隔方式，檔案作為一種史料，蘊藏著進入或理解歷史的鎖鑰，而人類理解歷史最直接的方式便是順著時序去解讀，這也是多數歷史研究者所習慣的方式。以時間先後保存，理論上將能體現事件發生

3 薛理桂，1998，《檔案學導論》(初版)，頁181-184。

的先後順序，可說是檔案資料的天然肌理。第三種脈絡則是前述兩者的進階，即依據事件的相關度進行整理，以某個議題或歷史事件為核心，將相關檔案史料打破各種原先的限制，重新加以編排。其優勢在於用事件的角度切入，能將散見於各處的資料集中化，賦與零碎的檔案一個初步的整理，給予一個粗略的「故事軸」，以利日後研究者在觀察時的需要。

總而言之，上述這三種方式符合了研究者最直觀的考量，亦反應著構成歷史的基本要素，機關脈絡代表著和檔案相關人、物的保存，時間脈絡則代表著變化，事件脈絡則代表著將檔案串連的故事，這些都是當我們在談論過去，或從事人文研究時所不能或缺的因子。我們或許可以說，檔案學的三種檔案脈絡，其實便是為日後人文研究進行基本清理、編排的預先準備。

（二）傳統檔案脈絡和研究之間的關係

上述的討論其實已提示了檔案學的保存邏輯，是和歷史研究的發展相契合的，並出於進一步歷史研究的需要所建立起的保存形式。事實上檔案學於19世紀末開始萌芽發展之際，正是和近代史學的生成——蘭克典範（the Rankean paradigm）相輔相成的。19世紀德國史家利奧波德·馮·蘭克（Leopold von Ranke, 1795-1886），反對當時流行的浪漫主義（Romanticism）或歷史主義（Historicism），強調歷史學應該是可以客觀、科學的學門，不受制於當代政治與任何哲學系統，亦不為預知未來，而是要探求真實發生的過去。這樣的信念帶來了深遠的影響，成為近代歷史學的主要骨幹。要建立科學的史學，在蘭克看來最重要的關鍵便是對原始史料的直接引用與批判，尤其是對一手檔案的使用，這成為後來蘭克典範的核心之一。⁴事實上，將文獻學的方法引用至近代歷史學研究之中，可被視為蘭克最大的貢獻。鑑定史料的真偽，仰賴「最純正、最直接的文獻」，全面的對相關史料進行檢索，直探檔案，成為近代史學的基本程序；即便在今日，這項重要的突破，仍可被視為歷史學研究和教學的根基。⁵

曾赴德國求學的傅斯年，也受到蘭克史學思想的洗禮，在其於1928年發表的〈歷史語言研究所工作之旨趣〉一文中便直接提出了「近代的歷史學只是史料學」的說法；史家的工作在於使用自然科學的方法，將一切所能觸及的史料進行整理。這種以科學治史的態度，和盡量保存檔案完整脈絡，使其不受後人編排所污染的態度，可說遙相呼應。當然這樣的價值觀，似乎隱隱暗示著在作為史料的檔案中有著

4 此段關於蘭克本人史學著述和思想的大致內容，整理自汪榮祖，〈回顧近代史學之父蘭克的史學〉，收於氏著，2002，《史學九章》，臺北：麥田，頁49-70。這並非說蘭克史學只是一味的主張實證、客觀而已，關於這部分在下文將會提及。這中間涉及許多複雜討論，已為史學史或史學方法領域的重要課題，非本文所能詳述。

5 Richard J. Evans 著，潘派泰譯，2002，《為史學辯護》，頁22-28。

「不證自明」的過去藏於其中，史家的工作只是讓檔案史料自己說話。上述蘭克或傅斯年的觀點，有其發展的時代背景，在成為史學研究典範的同時，也必然面對了許多質疑的聲浪，使得近代史學對其典範產生了深刻的反省，對蘭克的論述本身也有了重新認識的機會。事實上，長於檔案使用的蘭克從未認為「史學即史料學」，而是主張「歷史要寫得像過去已發生的事一樣真實」，前者是後者層層錯譯之後的結果。他所謂的客觀、科學的史學，絕不只是檔案的堆砌排比，史料仍需要靠史家主觀的選擇，在不違背對個別史實的客觀研究中，從整體的潮流加以掌握。⁶

總而言之，無論蘭克或傅斯年的主張都有其時代的針對性，不能僅從字面以僵化的教條方式看待，但無論如何，他們對史料求真的尊重是任何歷史研究者都該遵守的基本守則。這樣的原則一樣適用於所謂「後現代情境」下的史學研究，Richard J. Evans在他著名的《為史學辯護》一書中，便指出後現代史學也許可以從理論面對史學傳統帶來一些思考的刺激，但終究只是理論，史學的方法是不會改變的，也就是「由蘭克所設下，並在他之後被人們以多種方式發揚光大的那些鑑定史料的法則」。無論是哪一派別的史學研究者，他們大部分的心力都投注在「確定那些個別事實，重建那些個別事實，根據史料去研究，越穩固越好」這是連後現代史學最熱情的擁戴者都必須遵守的金科玉律。⁷

歷史研究本身便是在一個個不同的檔案和史料之間，找尋新脈絡，藉此進行結構性分析或訴說被時光淹滅的故事。檔案本身並不會說話，而是如同一個個散落的接點，需要靠史家的目光和關懷加以串連編排，並揭示其意義。檔案自身的脈絡是史家的參考座標，是研究的起點而非結束；史學研究所從事的，便是從檔案的既有脈絡中，在不違反歷史事實的原則下，以新的視角建立新的脈絡，挖掘出過去所尚未被世人所知悉的面向。

近代史學對史料的重視，和檔案學對檔案完整的重視是無法分開的；理論上，史家的工作便是在檔案學家所清理好的戰場上努力奮戰。但這並不表示尊重檔案真實的原則，就該被既有結構所僵化和限制，所謂的研究本身即是在尋找、建造出新脈絡的過程，一個好的研究者應該尊重檔案的既有架構，又不當受其限制，如是才能開創出具有新意的研究。此外，紙本的檔案存放原則本身並非沒有矛盾，我們將在下節中進行討論，並進一步指出，數位化正是化解這種矛盾的解答，且數位化的檔案處理將會更貼近研究者的思考方式和需求，不只是清空戰場，更宛如提供了無堅不摧的神兵利器。

6 參見汪榮祖，2002，〈回顧近代史學之父蘭克的史學〉，《史學九章》，頁62。

7 參見Richard J. Evans著，潘派泰譯，2002，《為史學辯護》，頁144-146。

三、傳統檔案保存的盲點與問題

(一) 何謂檔案

在上節中，我們大致將傳統檔案的保存歸納成依形成單位、依時間、依事件的三個脈絡，並指出在尊重傳統檔案真實的原則下，以研究者的問題意識，重新勾勒出檔案彼此間新的脈絡是所有歷史研究必經程序。接下來我們必須進一步回答，以數位的方式保存檔案，在這樣的既有關係中，能夠帶來什麼新的可能？我們將先指出傳統檔案保存的不足之處，以此為基礎突顯出檔案數位化的優勢，並進一步說明以數位化的方式整理、留存檔案所具有的新潛力。

在開始進行討論之前，我們或許須先思考檔案究竟是什麼？薛理桂在1998年版的《檔案學導論》裡，在綜合各家對檔案的定義後給予「何謂檔案？」的問題一個狹義的答案，即：

政府單位在處理公務的過程中所產生的文書，在歷經一段時間後，這些文書已不再使用，經鑑定其具有長期保存價值後，予以保存，以供使用。

他指出在狹義的定義中，包含著產生單位（政府）、時效（非現行文書）、價值（具長期保存的價值）等三個要素；廣義的定義，則是不限政府單位，私人機構、家庭或個人所產生的文書，都可以視為檔案之一種。⁸他更詳細歸納了檔案的文種特性，分別是：唯一性、多元性、公平性、可靠性、自然性和關聯性。⁹我們可以再將這六種特質，粗分成三方面：一是對檔案真實性的信任，特別它是機關在運作過程中自然生成的產物，除少數作假的特例情形，檔案都應該是最真實、可靠的稀有史料。其次則是檔案種類的多元性，尤其在今日，從紙本式的手稿、公文書，到錄音帶、錄影帶、照片、地圖、工程圖、微縮片、電腦磁片、光碟等，檔案的種類可說無所不包。最後則是檔案來源的多源性，因為政府機關內的業務常會有跨部門的情況，有時除了和同機構內不同的單位關聯，甚至還可能會與機構外其他單位結合，形成一份檔案文書多種不同來源的情形。

這樣的定義其實是取其最安全的方式，即以機關為主軸的方式界定檔案，再按時間的方式進行排列，此種作法可以說是最沒有爭議的作法，但仍有其問題存在，首先就如同上段所說的，等於是種「無作為的作為」，一旦數量龐大超過一個限度，人們將無法有信心能夠詳閱其中全部的資料，造成研究者將大部分時間耗費在史料搜集、考證，使得所當關注的研究焦點無法顯現。這也就是前面在討論秦國

8 參見薛理桂，《檔案學導論》，頁3-4。

9 參見薛理桂，《檔案學導論》，頁6-7。

經的檔案脈絡時，作為積極的檔案保存者，會試圖用「事件」去對檔案進行區分的原因。檔案必須出版才能為研究者廣泛的使用，然而一旦經歷出版或編輯的程序，特別再加入「事件」為參考座標後，便很可能（甚或必然）改變了所謂檔案的原有脈絡，無論來源、全宗或原始順序都面臨重新的編排與調整，這便形成了理論和實作間的矛盾。其次則是檔案多元化和多源化的衝擊，各種類型和各種來源的檔案，互相雜混，特別在以「事件」為核心的檔案整理過程中，這樣的問題更加激化。凡此，皆形成了傳統紙本檔案保存方式的盲點所在。

