

- Kirschenbaum, M. G. (2009). Done: Finishing Projects in the Digital Humanities. *Digital Humanities Quarterly* 3(2).
(<http://digitalhumanities.org/dhq/vol/3/2/000037/000037.html>)
- Knowles, A. K. ed. (2002). Past Time, Past Place: *GIS for History*. Redlands, CA: ESRI.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team et al. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331, 176-182.
- O'Donnell, J. J. (2009). Engaging the Humanities: The Digital Humanities. *Daedalus* 138(1), 99-104.
- Rosenzweig, R. (2003). Scarcity or Abundance? Preserving the Past in a Digital Era. *American Historical Review* 108(3), 735-62.
- Roush, W. (2005). The Infinite Library: Does Google's Plan to Digitize Millions of Print Books Spell the Death of Libraries; or Their Rebirth. *Technology Review*.
- Spaeth, D. (2003). Research and Representation: the M. Phil in History and Computing. *Computers and the Humanities* 37(1), 119-127.
- Staley, D. J. (2003). *Computers, Visualization, and History: How New Technology Will Transform Our Understanding of the Past*. Armonk, New York: M.E. Sharpe.
- Stone, L. (1971). Prosopography. *Daedalus* 100, 46-79.
- Stross, R. E. (2008). *Planet Google: One Company's Audacious Plan to Organize Everything We Know*. New York: Free Press.
- Srinivasan, P. (1997). Digital Library Projects in the United States. *Bulletin of Information Technology* 17(6), 15-21.
- Zorich, D. (2008). *A Survey of Digital Humanities Centers in the United States*. Washington, D.C.: Council on Library and Information Resources.

Part I. 歷史學與資料庫的交會

When Historians meet Databases

數位化歷史資料庫與歷史研究——
以明清檔案、淡新檔案、日治法院檔案等資料庫為例

Digital Archives and Historical Studies

史料整體分析工具之幕後——
介紹「臺灣歷史數位圖書館」的資料前置處理程序

The Document Processing Workflow of THDL, Taiwan History Digital Library

臺灣歷史數位圖書館與歷史研究的實際應用——
以「淡新檔案」為例

Actual Application of Taiwan History Digital Library and Historical Studies: Case Study on Tan-Hsin Archives

數位化歷史資料庫與歷史研究—— 以明清檔案、淡新檔案、 日治法院檔案等資料庫為例

王泰升 *

摘 要

本文擬分別就臺灣大學「明清檔案」、「淡新檔案」、「日治法院檔案」等三個「歷史資料庫」，說明數位化資料庫的出現，增加了歷史研究哪些創新的研究方法，進而使得探究什麼樣的歷史議題成為可能或更加具可行性。同時欲指出使用資料庫時，若對資料本身的性質不了解，將可能對歷史解讀造成什麼樣的問題。總之，運用足以廣納史料並具有檢索功能的數位化資料庫，對歷史研究是個突破既有成果的契機，但是所引發研究上的陷阱也不少，值得研究者引以為戒。

* 國立臺灣大學法律學院特聘教授，中央研究院臺灣史研究所與法律所特聘研究員。

Digital Archives and Historical Studies

Tay-sheng Wang *

Abstract

National Taiwan University created three historical document databases, namely, Ming-Qing Archives, Tan-Hsin Archives, and Taiwan Colonial Court Records Archives. In the current paper, the process of incorporating the new ways to conduct research to recent historical studies to make new research issues possible is presented. Moreover, arguments with regard to the problems of misinterpretation due to ignorance of historical knowledge when applying databases for research are put forward. The availability of more comprehensive digital resources and searchable databases has brought about opportunities as well as challenges to historical research.