（二）紙本的檔案保存方式的盲點

人們常常不自覺地預設紙本保存方式、尤其是以印刷出版方式的客觀性。這有其歷史脈絡，印刷術的出現對人類的歷史帶來巨大的影響，而對書本的尊重更是自古已然，無論中外。這種對書本的尊重和習慣，會使我們忽略了將檔案以書本的方式編輯出版，本身便是一個重新整理的過程，它對檔案既有脈絡所帶來的殺傷力，並不會小於數位化過程，甚至是更為巨大的。紙本對檔案的整理、出版，不出依機構、時間和事件三種方式，並以依事件分類者為大宗，因為依單位和時間整理，雖保存檔案的原貌，卻無法帶來研究上的便利；一頁頁的翻閱，或能使研究者心安，卻不見得能轉換成研究效率和質量的提升，也因此依事件整理成為最受檔案整理者重視的分類型態。

依事件積極對檔案進行整理的方式，並非今日才出現的思維，在檔案形成之初便有類似的認知。以清代的檔案管理為例，清代所謂的「專案檔」，是有深遠的歷史傳統的。專案檔指的是清代軍機處為了便於查考舊案，依例會將定時經辦的事務按照內容另行抄錄，加以儲存。精研故宮檔案多年的莊吉發，將軍機處的專案檔定義為「是以事為綱，逐日抄繕成冊，其每一種檔冊，僅關一類之事，並不雜載」。他將目前國立故宮博物院現藏的清代專案檔約略分為三類：「一類是清廷用兵隣國整理邊界的檔冊，如緬檔、安南檔、廓爾喀檔等；一類是剿辦秘密宗教及髮捻等內地民變的檔冊，如東案檔、東案口供檔、林案供詞檔、剿捕檔等；一類是平定少數民族之亂的檔冊，如金川檔、苗匪檔、剿滅逆番檔、剿捕逆回檔等。」¹⁰換句話說，即便在檔案最原初的生成階段，將檔案從機關等原生脈絡中抽出，改以事件為主軸的脈絡，已是檔案彙整的常態現象。這反映了檔案必然的現實，如果不對原生狀態進行加工、「後製」，要對其進行縱向（檔案自身的前後順序性）或橫向（跨機關單位的相關性）串連，都是件十分困難的事，反而無法呈現檔案內所記錄之事件的全貌。

這樣的整理傳統至今已演變成一種專門的學問，即所謂「檔案文獻編纂學」，徐紹敏於2001年所再版的《檔案文獻編纂學》一書中詳盡列出了檔案編作工作的六

10 莊吉發，1983，《故宮檔案述要》，頁237。

個步驟，分別是：（1）確定編纂題目；（2）編寫選材大綱；（3）查找和挑選檔案；（4）正文的加工和編排；（5）編寫彙編內參考材料；（6）彙編的校審和出版。¹¹事實上該書便將每個細節加以說明，從如何選題、編前研究、查找和挑選檔案、內容考訂、到編排的體例等，無不鉅細靡遺地加以介紹。這樣詳細的說明當然顯示了編排檔案時的負責態度，但同樣地，這麼小心翼翼、如履薄冰的態度，也正反應了檔案重新編排時的高度脆弱性和危險性。因為依事件關係替研究者對檔案進行梳理，不得不面對檔案編纂和分類的主觀危險。所謂的史料編輯本身，便是一種具主觀的選擇，最嚴重的情況，還是刻意為特定目的或動機服務的選擇。《檔案文獻編纂學》一書便反覆強調進行彙編的政治性動機和意義。這種為特殊目的編纂方式所帶來的影響，還比較容易避免，只要具有相當警覺心的研究者都應能迴避，更麻煩的則是無心的主觀流入，其動機出於真正對學術需要的整理，但卻在無意中，將編纂者自己理解過程的預設不自覺地帶入檔案，以自己的架構替代了檔案的原始架構。一旦整理完畢編印成冊後，檔案史料的解釋空間很可能便被凝滯於某一點上，窄化了解釋的可能，甚至形成一種誤導。以下即以戴炎輝先生整理的《淡新檔案》為例，說明紙本編輯者分類整理後對檔案所造成的影響。

《淡新檔案》是指清朝乾隆41年（1776）至光緒21年（1895）間淡水廳、臺北府及新竹縣的行政與司法檔案，該批檔案於日本時代由新竹地方法院承接，轉送覆審法院（即高等法院），最後再轉贈臺北帝國大學文政學部，供學術研究之用。日本時期因為該批資料主要是清朝治臺之相關檔案，故命名為《臺灣文書》。戰後移交臺灣大學法學院，由法律系戴炎輝教授主持整理工作，以該批資料所涵蓋之地域改命名為《淡新檔案》。戴炎輝對該檔案的整理功不可沒，在面對如此龐大、無章法的資料，他進行了一系列基於近代法律學架構的整理，將之分成行政、民事、刑事三編，其下再細分成16類，類以下再區隔成102款。¹²這一巨大的工程，對臺灣史的研究帶來深遠的影響。

戴炎輝的分類方式，確實提供了理解《淡新檔案》時切入的便利性，但卻無法除去以「後見之明」的方式解讀史料的必然問題，他引用西方現代法學觀念詮釋清朝臺灣的衙門運作，其侷限十分明顯。諸如分類中所見的「雜事款」（分別出現在行政編總務類、行政編財政類）、「其他款」（出現在刑事編總務類）的分類細目，不明確的題名，正顯示出現代法學觀念無法完全包涵《淡新檔案》的內容。原本依照吏、戶、禮、兵、刑、工的傳統六房分類的《淡新檔案》，其原初脈絡在戴炎輝的分類中喪失，如學者林文凱所指出的「戴炎輝這一新的分類與命名整理，無法完整

11 徐紹敏主編，2001，《檔案文獻編纂學》（第二版），頁8-9。

12 關於《淡新檔案》的詳細資料與介紹，可參見國立臺灣大學圖書館為該檔案所建設的相關網頁，內容十分詳盡：http://www.lib.ntu.edu.tw/cg/events/TH/tanhsin_page_01.html。上列關於《淡新檔案》的簡介即為整理網頁資料而得。

呈現出一個行政與訴訟事務在地方官府的整體統治中原有的體系性」。尤有甚者，這樣的分類方式也易造成研究者在使用《淡新檔案》時的盲點，諸如滋賀秀三、黃宗智與艾馬克等學者，在進行有關淡新檔案的土地法律文化研究時，就都僅以民事門的土地控案為分析，不自覺的接受戴氏分類法中三權分立的體制，忽視了當時土地政策、土地行政與土地訴訟間的真正關聯，難以清楚詮釋出清代土地的法律文化。¹³

當然，需要特別強調，對《淡新檔案》進行重新分類，很可能是戴炎輝不得不然的選擇，因為當戴炎輝挖掘出這批史料之時，保存狀態極差，原有的分類應早無留存。¹⁴選擇以西洋法的分類進行處理，是在別無他法下的選擇。《淡新檔案》的例子，說明了紙本的檔案保存絕非萬無一失，編輯者的主觀選擇，會直接影響到檔案的整理與出版，更會間接的影響到以後的研究取徑。這形成了一個積極檔案保存者的兩難，一是選擇最安全的方式，以所出機關或時間的方式進行整理，但這樣卻失去了使用者在運用該檔案的便利性；另一則是積極的以事件或較現代的觀點，用使用者比較好理解和操作的方式進行編排，但卻又不得不冒著污染檔案的風險。

然而，這樣的矛盾正突顯了檔案史料數位化的優勢，也說明了數位化的必要。在下文中將進一步說明，紙本檔案本身所存有的脈絡困境，在進入數位化的時代，在看似加劇的同時，卻正提供了解答的良方；此外數位化後的檔案在與一強大的系統搭配後，將為檔案的脈絡帶來新的觀察與研究可能。

四、數位時代檔案研究的新可能

在說明紙本檔案亦有盲點之後，在本節中將指出檔案數位化後，將會帶來新的檔案編排和保存的型態，這種新的型態不僅可以充分回應對數位化檔案脈絡的質疑，並能從更寬廣也更關鍵的角度，體現數位化的方式對檔案進行處理在研究上的優勢。種種質疑之聲，有些是承繼紙本的缺陷而來，有些則是因數位保存的特性而產生的，在檔案數位化後，前者得到顯著的改善，後者在經過強有力系統的處理後，則變成了研究者在數位時代的優勢；在下文中將指出，數位化的檔案可以更進一步解決紙本的缺陷，並完成紙本檔案所無法完成的功能，為人文研究帶來全新的視野。

13 林文凱，2004年12月，〈清代晚期臺灣的土地法律文化——淡新檔案內淡新地域漢墾莊在十九世紀四十年代以來土地抗租案件為主的分析〉，發表於臺灣社會學會等主辦「2004 臺灣社會學會年會暨『走過臺灣——世代、歷史、與社會』研討會」。

14 據戴炎輝孫子戴桓青的回憶：「光復之後臺大法學院圖書館的陰暗角落裡有一批殘破的清代檔案，原是臺北帝大的收藏，因蟲蛀水浸無人問津幾乎變成廢紙，祖父看到了卻如獲至寶。……在還沒有影印機與電腦的時代，祖父傾10年之力修補整理……許多蟲蝕水浸闕漏之處，是祖父參酌考證之後再加以褙裱補寫才維持完整。」可以想見當年在處理這批檔案時的艱鉅。見戴桓青，2008，〈四代百年臺大情〉，《臺大校友雙月刊》，第60期，頁24-25。

檔案數位化的優勢不僅只是儲存方式的不同，更重要的是，數位化檔案優異之處在於它是和「系統」相結合的，一套功能強大而完備的資料庫系統，除了能彌補過往檔案處理的不足，並能為人文研究帶來新的觀察面向與研究功能。