一、前言：數位化技術對歷史研究的衝擊

歷史學的研究之不同於小說創作，在於所為論述須立基於史料。參考史料的數量，將影響到產出的歷史論述之內涵，因為看得愈多，遺漏相關史實的可能性愈低，甚至對於某特定歷史／社會現象之屬常態與否的疑問，也可從發生頻率見其端倪。晚近資訊工程學上數位化技術，將數量眾多的史料納入單一資料庫中，並以層次分明的查詢系統，讓使用者得以盡可能檢索出欲閱覽的史料。這使得可參考史料之數量過少，不再成為歷史研究的重大障礙；反倒是如何從眾多史料中尋找出「適切」的史料，才是真正的考驗。

按史料之「適切」與否，一方面視所設定的研究議題而定，另一方面有賴於能解讀出史料本身的意涵。而如何設定一個可善用眾多史料的議題？是否讀得懂這些史料？已非數位化的資工技術所能解決，須回到歷史學研究的基本功，包括對整體歷史情境的掌握，以及對當時語言與各項制度的理解。切忌「看一個影，生一個因」（臺灣福佬話諺語）：單從數位化歷史資料庫看到一個影像／圖檔，就得出一個歷史論述，或者根本錯讀史料的文義，說出「無影之事」（出自〈淡新檔案〉內呈狀上用語，可用臺灣福佬話或臺灣客家話來唸）。

本文擬分別就臺灣大學所開發的「明清檔案」、「淡新檔案」、「日治法院檔案」等三個數位化的歷史資料庫，舉例說明數位化資料庫的出現，使得歷史研究增加了哪些新的研究方法，或使得探究什麼樣的歷史議題成為可能等等。同時欲指出使用這些資料庫時，由於對資料／史料性質的不了解而在歷史詮釋上可能造成的偏頗，或者資料庫本有其使用上的極限。

二、明清檔案資料庫：「你是我（研究者）的眼」

（一）在數量龐大的史料中依議題進行搜尋

首先擬討論的是臺灣大學所建置的「臺灣歷史數位圖書館」（<http://thdl.ntu.edu.tw>），這套系統收入眾多明朝尤其是清朝檔案的「明清檔案」（網頁上稱「明清臺灣行政檔案」）。其資料來源包括：1. 臺灣歷史博物館籌備處向中國第一歷史檔案館購得的軍機處檔案微捲。2. 電子資料庫：包括中研院史語所《內閣大庫》線上資料庫，以及國立故宮博物院出版的《清宮奏摺檔臺灣史料》光碟資料庫。3. 期刊：中國第一歷史檔案館主辦的《歷史檔案》。4. 已出版書籍：資料庫之檔案來源大多為已出版書籍，諸如《臺灣文獻叢刊》、《宮中檔乾隆朝奏摺》、《鄭成功檔案史料選輯》、《天地會》、《明清史料》、《乾隆帝起居注》、《雍正朝漢文諭旨匯編》等諸多書籍。5. 故宮博物院所藏之軍機檔、月摺檔、宮中檔原件，諸如《乾隆朝軍

* Distinguished Professor, College of Law, National Taiwan University.
Research Fellow, Institute of Taiwan History and Institutum Iurisprudentiae, Academia Sinica.

機檔》、《道光朝宮中檔》等。

重要的是，此一資料庫將這些歷史文件的內容全文輸入，讓逐字檢視與檢索成為可能。在沒有這項數位化之前，研究者很難在短時間裡將上述所有資料來源內的文書，全部看完，並從中依特定議題而擷取相關史料。如今此一資料庫已充當研究者的「眼」，而將其全部「看完」了，只待研究者指示該資料庫抽取出哪些它「看過」的資料。

為了善用此一資料庫已「看過」那麼多清朝官府內部檔案或官員所為的言說（discourse），筆者嘗試以之省視學界長期以來有關「傳統中國法是否『依法審判』？」的爭議，¹ 因為過去恐怕少有人、甚或沒人能同時看過那麼多清朝與審案有關的文書。不過，歷史本來就是出於某種認識目的，或某一特定視角底下的產物，故研究者須以其設定的認識目的或問題意識為基準，從眾多史料中擷取所需者。特別是像上揭有關「傳統中國法」的議題，由於其施行時間跨越中原數朝代，各朝代的情形可能不一，且因中華帝國所轄地域廣闊，各地域社會的情形也可能不一。筆者基於為臺灣社會探尋歷史淵源的研究目的，採取「將中國連結到臺灣」的研究取徑，故擬追問的是：就清朝（時間）之統治臺灣（地域），主政的清朝政府及其官員對於在臺灣的「司法」案件（亦即相當於今日司法的概念所涵蓋的案件），是否存在全然依照大清律例等官府成文規定做成裁斷的觀念和實踐？並以該實況作為清治下臺灣社會歷史經驗的一部分。