（一）數位時代檔案的收集與整理的新樣貌——再論「何謂檔案」

數位化檔案給予研究者最大的不信任感，如前已提及的，一是集中在是否會破壞檔案的原始脈絡上，在前文中我們試著說明所謂「檔案的原始脈絡」本身便是一項迷思，即使是在紙本的年代，為了增加檔案研究的便利，讓使用者不會迷失在史料的浩瀚無垠之中，將檔案從原生的機關、時間脈絡中抽離，一直是檔案使用的常態。之所以為常態，正因為這是符合使用者需要的最佳方式，然而問題便在於要進行這樣的整理工程，需要花費大量的時間、人力和物力，而且經常會有掛一漏萬的風險。二則是檔案從紙本經過數位處理之後本身性質的改變，其新的特性，會讓習於紙本的使用者感到不安，因為一旦數位化，檔案變成了看不見的位元，不再令使用者能親身接觸。以下我們試著以檔案的多元／源化和檔案的原子化兩者為例，¹⁵加以論證，希期去除掉這類直覺式的恐慌。這些數位化後新的特性，經過精細的系統的處理後，不但不會造成盲點，反而會成為新的研究利基，創造出一種全新的檔案使用邏輯和視野。

首先是檔案的多元／源化，如前述這是在紙本時代即有的難題，過往在思考檔案時往往由印刷的角度去思考，非文字的資料如影音等視聽資料，除非經文字化處理，否則必難以同時保存，進入數位時代後，這樣的問題便不再存在。¹⁶多源化則是指檔案來源的不同，這可說是傳統檔案形式所有的形式之延伸，不單只是指跨機關的檔案整理，更涉及了許多原本就不具連貫性的資料文件，如「臺灣歷史數位圖書館（THDL）」¹⁷所藏的古契書就是最明顯的例子，該資料庫收集了從清代到日本時期臺灣地區的土地契約文件，其內容多半為古地契，記載著某塊土地地權的轉移或交易情況，其總量截至目前為止共有39,255件，¹⁸每月仍不斷增加中，其來源共有

15 這並不表示說檔案的數位化所面臨的新問題，僅多元／源化和檔案的原子化這兩點而已，除了這兩點之外，檔案內容的「質」的問題，亦即其數位內容和後設資料的正確度，仍是史學研究者在使用上，所感到擔心的。莊樹華在一次研討會中就指出，這種「質」的殘缺，往往比檔案「量」的不齊，帶來更大的風險。莊樹華，〈數位檔案運用的契機與風險——以近代中國外交檔案為例〉，發表於國立政治大學圖書資訊與檔案學研究所主辦，「數位檔案加值與教學應用研討會」（2010年5月27日）。此外，還有檔案Silos化的問題，當數位化如星火燎原般四處擴散時，如何不會各自為政，乃至重複，也成為檔案數位化後的新難題。但這些問題都比較偏向於技術性和整合性，和本文從研究的角度提出的質問略有不同，故無法於此處進一步加以細論。

16 以目前國史館資料庫中「元首文物」的檔案分類而言，蔣中正、蔣經國、李登輝等總統文物，其類型便包含了文件、底片、照片、視聽，同時存在一個資料庫中，供使用者觀察。

17 臺灣歷史數位圖書館（THDL）網址為：<http://thdl.ntu.edu.tw/>。

18 2012年10月12日。

100多個，除了已出版成冊的書籍外，還有許多原件資料，其數量和延伸性皆非過去紙本時代所能想像。

檔案原子化的問題，在前文已稍涉及，關鍵在於數位化的過程所給予人們的想像，和紙本整理的過程正好相反。過往紙本的整理過程是將一份份分散的檔案，透過抄錄、打字彙集成書，檔案的數位化則是將散落或已編輯成冊的檔案，再一份份散解轉換成數位的檔案；中間更涉及了由實體變成虛擬的過程，加深了人們心中主觀的不信任感。如何在原子化的數位檔案之間，「回復」、「保有」檔案的既有脈絡就成為多數人在思考檔案數位化的焦點。¹⁹此外，數位檔案的原子化所涉及的另一個層面，對大部分人來說，數位檔案資料庫的使用，多半僅是針對全文或檔案描述進行關鍵字檢索，很容易形成「見林不見樹」的情況。²⁰總之，各種類型和複數來源的組合，造成數位化檔案不同於過往的集合型態；由實體轉變成虛擬的過程，也使得檔案數位化的保存予人更形分離的感覺，這種疏離感使人們恐懼無法掌握資料的全貌。這兩種思考的結合，造成了對數位化檔案的不信任。

然而，問題便在於這些對數位資料的擔心，其實是基於既有研究習慣後刻板印象式的恐慌，反而忽略這些憂慮所代表的，乃是一手檔案史料的本質，之前之所以無此困擾，並不是因為這些問題不存在，而是因為紙本印刷的編輯形式，掩蓋甚至割裂了檔案資料的原有特質與風貌，一旦檔案在應用上的可能性被壓縮，自然就不會存有風險，而檔案一旦經由數位化之後，從紙本的限制中獲得解放，無限的可能被釋放，重新回復其本色，是以那些對於數位化檔案的擔憂往往正是其長處所在。譬如多元／源化的問題，這本來就是一手史料的特性，多元化在史學研究多角發展的今日，諸如影視史學、圖像歷史早已成為新的研究重點，研究者原本就不應該自囿於文字的資料，反過來說如何結合文字和影像進行觀察才是困擾研究者的癥結所在，其起因便在於在紙本世界中，文字的邏輯和媒介都難以將影像納入，這樣的困難在數位世界中根本無法立足，因為文字和多媒體的互動，原本就是數位的常態。史料的多源化與其說是研究者的惡夢，或說對檔案脈絡的傷害，正好相反，史料的多源化反而應該是研究者的美夢成真，傅斯年對史學研究者查找資料理應「上窮碧

19 史學研究者洪麗完便曾強調「為了保持檔案資料的完整脈絡，入藏時不能隨意拆散（整批入藏），應注意資料本身內在脈絡」；作為一個專業的臺灣史研究者，她認為對龐大古文書的整理「社會、經濟、文化與政治、宗教」等分類，會較以單件契字的性質有意義。洪麗完，2007，〈評論稿〉，收於逢甲大學歷史與文物管理研究所、台灣古文書學會編校，《第三屆台灣古文書與歷史研究學術研討會論文集》，臺中：逢甲大學出版社，頁331。

20 莊樹華和龐桂芬便為文指出：「檔案的描述雖有助於使用者在大量的資料群中快速擷取資料，但過度依賴關鍵字詞所取得的資訊，往往會切斷了檔案彼此之間的關聯性，以致隱藏於關鍵字詞背後的概念形成，及思考過程之間比較模糊的地帶，恐怕都被湮滅。數位化的便利，對史學研究者是節省了資料收集的時間，可以投入更多的分析與思考，還是讓研究者迷失於資料取得的便捷與豐富，而忽略了歷史研究的深層思考，是值得深思的問題。」具體說明了關鍵字檢索對研究所可能造成的盲點。莊樹華、龐桂芬，2006，〈數位化時代下的檔案館營運——以中研院近史所檔案館為例〉，《檔案季刊》，第5卷4期，頁23。

落下黃泉」的經典名言，其所描繪的理想便是要從各種方面收集史料，讓支持研究的資料能「多源化」。再引前述 THDL 內的古契書資料為例，在過去要能同時看到那麼多的契書，是件十分困難的事，即便有一「理想的」圖書館能收集所有編印出版的契書史料，全部攤開於研究者眼前，對研究者來說反而是種亂無頭緒的巨大負擔與挫敗。而今日只要透過網路的連結，使用者便能在自家電腦上立刻接觸所有的古契書，不用再舟車勞頓去尋覓那「理想的」圖書館。尤有甚者，數位化不單解除了本本古契書史料集的紙本限制，同時更開啟了有效重構並觀察史料脈絡的可能。不只是多源，而且是挖掘出潛藏條理枝繁葉茂的多源景象，令使用者可以依據其研究的尺度，對檔案資料進行閱讀，使多源不再是難能可貴的理想，更不是使用者的負擔，而是使用者能輕易使用的利基。

檔案史料原子化的問題，則可分成兩個層面，一是在研究流程上，「見樹不見林」的問題，並非是數位時代所獨有，而是紙本時代即有的現象，研究者僅翻閱可能和研究有關的部分，只選取關鍵字「拿了就跑」的割裂式引用，一直是被批評、教導學生避免的陷阱。反過來說，之所以如此，便在於人有限的精力和時間在面對龐大的史料量時，無法負荷，這正是資訊技術所能輔助之處。是以，割裂式的引用是從紙本時的延續，而有效的資訊技術正是可能的解答之一。在另一個層面上，檔案資料在相當程度上，其本質便理應被原子化的對待和處理，紙本的編輯形式硬生生改變、重組檔案脈絡出版成書，反而對檔案史料帶來難以估計的傷害。前述《淡新檔案》即為一例，重新以後人的眼光編排雖是一種不得不然的安排，但卻也必然對檔案的理解帶來侷限的可能，如果能維持其個別檔案原子化的型態加以保存，並能維持一種可自由調整關聯的串連方法，無疑才是最佳的整理方式，而這是在紙本世界所難以達成的。《淡新檔案》絕非歷史上的孤例，事實上紙本保存的脆弱，讓很多檔案資料在流散後重輯的過程中，都面臨了一樣的難題。如在宋代研究中十分重要的史料《宋會要輯稿》為例，該書即是在《宋會要》亡佚後，從《永樂大典》中輯出，最初的編纂者徐松，便以無法有效回復《宋會要》的原貌，只能自行加以重構，而後又面臨多次的轉手、修正和割裂，今日的《宋會要輯稿》和《宋會要》之間有著再多的考證仍難以填補的差距。²¹ 然而一旦還原其原子化的風貌，輔以各種脈絡觀察的可能，一來或可析釋後世修刪者的詮釋，二來能讓《宋會要輯稿》的內容依據時間、空間或其他研究需要自由編排，雖然無法恢復原貌，但或許更能切合《宋會要》作為「政書」的原意。

總而言之，一旦當我們重新去深入思考檔案資料的本質，檔案唯有經過數位化後，才能彰顯出檔案的本質。紙本編排方式的穩當，仍是建立在對檔案資料之脈絡和可能性的壓縮上。是以，當檔案數位化重新呈現其風貌時，無論在研究和應用