為處理這項議題，乃以「律」和「例」為關鍵字，從本資料庫檢索出分別多達4000多筆與1萬多筆有關審案的檔案。接著的目標是，從這些數量仍相當大的檔案中，挑選出與清治臺灣的審判有關連者，包括審案者曾來臺灣任官或擔任監督臺灣政務的上級官員、犯人為臺灣各地人士或曾來臺灣停留一段時間、案件發生地點為臺灣某地等。如何檢索出符合這項需求的檔案呢？若以（全文暨所有欄位：）「臺灣」加（標題：）「律」，可檢索出53筆，同樣的以「臺灣」加「例」，則檢索出382筆，但這樣是否已找出所需，其實並沒把握。或許選用特定的人名，例如曾在臺任官者的姓名，來進行查詢。不過，在未深入了解這些傳統中國官方文書的書寫模式或知識體系之前，其實並不容易確知哪些是用來檢索出「與清治臺灣的審判有關連」

者最恰當的「關鍵字」。

因此，在一項屬於「先導型」的研究計畫中，最後採用最有把握、但最費時之逐一閱讀方式，確認其是否「與清治臺灣的審判有關連」。亦即就前述用「律」和「例」為關鍵字所檢索出的幾千筆檔案逐一檢視，根據上述「與清治臺灣的審判有關連」的定義，判定可否作為分析的對象。受限於該計畫之執行期間僅一年，當找出500份／筆具備該關連性的檔案時，即不再繼續搜尋下去。² 不過，對此進行類型化分析時，因一份檔案可能包含數案，並可能符合了數個類型，以致可能會被重覆地計算，故所有被納入各個類型的檔案，在統計數據上總共為511件。以下即列出這些類型，以及特定類型在全體中所占比例，並詮釋其意涵。當然這僅僅是對前揭議題的初步分析結果，若欲求其詳盡，仍須另以較大規模的研究計畫探究之；惟無論如何，從前述研究過程多少可體會：使用數位化歷史資料庫進行歷史研究，不是想像中那麼簡單，按「關鍵字」的設定是很關鍵的，而這往往需要對所收錄檔案／文書具有「關鍵的」知識。

（二）就挑選出的史料進行整理與解讀

1. 全然依律例斷罪

這511件中，有381件為司法官僚全然依律例斷罪者，占多數（74.55%）。其可再細分：（1）只引用律例名稱斷罪，計有253件（49.51%），以及（2）引用律例之條文內容斷罪，共有128件（25.04%），數量較少。兩者詳言如下：

1.1 只引用律例名稱斷罪

在此類型，司法官僚僅一筆帶過律例的名稱，直接指出應依某項律例，判處犯人絞刑、流放或其他懲罰，較少先論及律例條文的詳細內涵，再說明適用該條文後所得到的斷案結果。

於茲舉兩例說明之。其一，嘉義縣民黃結從賊逆倫案，司法官僚認為：「查黃結從賊逆倫，罪大惡極，應依故殺期親尊長凌遲處死律，應凌遲處死。（本文底線為筆者添加，以下同）」³

其二，謝笑於臺灣新化番仔溝糾眾械鬥案，司法官僚認為：「查謝笑以內地民人，膽敢于海外重地倡議傳帖、糾眾械鬥、焚殺傷人，寔屬兇惡渠魁；謝笑應請照

1 日本學者滋賀秀三認為清朝官府的聽訟，本質上是一種「教諭式的調解」，審判中所謂的「普遍性判斷基準」是情、理、法，民事紛爭固為如此，即使在刑法的適用中，也須斟酌情理法。華裔美國學者黃宗智卻不以為然，其認為清朝縣官的行動是受到成文法律的制約的。另一位日本學者寺田浩明則指出，清代中國本來就不存在一種「秩序樣本」，可以作為基準、驗證現實與其是否相符，故官員是否依照作為客觀標準的法來斷案，乃是現在研究者出於當今的法律概念而設定的議題，對清朝的人們並無意義。筆者較贊同寺田浩明的看法，故將提問改為：「清朝政府或官員是否存在『全然依照大清律例等官府成文規定做成裁斷』的觀念和實踐？」並持否定的見解，不過亦承認在多數情形，司法官僚願意且實際上遵守皇帝所頒行的那些成文規定。相關的論述及評述，參見王泰升、堯嘉寧、陳韻如（2004：47-50）。