21 參見陳智超，1995，《解開《宋會要》之謎》。

上都能開發出全新的可能。檔案理應多元、多源和原子化，那些令人擔憂的難題，與其說是針對檔案的還原，倒不如說是在檔案從紙本限制中解放後的駕御。因此，如何駕御數位化的檔案資料，才是數位人文真正的關鍵之處，而要能達到這樣的目的，其解答便是「系統」。

（二）數位化檔案的核心方法論：系統

如前述，數位化的檔案要能發揮其研究價值，最關鍵之處便在於「系統」。傳統檔案整理過程中，檔案的品質和整理分類決定了檔案的使用方式和難易，面對數位檔案則需要非常不同的思維方式。一個非常基本但常常被忽略觀察是數位檔案必須經過中介的媒體才能夠被使用。換句話說，數位化檔案無法單獨存在，使用者一定要透過一個軟體系統才能對檔案進行有組織的研究與使用。這個觀察雖然看起來是老生常談，但卻是數位檔案和傳統檔案最關鍵的差異。也就是說，談數位檔案就不能不談檢索系統，而檢索系統的設計與功能也就決定了數位檔案的可用度，所以我們在探討數位檔案時就不能僅僅從檔案的層面看問題，而是要把檔案的系統當成和檔案一樣重要的一個環節。

既然數位檔案和系統是密不可分的，我們需要花一些篇幅探討一個數位檔案的系統應該，或說能夠，提供什麼。

傳統的檢索系統，包括搜尋引擎、圖書館自動化系統，以及大部分的文件查詢系統等皆以求準率（precision）和求全率（recall）作為評斷系統可用度的指標。直觀的來說求準率指的是檢索結果與使用者的需求的相關度，求全率則是檢索結果包含使用者期望得到的所有文件的比例。既然研究者最關心的是相關的文件有沒有被找到，一般認為一個針對研究者需要所設計的檢索系統需要高的求全率。高的求全率當然是需要的，但它是不是研究者唯一的需求呢？

當我們進一步檢視檢索系統的設計時，會發現這些傳統系統背後大都有一個假設，就是系統內所收集的文件之間沒有相互的關聯。換句話說，當使用者下了一個關鍵詞，系統將搜尋到的文件用系統內建的排序函數（ranking function）排序後依照系統認為的相關順序條列給使用者。但是因為系統內的排序函數和使用者對相關的認知可能有很大的差異，所以所回傳的文件的順序往往不符合使用者期待，而給使用者一種支離破碎的感覺，這也是為什麼數位檔案的使用系統會給人一種破壞原有檔案脈絡的印象。

但是對一個研究者而言，檔案中的文件是有關聯的；當一個研究者使用一個檔案系統尋找資料時，除非他很清楚自己要找的是某一個特定文件，不然的話，通常都會要研讀所搜尋到的一堆文件（而不是單一的一件）。而在研讀這些文件時，要

找的也就是文件與文件間的脈絡。換句話說，對研究者而言，文件除了本身的意義外，更有文件間經由脈絡的串連所共同產生的意義，這些脈絡往往才是研究者在尋找的，而文件本身即是支持這些脈絡所產生的意義的證據。

所以一個以研究需求為出發的數位檔案檢索系統需要考慮到文件間的脈絡問題，其建置的基本邏輯除了要保持所收錄的檔案本身原有的脈絡外，還需要具有開放性，能夠產生與呈現各種不同連結可能的多元脈絡；一個符合使用者需求的檔案檢索系統應盡量發掘文件間各種關聯，並提供使用者一個觀察這些脈絡的環境。

既然檔案中的文件不是單獨存在的，一個檔案檢索系統理論上來說應該可以產生及呈現檔案中任何一個文件子集合中的重要脈絡，事實上也是如此。但是為了說明的方便起見，我們可以把任一文件子集合想成一個檢索成果的集合（a query return as a sub-collection），而檔案檢索系統的一個重要目標就是提供各種方法讓使用者對檢索成果的集合中的檔案文件間脈絡做觀察與分析。

以上我們指出一個檔案的檢索系統不能僅被視為檔案數位過程中的「附加物品」，而應當成是數位檔案中重要的一環；任何一個系統的建置，都應依據檔案的特性，量身設計出所需要的系統。而這也才是建構數位檔案最核心的方法論，也才能讓數位化檔案的研究潛能真正發揮。

（三）數位化檔案系統所能提供的檔案脈絡

既然一個檔案檢索系統的任務不僅是讓使用者可以檢索到需要的文件，而且更要提供文件間的脈絡，我們就需要進一步分析一個檔案系統應該提供給使用者怎樣的史料脈絡。我們可大致將這些脈絡歸結成六類：（一）檔案原始的脈絡；（二）重組原始脈絡產生的多重脈絡；（三）鳥瞰型的脈絡；（四）子文件集的縱觀脈絡；（五）文件間統計型的脈絡；（六）文件間隱藏的新脈絡。以下將用國立臺灣大學數位典藏研究發展中心目前所發展的檔案檢索系統做例子，²²說明這六種脈絡的意義及如何達到。

1. 原始脈絡的保存——保持原本脈絡的瀏覽功能

無論數位化檔案的系統功能多麼強大，其基礎仍還是構築在史料保存的真實和完整上，誠如論者已指出的「檔案的使用者絕多為史學研究者。史學研究者對資料的掌握，既要求完整又求正確」，再加上許多單位在檔案數位化後，基於保護史料的原則更不希望使用者閱讀原件，所以資料庫要提供檔案的瀏覽功能，應讓使用者

22 臺灣大學數位典藏研究發展中心的首頁為：<http://www.digital.ntu.edu.tw/>，除本文所提到的相關資料庫外，該中心網頁還提供了豐富的訊息和資料庫，歡迎讀者前往使用。

可以留有傳統的閱讀習慣。²³如果從研究的角度來看，即是要保留檔案的原始脈絡，讓研究者在開展自己論述的同時，可以隨時回歸到史料最原初的狀態，掌握檔案的時間、機構、全宗、相關事件等等資料。以「臺灣省議會史料總庫」為例，²⁴使用者可以在不輸入任何關鍵字的情況下，直接點入「會期瀏覽」，按會期一頁頁的閱讀。此外如圖1所示，「類別瀏覽」功能則可以讓使用者一目瞭然檔案的歸屬。當使用者在閱讀「三七五地租」的相關檔案時，可以立刻知道其原本是屬於民政事務>地政>地權；當在閱讀「菸酒公賣」的資料，可以馬上知道是在「質詢」的類目下。

2. 重組原始脈絡產生的多重脈絡——既有知識結構的重組

紙本印刷的最大限制是只能提供使用者一種脈絡，這個限制使得檔案只能以一個線性的方式發展。這個限制在數位的世界可以被完全打破，而允許使用者將原有的脈絡重組成多重的脈絡。

一般來說，一個整理完好的檔案都有它本身的分類（樹狀）結構，而這個知識結構一旦固定後，檔案中的每份文件也就被這個分類固定住。我們可以舉 THDL 研究工具集中「清代臺灣官職表」系統為例，²⁵該系統以鄭喜夫編纂，國史館臺灣文獻館所出版的《臺灣地理及歷史·官師志》為藍本，這本書是研究清代臺灣官制重要的參考書籍，當我們試圖將這本書由紙本轉換為數位系統時，不只是把資料放上網路，也不僅提供檢索。更重要的，這些資料變成立體的，有結構的，彼此連結的。從原書的目錄即可知，該書是以官職作為分類，按年代先後，依序記錄曾任該職位的人名、就任時間與原委、離職時間與原因。這樣的安排堪稱詳密，但卻無法方便的對一個人仕宦生涯進行觀察，即以人作為出發點，而只能被限定以職位作為唯一的視角，而且還是沒有上下隸屬關係的孤立記載，官職背後所具有的變動或其他觀察可能都被排除了。

但轉化為數位檔案後，即能打破紙本的凝滯，以研究者的需求靈活排列，進行資料組合與調整，並以樹狀結構還原官職原本應有的歷史結構（見圖2）。以譬喻形容的話，前者檔案就像拼圖，後者則像積木，一片片拼圖雖然看似可以分開，但一定還是要合在一起才有意義，而其拼湊的方法也只有一種，就是原來設計者所畫出的圖像。積木則大不相同，它可以依照各種不同的想法和需要，而拼造出各式各樣組合。

23 莊樹華、龐桂芬，〈數位化時代下的檔案館營運——以中研院近史所檔案館為例〉，《檔案季刊》，第5卷4期，頁26。

24 「臺灣省議會史料總庫」網址為：http://ndap.tpa.gov.tw/drtpa_now/。更多關於該資料的介紹可參閱項潔、董家兒、蕭屹靈，2008，〈臺灣省議會檔案數位典藏系統之研發與建置〉，《臺灣省諮議會會訊》，16期，頁33-46。

25 該系統網址為：http://140.112.30.230/Career_tb/。

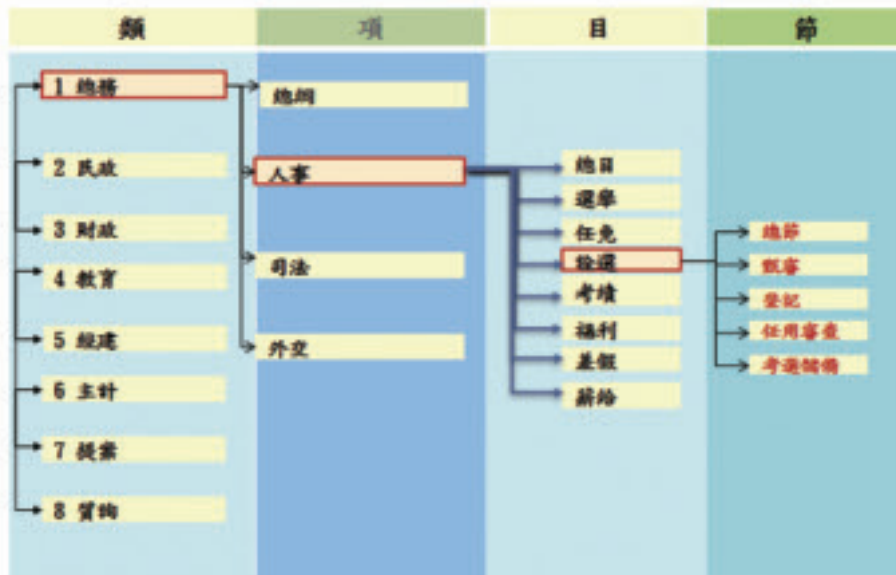


圖 1-1 臺灣省議會史料總庫的原始結構

臺灣省議會史料總庫

Digital Repository of Taiwan Provincial Assembly

首頁與檔案 | 網站說明 | 檔案簡介

類別瀏覽 | 資料瀏覽 | 搜尋

檔案類別瀏覽

- 總務 (17838)
 - 院政 (19458)
 - 總務 (9975)
 - 地政 (4727)
 - 地目 (97)
 - 地權 (427)
 - 地價 (1119)
 - * 地價 (3298)
 - * 地價登記 (81)
 - * 土地取得稅 (529)
 - * 三七五地租 (37)
 - * 土地賦稅 (1988)
 - * 地價 (85)
- 地價 (34)
- + 軍警 (1752)
- 財政 (7973)
 - 教育 (1269)
 - 經建 (14283)
 - 工計 (3863)
 - 糧食 (31167)
 - 寶璽 (11142)
 - 總務 (13142)
 - 地目 (79)
 - 院政 (1628)
 - 財政 (1528)
 - * 總務 (3138)
 - * 財政主計 (174)
 - * 行庫 (181)
 - * 財源計畫 (14)
 - 建設 (1598)
 - 農林 (874)
 - 教育 (629)
 - 交通 (1818)
 - 總管理 (2257)
 - 其他 (78)

- 議事錄 / 公報與新聞
- 報告 (3386/3210)
- 會議紀錄 (3881/4996)
- 調查報告彙編 (2082/2831)
- 宣言聲明 (331/78)
- 提案 (82418/139449)
- 質詢 (20444/91320)
- 公文 (5/1706)

搜尋字:

搜尋目標: 檔案 公報 議事錄

類別瀏覽

欄位檢索輸入說明:

- 日期檢索請按格式：西元年(4碼)-月(2碼)-日(2碼)
例：1945-06-23 或 1952-02
- 檢索條件組合提供三種邏輯字
 - AND：使用符號 & 或 and 連結條件。ex: A&B&C
 - OR：使用符號 | 或 or 連結條件。ex: A|B
 - NOT：使用符號 - 置於條件之前。ex: -A

圖 1-2 臺灣省議會史料總庫的原始脈絡保存



圖 2 清代臺灣官職表的樹狀結構呈現

3. 鳥瞰型的脈絡——文件集整體意義的顯現

當我們在面臨大量的檔案資料時，好的檔案系統可以提供使用者立即的鳥瞰脈絡，對檔案的構成有即時而大致的理解，如圖3為例，透過THDL的整理，可以將龐大而多源的明清檔案內所收的文件，立刻於時間軸劃出年代的分布。假設某個年份的文件數量越多，則表示當年某些事件對臺灣政治社會的震撼越大；是以對臺灣影響最大的前六件事件依序為：（1）1787年林爽文事件；（2）1884年中法戰爭相關事務；（3）1895年中日甲午戰爭相關事務；（4）1874年牡丹社事件；（5）1806年蔡牽之亂；（6）1833年張丙事件。然而倘若和圖4明清檔案中由《清實錄臺灣史資料專輯》中的文件分布進行比較，可以明顯發現林爽文事件壓倒性的勝過其他事件，兩相比較，或者可以假設清廷官方對於林爽文事件的重視程度，遠超過中法或甲午等戰役。此外，影響臺灣西部開墾甚為重大，並造成族群劇烈遷移的郭百年事

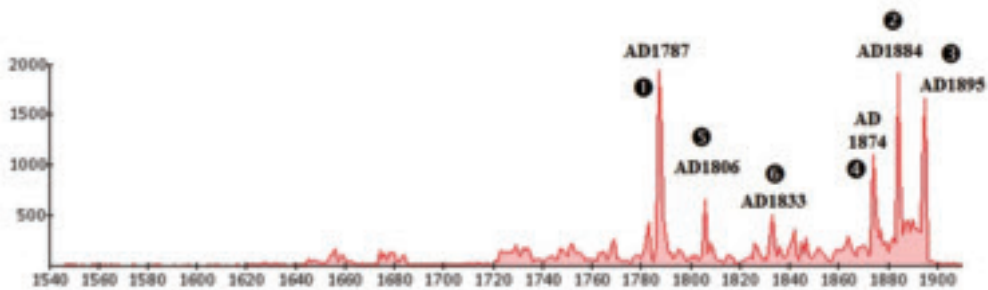


圖3 明清檔案文件在年代上的分布圖

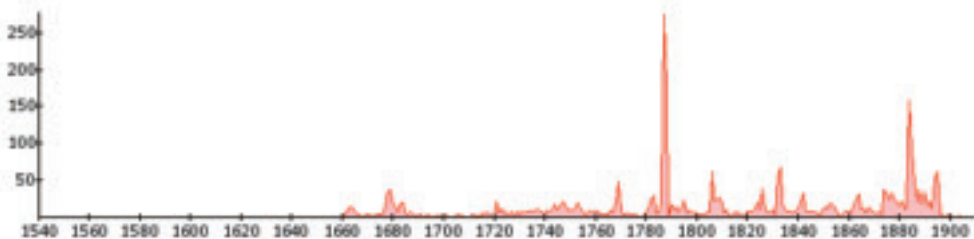


圖4 明清檔案中《清實錄臺灣史資料專輯》文件在年代上的分布圖

件，²⁶卻未在明清臺灣行政檔案之中留下任何的紀錄，再度顯示了中央和地方在視野上的差異，而這種差異性是單憑人力一件件閱讀所無法勾勒的。前述的解釋當然還有待研究者進一步的析理，但要強調的，這種鳥瞰性脈絡的觀察方式，是紙本或傳統檢索系統所難以想像的。

4. 子文件集的縱觀脈絡——檔案原有知識結構的多重脈絡呈現

縱觀脈絡是一個檔案檢索系統應該提供的最核心的脈絡，也是我們方法論中最基礎的觀念。大體來說，每個檢索系統都允許使用者用查詢的方式尋找文件，而查詢的結果可能來自檔案知識結構的各個支系。這個查詢結果可以被看成原有文件集的一個子集合，而所謂縱觀脈絡就是指這個子集合在原來知識結構之多重脈絡裡所呈現的各種關係。這些關係中最明顯的是原本詮釋資料（metadata）的欄位中本就蘊含的。舉個例來說，如果年代是一個詮釋資料中的欄位，則年代可以用來作為子集合中文件分類的一個關係。我們稱這種呈現關係的方法為「後分類呈現」的技術。如在THDL中，使用者可以將檢索得到的子集合依年代、出處、作者、類型、地域等欄位做後分類，然後進行更進一步的脈絡觀察。

圖5是THDL中一個檢索成果的後分類呈現。其中左上角是檢索成果的地理分布；（針對古契書我們也做了地理資訊的分析，²⁷這些座標資料也是後分類的一個方法）中間是時間分布；左下是詳細的五種後分類，分別是年代、出處、作者、性質分類及地域。圖6顯示這五種後分類的內容。

但我們要指出詮釋資料的欄位並不是唯一的後分類方法。其他資訊（如子集中出現的人名、地名）也可以用來做後分類。後分類的目的是讓使用者對檢索成果得到的子集合中的文件用各種不同的角度觀察時可以看出的分布情形。這些分布往往是重要脈絡發現的開始。

我們再以「日治法院檔案資料庫」為例，說明後分類的效果。「日治法院檔案資料庫」的主要內容為臺北、新竹、臺中、嘉義等四個地方法院所典藏的日治時期各類司法文書，包括民刑事判決原本、民刑事案件登記簿、非訟事件以及強制執行事件等卷宗、公證書原本、有關法人等的各種登記簿、行政卷宗等，一共 5,645 冊，²⁸檔案內容分為「條」、「項」、「款」，在詮釋資料欄位另外以「冊」及以「案」

26 相關論述與分析可參見陳嵩明，2010，《清代臺灣界外武力拓墾：噶瑪蘭、水沙連、與竹塹東南山區之比較研究》，臺北大學社會學系碩士論文。

27 蔡炯民、歐仲翔、翁稷安、項潔，〈時空資料的地方再現：GIS與數位典藏的交會〉，《國土資訊系統通訊》，78期（2011.6），頁13-20。歐仲翔，〈使用者取向之歷史地理資訊系統：古契書與統計資料呈現〉，國立臺灣大學資訊工程學研究所碩士論文，2011。