2 感謝國立臺灣大學法律學院郭詠華碩士生執行此項搜尋任務，找出這500份檔案並進行類型化工作。
3 〈罕獲匪犯，審明正法〉，故宮博物院所藏《乾隆朝軍機檔》，文件序號040544。另見國立故宮博物院圖書文獻處文獻科編輯（1982b：722）。本案所認定之事實為：嘉義縣民黃結於乾隆五十二年間，欲加入賊盜李七等人時，被黃結的胞叔黃相得知與責罵，因而心懷忿恨。在黃結前往投靠李七後，將黃相綁至賊營殺害，並夥同李七等人屢次攻城打仗，殺死義民一人。

臺灣聚眾散筍殺人放火造意為首者斬決梟示例，斬決梟示。」⁴

1.2 引用律例之條文內容斷罪

有別於上揭之僅對律例名稱一筆帶過，本類型則為司法官僚先引述律例的條文內容，接著再適用該律例斷案。亦舉兩案為例：

其一，諸羅縣民黃佛毆傷楊鈴致死案，司法官僚於判詞中先引述律例的規範內容為何，然後再明確表示依該律例的規定應如何處置。其謂：「黃佛合依共毆人致死以致命傷為重下手致命傷重者絞律，應擬絞監候，查律載，他物毆傷人限貳拾日，又例載，他物傷限外拾日之內果因本傷身死，情正事實者，方擬死罪，奏請定奪等語。今楊鈴於乾隆拾伍年正月初肆日被毆，至貳月初貳日因傷身死，係在辜限外拾日之內，相應照例聲明，奏請定奪。」（中央研究院歷史語言研究所編，2009b）

其二，彰化縣民張標、謝志等人糾結為天地會案，司法官僚於判詞中先引述謀叛的律例內容，再依照該律例來處置張標等人。其謂：「查律載，謀叛但共謀者不分首從皆斬。又例載，異姓人歃血訂盟焚表結拜弟兄，照謀叛未行律為首者擬絞監候，為從減一等，若聚眾至二十人以上，為首者擬絞立決，為從者發雲、貴、兩廣極邊煙瘴充軍各等語。此案前據該鎮道拿獲張標等三十六犯，審明將糾夥立會之張標等二十八犯擬斬，先行正法，聽糾入夥之林三元等八犯擬遣具奏，經臣部核覆，將林三元等八犯改擬絞候，並令將逸犯謝志等十三名嚴緝務獲，奏結在案。」（中央研究院歷史語言研究所編，1985a：856-858）

2. 不全然依律例斷罪

於相關的 511 件檔案中，有 130 件（25.44%），亦即大約整體的四分之一，司法官僚就其所認定之事實，不全然依律例來斷罪。不過其類型相當多，以下分別列出並舉一案例示：

2.1 原應依甲律例，但基於某些因素，改依乙律例斷罪

此類型的檔案共有 28 件。例如在彰化縣陳土葛被匪殺搶斃命案，司法官僚謂：「查例載：謀財害命、得財殺死人命造意首犯，斬決；……查該犯怙惡不悛，殘忍已極；律止斬決，尚不足以蔽辜。又況新開之路，商旅近已絡繹往來；竟敢肆行殺搶，若不嚴重重辦，誠恐效尤漸起，勢將荊棘道途。再四籌思，不得不變通辦理。是以擬請按照軍法斬梟，傳首犯事地方懸桿示眾，方足以昭炯戒而快人心。」（臺灣銀行經濟研究室編，1963b：65-69）在此明示原本依例應「斬決」，但「變通辦理」

而改依軍法「斬梟」。

2.2 該當律例中甲部分，卻依同條律例中乙部分斷案

此類型的檔案共有 6 件。例如司法官僚認為該犯行符合律例的主犯部分規定，但卻依同條律例中關於從犯的規定，來處置犯人。在洪則等人偷渡至臺灣案，司法官僚表示：「謝瑚、謝巖合夥承買私船，隨同攬載，僅照為從減等擬流，不足蔽辜，請從重照包攬過臺為首充軍例，發近邊充軍，到配杖一百折責安置。」（國立故宮博物院圖書文獻處文獻科編，1982a：658-660）

2.3 有可能該當甲律例，但逕論以乙律例之罪

此類型的案件共有 23 件。例如黃江珠等人於臺灣加入天地會案，閩浙總督伍拉納上奏乾隆皇帝的奏摺中謂：「黃循漏輾轉傳授，蔓延煽惑，實屬目無法紀，未便仍照僅止入會辦理，致無以示懲創，應請從重比照左道異端煽惑人民為首絞律，擬絞監候，請旨即行正法，以昭炯戒。」（中國人民大學清史研究所、中國第一歷史檔案館合編，1980-1988：424-427）在此其暗示有另一個處罰較輕的罪名，但為了「從重」還是處以另一較重之罪名。

2.4 引用並未被列入律例中的前案來斷案

此類型的案件只有 2 件，但亦值得注意。例如在陳步梯力守危城、兵單城陷案，閩浙總督邊寶泉就澎湖失守一事，在上奏光緒皇帝的奏摺中以「情有可原」，主張不應完全依照律例來斷案，其表示：「陳步梯……較周振邦情有可原，與鄭膺杰事實相若。可否援照成案，仰乞天恩量予末減，出自鴻慈逾格。」（戚其章主編，1989-1996：2394）

2.5 不依照律例所定之罰則，而是減輕懲罰

此類型的案件共有 30 件。例如在臺灣鄭文遠謀反案，刑部議奏：「臺灣叛賊鄭文遠等家口，應分別定罪」。得旨：「凡謀反大逆以及謀叛重罪，均無可宥，按律凌遲處死。正法之祖父、子孫、兄弟及同居之人、期親及伯叔父、兄弟之子不限已未析居，男十六以上不論篤疾廢疾，皆斬。但此案事起倉卒，遠隔海洋；親屬人等，有身在臺灣者、亦有身在內地者。若概從誅戮，情堪憫惻。除身在臺灣者依律正法外，其內地者從寬免死，解部給與功臣之家為奴。」（臺灣銀行經濟研究室編，1969：307-308）按依律文，應一概處死，但刑部主張人在內地者不依律處死，皇帝亦同意。

2.6 給予超過律例所定程度的懲罰

此類型的案件共有 34 件。例如在蔡牙等包攬民人、運送貨物偷渡至鳳山縣案，

⁴ 本案所認定之事實為：晉江縣民謝笑搬遷到臺灣新化番仔溝附近之四張犁後，與附近三塊厝的漳人黃添起衝突，後來演變為謝笑等泉州人與黃添等漳州人的械鬥案（中央研究院歷史語言研究所，2009a）。

司法官僚先表示：「查例載，閩省不法棍徒，充作客頭，在沿海地方引誘偷渡之人，包攬過臺者，為首發近邊充軍，舵工人等知而不舉，杖一百，徒三年，其偷渡之人照私渡關津律，杖八十，遞回原籍。」但其並不依此斷罪，而謂：「舵工蔡牙，聽從在逃之王金山載貨偷渡，該犯希圖漁利，膽敢包攬客民，附載過臺，殊屬不法。臺灣現當嚴查海口之時，該犯猶敢圖利偷載，若僅照棍徒引誘偷渡本例擬軍，不足示懲，蔡牙應如該提督等所奏，從重發往黑龍江給與披甲人為奴……。」（中央研究院歷史語言研究所編，1985 a：861-863）

2.7 該當某律例所定之罪，但是基於某些理由而不予議處

此類型的案件共有 4 件。例如在諸羅縣之翁雲寬縱容莊佃羅瓦焚搶案，司法官僚對於已構成誣陷之行為，認為應不予議處，而謂：「王最、王董、王因、林送告詞雖有失實，但當家物猝被焚搶，人口失散，情急懷疑，且因被翁雲寬莊佃羅瓦、翁柔等搶殺，隨控其莊主，又因目不識丁，致被做詞人添寫，尚屬情有可原，均請免其置議。」（中央研究院歷史語言研究所編，1985 b：1987-1989）