28 項潔、蕭屹靈、董家兒，2009，〈日治法院檔案數位典藏系統之研發與建置〉，收於王泰升主編，《跨界的日治法院檔案研究》，頁83-148。



圖5 THDL中以「石岡」為檢索詞的檢索結果

年代後分類

出處後分類

作者後分類

契書性質後分類

| 年代 (DT) | 分類 |
|----------------|-------|
| 總計 | 13604 |
| 清康熙二十四年 (1685) | 1 |
| 清康熙二十五年 (1686) | 1 |
| 清康熙三十六年 (1697) | 1 |
| 清康熙三十八年 (1699) | 1 |
| 清康熙四十年 (1701) | 1 |
| 清康熙四十九年 (1710) | 1 |
| 清康熙五十五年 (1716) | 1 |
| 清康熙五十六年 (1717) | 1 |
| 清康熙五十七年 (1718) | 2 |
| 清康熙五十八年 (1719) | 1 |
| 清康熙五十九年 (1720) | 1 |
| 清雍正元年 (1723) | 1 |
| 清雍正三年 (1725) | 1 |

| 出處 (CL) | 分類 |
|-----------------------|------|
| 臺灣總督府檔案抄錄本的 | 2903 |
| 文書 (17年保存公文檔案 (國中圖9)) | |
| 臺灣總督府檔案抄錄本的 | 2074 |
| 文書 (永久保存公文檔案 (國中圖9)) | |
| 臺灣總督府檔案抄錄本的 | 804 |
| 文書 (土地家屋公文檔案 (國中圖9)) | |
| 臺灣總督府 | 378 |
| 大臺北古契字二集 | 312 |
| 方方地古文書輯抄匯編 | 300 |
| 大臺北古契字三集 | 248 |
| 台灣中部平原漢族古文書 | 210 |
| 大臺北古契字集 | 240 |
| 臺灣總督府檔案抄錄本的 | 228 |
| 文書 (高等科野公文檔案 (國中圖9)) | |
| 詳細收據類本文書 | 224 |

| 作者 (AU) | 分類 |
|---------|----|
| 不詳(典主) | 24 |
| 不詳(賣主) | 13 |
| 事件無相關資訊 | 9 |
| 謝榮(典主) | 7 |
| 謝榮(賣主) | 7 |
| 高顯(賣主) | 7 |
| 李族典(典主) | 7 |
| 陳水清(賣主) | 6 |
| 胡德吉(賣主) | 6 |
| 胡序(賣主) | 6 |
| 王峰(賣主) | 6 |
| ccc(賣主) | 6 |
| 郭紀華(賣主) | 5 |
| 林朝(賣主) | 5 |
| 朱祥光(賣主) | 5 |
| 謝典(賣主) | 4 |

| 契書性質 (DT) | 分類 |
|-----------|------|
| 買賣契 | 6797 |
| 其他 | 4437 |
| 典契 | 3009 |
| 贈與契 | 273 |
| 證明協議(協議) | 259 |

地域後分類

| 地域 (GE) | 分類 |
|---------|-----|
| 竹堑一堡 | 848 |
| 大廳庄 | 89 |
| 望山庄 | 88 |
| 鹿寮坑庄 | 68 |
| 鹿港公庄 | 43 |
| 新埔庄 | 42 |

圖6 「石岡」檢索成果的年代、出處、作者、契書性質、及地域的後分類

為單位分別建置的情況下，共有 307,139 案的詮釋資料，200 萬餘張影像。這批資料除了數量龐大與保存現況甚差外，內容本身在當初形成的時候也因來自各個法院而有差異，如果要對這批資料做綜合性的研究會十分困難。

「日治法院檔案資料庫」和我們建置的其他資料庫一樣，都用檢索成果後分類作為輔助使用者觀察縱觀脈絡的機制。²⁹ 我們以關鍵詞「離婚」為例。以離婚做關鍵詞檢索可以找到 2,740 筆資料，從法院所屬後分類可看出新竹、臺中、臺北、嘉義地方法院均有（分別為 204、875、1,108、553 筆，其中民事判決原文分別為 204、875、1,103、553 案，其餘六件為公正證書），年代分布從明治 38 年（1905）的兩案到昭和 20 年（1945）的五案，每年都有，其中在大正 9 年（1920）起進入每年超過百案的高峰期。雖然在昭和 8 年（1933）後案件數量回復兩位數字，但一直到太平洋戰爭開始（1941）才降到每年五十案以下。這種透過檢索得到的子集合中的文件可能散布在檔案的各處，這些文件中的縱觀脈絡透過後分類可以輕易被觀察到，但是不用後分類的機制則不易被察覺。針對這 2,740 筆資料做進一步的觀察可以發現，其中 85% 的離婚案件的原告是（臺灣）女性。這個現象透過後分類可以立刻被觀察到。如果沒有後分類的觀念則極為困難。

後分類的觀念和 *faceted search*³⁰ 類似，但基本的出發點有很大的差異。*Faceted search* 後分類的目的是讓使用者在使用不夠精準的關鍵詞檢索的情況下可以比較容易的從大量的檢索結果中找到想要的文件，所以它的設計還是以找到某個特定文件（也就是作為加強求準率的效能）為主。我們的後分類的目的則是提供使用者所檢索到的子集合中的文件之間的關係。因為不知道使用者所想的脈絡是什麼，所以要盡量提供各種不同的後分類及各種呈現的方法，以讓使用者可以對他所需的做觀察。

又或者如「總督府抄錄契書 GIS 工具」，如前述在面對大量的古契書資料，使用者往往不知如何進入，一旦輔以 GIS 系統便可以立即掌握契書所在的位置，同時還可以再依文件種類和時間的過濾，對土地開發有概略的理解。以姜秀鑾對新竹一帶的開發為例，圖 7「姜秀鑾對新竹開發示意圖」是以道光 14 年（1834 年）12 月，淡水廳同知李嗣業任命的金廣福墾隘的粵籍頭人姜秀鑾為關鍵字，分別檢索 1835-1836 年、1835-1838 年、1835-1850 年、與 1835-1895 年這四個時間區段，觀察與姜秀鑾有關之契約文書在「總督府抄錄契書地理資訊」雛型系統上的座落分布，從圖中可以看出這些契約文書的座落分布，最早由今日的新竹縣的竹東市開始，逐漸向北埔、寶山、及峨眉三個鄉拓展。除了讓我們可以依時間序列與地理分布兩者關

29 「日治法院檔案資料庫」的網址為：http://tcra.lib.ntu.edu.tw/tccra_develop/。相關研究範例，可參見項潔、蕭屹靈、董家兒，2009，〈日治法院檔案數位典藏系統之研發與建置〉，收於王泰升主編，〈跨界的日治法院檔案研究〉，頁 83-148。

30 參見 Vanda Broughton, 2006. The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval, *Aslib Proceedings*, Vol. 58 Issue 1/2, pp. 49-72.

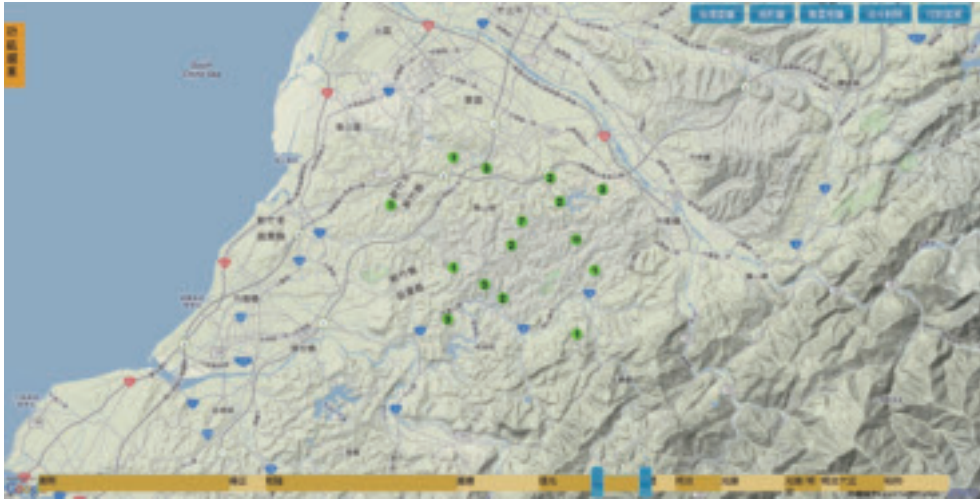


圖 7-3 1835 年至 1850 年，以峨嵋鄉為基地，持續開發峨嵋溪流域

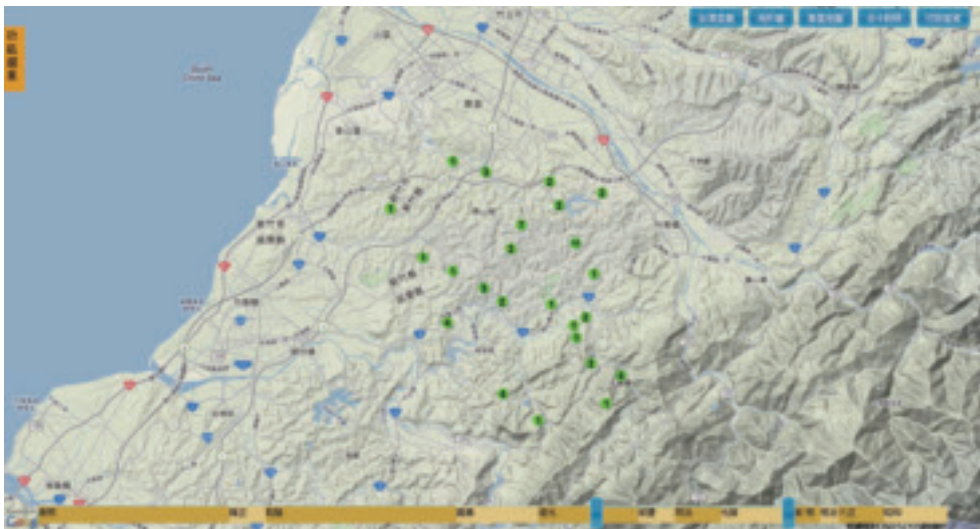


圖 7-4 1835 年至 1895 年，可看出金廣福開發範圍

係，約略對照出當年金廣福整合新竹東南山區的官民隘，向新竹東南山區推進的拓殖開發之情形，更可進一步探索其拓展情形是否受其他地理環境或政經社會條件所影響。若依文書種類進行檢索後分類，也可以看出不同文書種類在時間序列上所顯示的差異，尤其1886年（光緒12年）以後，福建臺灣巡撫劉銘傳推行清賦，裁廢墾隘，金廣福正式走入歷史，雖然還有與姜秀鑾相關的契約文書出現，但大概為與分割或轉手相關的種類，而非首墾契書了。³¹

5. 文件間統計型的脈絡——文件子集中文字及數量的分析

系統如何幫助研究者，在文件中清理、建立出統計型的脈絡，THDL中的詞頻分析，以及「臺灣歷史數位圖書館工具集」中的前後綴詞分析工具是最好的例子。詞頻工具是指詞彙在檢索結果中出現的頻率，利用詞夾子程式，³²清理、擷取出THDL檔案中的人名、地名和其他專有名詞，再統計出現的頻率，亦會使研究者在檢索之後，能從事更有效與簡單的再觀察對象。如朱一貴在不同年代的高峰，若再加入詞頻進行分析，便會發現和朱一貴最常出現的人名竟不是和事件直接關聯的藍廷珍、施世驃等人，反而是林爽文、常青等，在另一亂事中的人物，在經由對史料內文的解讀後，我們可以初步推論當乾隆在面對林爽文之亂時，其祖父康熙在平定朱一貴之亂的種種戰略與思考，成為他重要的參照點。可能正是因為林爽文事件的存在，使得朱一貴之亂重新被提出、詮釋，構成新的歷史記憶。這樣交插、跨研究對象的比對和討論，是單從紙本閱讀經驗中所難以察覺的。³³

前後綴詞和詞頻工具相同，亦是應用詞夾子程式，製作出一個簡易的工具介面，讓使用者可以對THDL所藏的檔案文件，進行關鍵字的前綴詞的分析。³⁴可以查詢所輸入的關鍵字，其左方（即關鍵字之前）或其右方（即關鍵字之後），所出現的字串，並統計其出現的篇數和次數。在檢索的過程除了可以指定字串的長度外，還可以指定全文需含有的詞彙。比如說我們可以輸入「銀」為欲分析的關鍵詞，觀察其前方三個字數的綴詞，藉以觀察明清檔案中和臺灣有關的銀圓種類，其結果為：佛銀（含佛面銀、佛銀、佛頭銀、番佛銀、佛番銀、清水銀）9,127件、番

31 參見蔡炯民、歐仲翔、翁稷安、項潔，2011，〈時空資料的地方再現：GIS與數位典藏的交會〉，《國土資訊系統通訊》，78期，頁13-20。

32 關於詞夾子程式的設計與應用，可參見謝育平，2010，〈同位詞夾子：主題式分類詞庫萃取演算法〉，《數位人文研究的新視野：基礎與想像》，頁133-162；謝育平、楊龍廉、趙建宏、黃銘立、古馮文、林郁智，2009，〈使用詞夾子建立中文典籍分析加值服務〉，發表於「銘傳大學2009資訊科技與實務研討會」；張尚斌，2006，〈詞夾子演算法在專有名詞辨識上的應用——以歷史文件為例〉，臺北：國立臺灣大學資訊工程學研究所碩士論文。

33 涂豐恩、杜協昌、陳詩沛、何浩洋、項潔，2010，〈當資訊科技碰到史料：臺灣歷史數位圖書館中的未解問題〉，《數位人文研究的新視野：基礎與想像》，頁21-44。

34 該工具網址為：<http://thdl.ntu.edu.tw/SimpleTools/TermPat/TermPatSimpleUI.php>。

銀866件、龍銀842件、劍銀210件、花邊銀（含花銀、灼銀）76件、紋銀301件、六八銀157件。可知主要皆為外國銀，分布如圖8。

6. 文件間隱藏的新脈絡——藏於語意（semantics）中的脈絡

發掘研究者所未察覺的新脈絡，可以說是系統最終極、最理想的功能，意即系統不單只是被動的觀察工具，而是能夠更主動的替使用者發現具深刻意義的研究議題。這樣新的脈絡往往藏於語意之中，單憑搜尋、檢索難以發現。同樣收錄在「臺灣歷史數位圖書館工具集」的土地移轉圖和引用關係圖兩樣工具就是最好的例證。土地移轉圖重新將古地契間的脈絡進行重建，³⁵利用資訊科技完成這項用人力幾乎不可能完成的任務，清理出THDL中來自100種不同出處超過30,000件的地契文書，析理出土地的身世。它重建了土地的上下手契關係，利用資訊技術自動擷取契書的標題、人名、四至、土地面積、賣價、地號、時間等特徵，在兩兩比對後找出契書間的關聯。接著再把上述一對或一組的上下手契與鬮分契串連起來，就成為某一塊土地在不同地主手中的轉移圖。在三萬多件契約中，出現了2376個土地移轉圖（最大的有103份契約），每一個圖都反映著一塊土地的身世。如圖9這是一塊苗栗永和山地區的土地交易情形，圖中主角為率先開墾此地的廖姓家族。這個圖在社會史、經濟史和家族史所呈現的意義，有待研究。

THDL《明清臺灣行政檔案》引用關係圖利用明清行政檔案當中，臣子的奏事文書、與皇帝諭旨文書之間相互的引用關係，將數量龐大卻零散的行政檔案，連結成反映詳細政治決策過程的「引用關係圖」。³⁶它打破了原有的檔案脈絡，即按檔案所屬部門的分類，如它是來自宮中檔，還是來自內閣大庫；或者，它是直通皇帝的奏摺，還是要經過六部的題本；改以奏摺與上諭間彼此間的引用和回應，去重建事件處理過程中官方的往返與運作。其反映了清帝國內訊息的傳遞和流通，以及政策形成的決策過程，更重要的，它也顯現了面對同一事件，不同地區、不同層級的官員，出於自身立場或利益考量的迥異看法，這些新的脈絡都有待研究者去進一步的考察。表1為引用關係所涉檔案最多之前二十件，及其大致內容。

35 土地移轉圖查詢系統網址為：<http://140.112.30.230/GRAPH/INDEX/newindex.php>。其數位化的方法和過程可參見黃于鳴，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。對於古契書的數位整理可參見項潔、陳詩沛、杜協昌，〈台灣古契約文書全文資料庫的建置〉，2009，《第三屆台灣古文書與歷史研究學術研討會》，逢甲大學歷史與文物管理研究所、台灣古文書學會主辦，臺中市：逢甲大學；盧家慶，2008，《台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。