2.8 無律例明文規定該如何處置，比附援引既存之法條來斷案

此類型的案件共有 3 件。例如在臺灣縣郭洸侯等合夥謀逆案，覆審之大學士臣穆彰阿等人認為，郭洸侯等人並非如同之前審擬此案的知縣所言犯下謀反等罪行，而應比附某「例」斷其罪，其表示：「律載：斷罪無正條，援引他律比附加減定擬各。……郭洸侯應比依直省刁民假地方公事強行出頭、逼勒平民約會抗糧、聚眾聯謀、或果有冤抑、不於上司控告、擅自聚眾至四、五十人以上、為首斬立決例，量減一等，擬杖一百、流三千里，仍從重發往新疆，酌撥種地當差。……」（臺灣銀行經濟研究室編，1963a：148-158）不過，由於律文本身已允許比附「例」來斷罪，此類型亦可謂屬於前述「全然依律例斷罪」。但就「不全然依律例斷罪」之件數，縱令將此類型排除，仍有 127 件（24.85%）。

在此可對上述類型化分析的結果，簡單地做個解讀，詳細詮釋尚有待他文為之。按在審斷犯罪的理由中，不少案件明白表示不遵從大清律例，可見傳統中國法中並無「須全依官府規定斷罪」的觀念與實踐，故與今日繼受自西方的法治觀念，以及司法審判須依立法機關所制定的成文法律的規定（參見憲法第 80 條）有別。兩者為相異的兩種法律生活方式。但是，從多數案件的審斷理由中，仍揭示或訴諸大清律例之內容，卻也證實司法官僚相當重視且實際上引用大清律例等官府規定斷罪，並顯現出皇帝及其官僚所頒行的成文規定，在整個帝國官僚體系內仍有其一定的**重要性**，足以某程度地約束當時的政治上作為。雖然承審官員可不遵從律例上之規定，但總需要多花費一番功夫來說服上級官員及皇帝。

這般的審斷案件，來自漢朝以降中原政權之兼採儒、法兩家，故以儒家倫理要

求及維護皇帝統治權威作為決定政務的判準。於是，審斷案件的終極關懷，乃是個案當事者利益的調和，以及朝廷或官府治理上的需要，而非如現今法制之以「發現」並「執行」成文規範為唯一要求。

（三）運用此資料庫時須面對的局限性

司法官僚亦有可能依情理而斷案，但因文書中完全未提及律或例，以致從上揭搜尋中，當然無法被檢索出來，也就是不在該 511 件中。若改用「情理」、「情」、「理」來搜尋，則可能因出現太多筆數而成為無用的檢索。其實，作為研究者此時可考慮使用其他論著或史料來源，不需一味執迷於資料庫。例如清朝著名的幕友汪輝祖，即曾經以人情、義理來處理缺乏可資適用之律例的案件，且這般的「捨律引禮」亦得到當時司法官僚的認同。（汪輝祖，1971：17-19；張偉仁，1989：21-22）

另一項宜注意之點是，由於本資料庫的資料來源之故，上述 511 件絕大多數屬於清朝中央政府所為的文書，而在當時僅僅重案方審轉至中央政府進行審判的制度下（王泰升，2009a：66-69），這些案件乃是所有進入官府衙門的案件中的**少數**。而當社會上發生兇殺、竊盜等刑案或財產上紛爭時，又可能只有少數或部分進入官府衙門。所以，若議題設定在探究清治時期臺灣社會的經驗，則不能僅僅憑藉此資料庫所收錄之史料，亦即不能只依靠這套資料庫。下述屬於清朝地方政府史料的淡新檔案，因此有登場的必要。總之，使用歷史資料庫時，須運用歷史學的知識，理解其所收錄的文書在當時整個法制乃至社會運作上的位置。