36 THDL《明清臺灣行政檔案》引用關係圖網址為：http://140.112.30.230/citation/list_mq_graph_8.php。

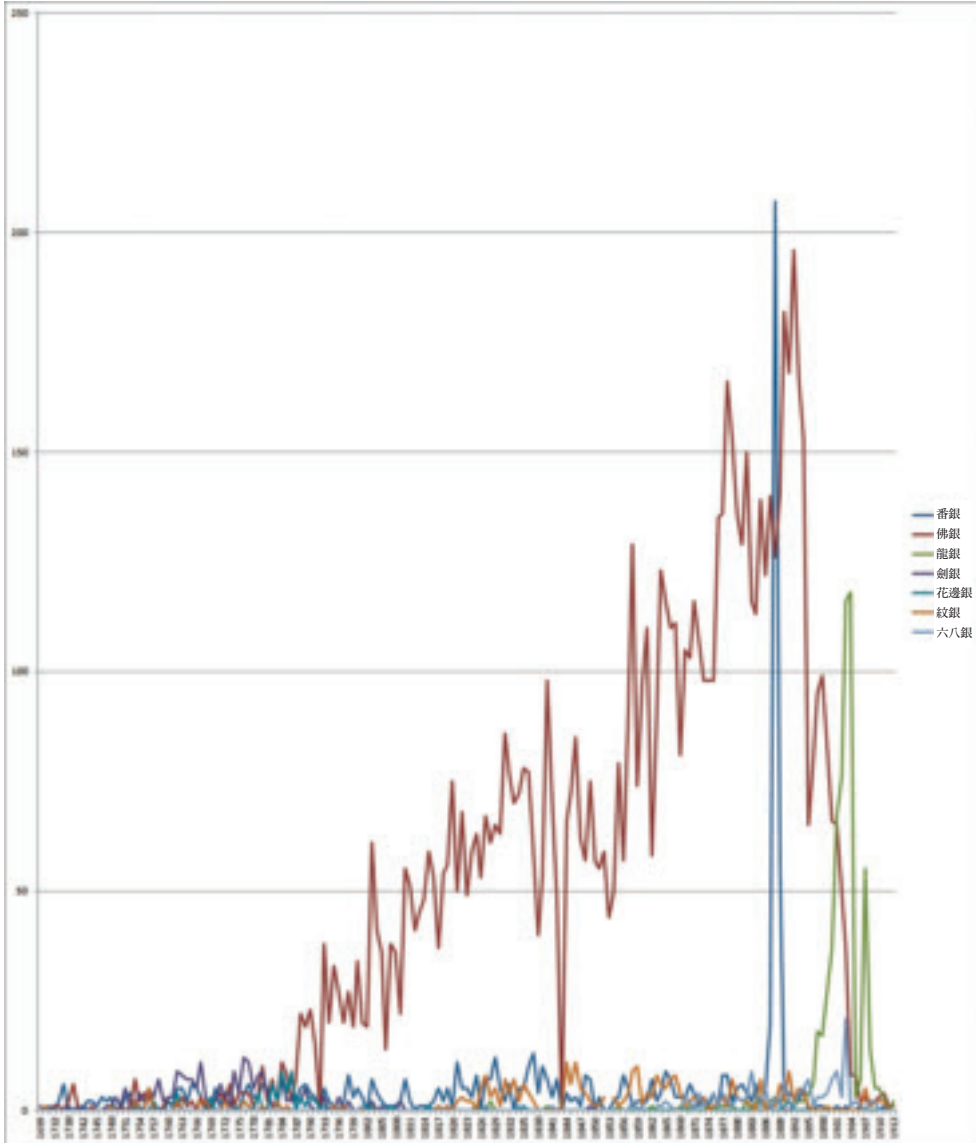


圖8 古契書中主要銀圓類別的年代分布

表1 引用關係所涉檔案最多之前二十件

| 引用關係圖 (檔案數) | 主要年代 | 內容 |
|----------------|----------------------------|--|
| 153, 126, 109 | 乾隆52年 (1787) | 林爽文事件產生的奏諭 |
| 107 | 乾隆53年 (1788) | 林爽文事件的奏諭 |
| 95 | 乾隆52年~同治6年 (1787-1867) | 林爽文事件引發清治臺灣的政策討論: 乾隆下令「嗣後凡遇有台灣道員缺, 俱自加按察使銜俾得自行奏事」, 日後臺灣道一職補授必定引用此諭旨。 |
| 85 | 嘉慶11年(1806) | 蔡牽事件產生的奏諭 |
| 83 | 同治13年(1874) | 牡丹社事件產生的奏諭 |
| 62 | 光緒10~11年 (1884-85) | 中法戰爭期間法軍攻打臺灣的基隆與滬尾產生奏諭 |
| 47 | 同治9年~光緒1年 (1870-75) | 牡丹社事件後日本兵船盤據龜山一帶 |
| 45 | 乾隆51~52年 (1786-87) | 林爽文事件剛發生時的奏諭 |
| 45 | 道光21~22年 (1841-42) | 中英鴉片戰爭期間, 兩艘英國船隻出現在淡水與大安, 被臺灣道姚瑩與臺灣鎮總兵達洪阿逮捕。 |
| 43 | 光緒11~16年 (1885-90) | 劉銘傳上任臺灣巡撫後開山撫番及其事後請功的內容 |
| 42 | 光緒11~14年 (1885-88) | 清廷檢討海防事宜, 臺灣身為「南洋門戶」因此也是重點地區之一, 包括臺灣建省。 |
| 37 | 同治13年~光緒1年 (1874-75) | 牡丹社事件後, 光緒皇帝點名多位官員提供海防籌備計畫。 |
| 35 | 道光12~13年 (1832-33) | 張丙事件 |
| 30 | 道光6年(1826) | 彰化、大甲、中港、後壠械鬥事件 |
| 29 | 同治4年~光緒19年 (1865-93) | 臺灣人事升遷的商議文件 |
| 29 | 同治7~8年 (1868-69) | 同治年間的中外衝突 |
| 28 | 康熙60年~道光12年 (1721-1832) | 巡臺御史存廢的討論 |
| 28 | 乾隆52~56年 (1787-91) | 林爽文事件過後恢復事宜的報告 |

五、結語

傳統的檔案保存方式有其自身的邏輯，是一個經過長期運作，於實務之中不斷改善後所得的學問，是值得重視並尊重的學術典範。然而，當保存方式由紙本演變成數位，印刷文字被二進位元所取代後，保存的邏輯必然要與時俱進，並非要完全否定傳統檔案保存重要性，而是要在尊重傳統核心的前提下，保留其核心價值，並賦與其數位的全新風貌，衝撞、生成全新的可能，呈現出數位化時代檔案保存的全新價值與脈絡。

檔案本身自然不會說話，必須要透過史家的詮釋才有可能彰顯其意義，紙本的檔案處理並不是完全沒有問題的，尤其是單一脈絡的編輯方式，很容易造成誤導和盲點。理論上，通過學者的勤勞與用心，在檔案間細細爬梳，仍然可以躲過編輯者的分類大刀，重新組合出自己的需求，但終究只是理論上而已，面對汗牛充棟的史料處理，對運氣的需要很可能勝過努力，更遑論過程中的費時費力。

數位化後檔案的新脈絡不但不是要否定檔案和史家之間的既有關聯，更是要加以拉近兩者的距離。檔案經過數位化處理後，不但能保存原有的收藏邏輯，更能賦與其不受篇幅和印刷限制的靈活和彈性；史家仍是史料最終的闡釋和解讀者，但通過資訊的處理，除了創造出一個令研究者更能夠觀察的環境，並析理出埋沒在史料中的新線索，使史家可以將心力投注在對史料的解讀。或許，當史料獲得和整理的問題被資訊技術解決之後，不再是困擾研究者的重擔，如何發展出鞭辟入裡的新論述，可能才是數位化時代的檔案脈絡所給予研究者的真正挑戰。總之，經由檔案的數位化，傳統檔案單一脈絡的限制被打破，檔案整理與檔案研究之間的矛盾亦隨之減輕，關鍵便在於建立一套以研究為主要思考取向的檔案檢索系統，透過檔案檢索系統的設計，檔案的豐富且多重的脈絡可以被發現、被呈現、被觀察。

西方重要史家勞倫斯·史東（Lawrence Stone），曾有一段對於史家使用檔案的鮮明描述，他說：

當你在檔案堆裡工作，你就遠離了舒適的家，你會很不耐煩，你焦躁不已，你會像發瘋似的塗塗寫寫。你總會犯錯，我不相信西方世界有任何一個學者做註可以做得十全十美。檔案研究是一項煩人的生計。³⁷

我們衷心希望，在數位化的檔案時代，史東所感受到的焦躁可以一掃而空，使用檔案可以變成一個令人感到愉快和刺激的過程，並大大的提升對檔案引用的正確率。就如同愛德華·卡爾那句經典名言：「歷史是歷史學家跟他的事實之間，不斷

37 轉引自 Richard J. Evans 著，潘派泰譯，2002，《為史學辯護》，頁 140。

交互作用的過程，是現在跟過去之間，永無止境的對話！」³⁸我們相信數位化工程，將會在史家和史實的交互作用中扮演起重要的對話，它將使史家能更為貼近、理解他所面對的史料，進而開展出一場深刻而愉快的對話。

38 轉引自 Edward H. Carr 著，江政寬譯，2009，《何謂歷史？》，頁 126。

參考文獻

- Edward H. Carr 著，江政寬譯，2009，《何謂歷史？》，臺北：博雅書屋。
- Richard J. Evans 著，潘派泰譯，2002，《為史學辯護》，臺北：巨流出版社。
- Vanda Broughton. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, Vol. 58 Issue1/2. pp. 49-72.
- 汪榮祖，2002，《史學九章》，臺北：麥田出版社。
- 林文凱，2004，〈清代晚期臺灣的土地法律文化——淡新檔案內淡新地域漢墾莊在十九世紀四十年代以來土地抗租案件為主的分析〉，發表於臺灣社會學會等主辦「2004臺灣社會學會年會暨『走過臺灣——世代、歷史、與社會』研討會」。
- 洪麗完，2007，〈評論稿〉，收於逢甲大學歷史與文物管理研究所、台灣古文書學會編校，《第三屆台灣古文書與歷史研究學術研討會論文集》，頁331，臺中：逢甲大學出版社。
- 徐紹敏主編，2001，《檔案文獻編纂學（第二版）》，頁8-9，杭州：浙江大學出版社。
- 涂豐恩、杜協昌、陳詩沛、何浩洋、項潔，2010，〈當資訊科技碰到史料：臺灣歷史數位圖書館中的未解問題〉，收於項潔編《數位人文研究的新視野：基礎與想像》，頁21-44，臺北：國立臺灣大學出版中心。
- 秦國經，2005，《明清檔案學》，北京：學苑出版社。
- 張尚斌，2006，〈詞夾子演算法在專有名詞辨識上的應用——以歷史文件為例〉，國立臺灣大學資訊工程學研究所碩士論文。
- 莊吉發，1983，《故宮檔案述要》，臺北：國立故宮博物院。
- 莊樹華，2010，〈數位檔案運用的契機與風險——以近代中國外交檔案為例〉，發表於國立政治大學圖書資訊與檔案學研究所主辦「數位檔案加值與教學應用研討會」。
- 莊樹華、龐桂芬，2006，〈數位化時代下的檔案館營運——以中研院近史所檔案館為例〉，《檔案季刊》，第5卷4期，頁23、26。
- 陳智超，1995，《解開《宋會要》之謎》，北京：社會科學文獻出版社。
- 陳嵩明，2010，《清代臺灣界外武力拓墾：噶瑪蘭、水沙連、與竹塹東南山區之比較研究》，國立臺北大學社會學系碩士論文。

- 項潔、涂豐恩，2011〈導論——什麼是數位人文〉，收於項潔編《從保存到創造：開啟數位人文研究》，頁9-28，臺北：國立臺灣大學出版中心。
- 項潔、陳詩沛、杜協昌，2009，〈台灣古契約文書全文資料庫的建置〉，發表於逢甲大學歷史與文物管理研究所、台灣古文書學會主辦「第三屆台灣古文書與歷史研究學術研討會」。
- 項潔、董家兒、蕭屹灵，2008，〈臺灣省議會檔案數位典藏系統之研發與建置〉，《臺灣省諮議會會訊》，16期，頁33-46。
- 項潔、蕭屹灵、董家兒，2009，〈日治法院檔案數位典藏系統之研發與建置〉，收於王泰升主編《跨界的日治法院檔案研究》，頁83-148，臺北：元照出版社。
- 黃于鳴，2009，《臺灣古地契關係自動重建之研究》，國立臺灣大學資訊工程學研究所碩士論文。
- 歐仲翔，2011，《使用者取向之歷史地理資訊系統：古契書與統計資料呈現》，國立臺灣大學資訊工程學研究所碩士論文。
- 蔡炯民、歐仲翔、翁稷安、項潔，2011，〈時空資料的地方再現：GIS與數位典藏的交會〉，《國土資訊系統通訊》，78期，頁13-20。
- 鄭喜夫纂輯，1980，《臺灣地理及歷史·官師志》，臺北：臺灣省文獻委員會。
- 盧家慶，2008，《台灣古契書自動分類與依分類定義契書角色》，國立臺灣大學資訊工程學研究所碩士論文。
- 戴桓青，2008，〈四代百年臺大情〉，《臺大校友雙月刊》第60期，頁24-25。
- 薛理桂，1998，《檔案學導論》，臺北：漢美出版社。
- 謝育平，2010，〈同位詞夾子：主題式分類詞庫萃取演算法〉，收於項潔編《數位人文研究的新視野：基礎與想像》，頁133-162，臺北：國立臺灣大學出版中心。
- 謝育平、楊龍廉、趙建宏、黃銘立、古馮文、林郁智，2009，〈使用詞夾子建立中文典籍分析加值服務〉，發表於「銘傳大學2009資訊科技與實務研討會」。