三、淡新檔案資料庫：檢索系統與原分類標準相互參照

（一）淡新檔案原分類表之作成

於臺灣學界已相當聞名的淡新檔案，乃是收納清朝自乾隆 41 年（1776）至光緒 21 年（1895）間，淡水廳、臺北府及新竹縣衙門內辦理政務所留下的檔案。該檔案係日治時期由司法機關承接後轉贈臺北帝國大學文政學部，戰後由臺大法律系戴炎輝教授命名及主持整理工作，將檔案內之文書分為行政、民事及刑事 3 門，門下並分類、款、案，例如 17325 案，代表是第 1 編、第 7 類、第 3 款、第 25 案，於案號下方還會記載該案的第幾件。1986 年，戴教授將此檔案移交臺大圖書館特藏組收藏，清點實得總案數為 1,143 案，共 19,281 件。數量最多的時期為光緒年間（1875-1895），共有 838 案（73.31%），其次為同治年間（1862-1874），共有 161 案（14.08%），兩者總計 999 案、占全檔的 87.40%。是以，淡新檔案主要記載著十九世紀，特別是其下半葉，北部臺灣地方衙門的運作實況。對於研究臺灣清治時期政

治、法律、經濟、社會等各方面，提供珍貴的第一手資料。也是世界著名的傳統中國縣級檔案，可作為研究傳統中國法制及其司法審判的重要憑藉（國立臺灣大學圖書館，2009）。

這些數量相當多、原僅是清朝地方衙門內存檔及作為辦案參考的文書，本身可能並沒有一套精密而有系統的編排方式，為便於今日使用，必須進行整編。換言之，現今使用此等文書的人，都已接受過現代意義的法律（以下簡稱「現代法」）的訓練或影響，故有必要將這些由受傳統中國法訓練或影響的人所書寫的文書，進行某種現代人可理解的分類和編排，以利其查詢。因此，肩負整編重責的戴炎輝教授只好以現代人所熟悉、卻不存在於清朝書寫當時的現代法概念，進行第一層次的分類，亦即行政、民事、刑事三個「門」。但是愈往下看，愈注意到該文書內所使用的文字，有些是現代法所無而屬於傳統中國法的用語，例如「錢債」等出現在「類」、「胎借」或「抗吞」等出現在「款」的名稱上。有時亦使用現代法上詞彙，作為「類」或「款」的名稱。更大的問題是，有些文書的內容，在現代法上實在找不到相對應的詞彙來描述。此時戴炎輝教授只好自己想個最恰當的詞彙，當作「類」或「款」的名稱。但在沒有共通知識或統一的定義下，使用者對於某詞彙的理解，可能與身為整理者的戴教授不同，以致對某「類」，尤其是某「款」之內容的想像，可能與戴教授所歸入者有一定的差距。

在此以戴教授所提出的「淡新檔案分類表」（以下稱「戴氏分類表」），當中的「民事」門為例來說明。其分類如下：

第二門 民事（共 224 案）

第一類 人事（共 13 案）

第一款 失蹤

第二款 結婚

第三款 離婚

第四款 收養

第五款 監護

第二類 田房（共 138 案）

第一款 租借

第二款 抗租

第三款 霸收

第四款 霸占

第五款 爭界

第六款 爭財

第七款 公業

第八款 用水

第九款 抄押

第三類 錢債（共 68 案）

第一款 買賣

第二款 典當

第三款 胎借

第四款 借貸

第五款 委寄

第六款 討物

第七款 抗吞

第八款 抗算

第九款 匿契

第四類 商事（共 5 案）

第一款 郊行

第二款 合股

第三款 倒閉

第四款 侵權

上述各「類」或「款」的名稱，不同於已繼受西方法／現代法的臺灣現行民法典上分類用語，且多數情形也非大清律例上的用語（清朝並無相當於今之民法典者）。使用者欲據此分類查詢出自己所需之史料時，只好望文而揣摩其意；除非自

已花大量的時間閱畢全檔，方無遺漏之可能。但依臺大圖書館特藏組估算，淡新檔案原件展開後，大約將近有 6000 公尺。如果每天讀 10 案，都不懈怠，大概要花 5 年的時間才能全部看過一遍。

（二）本資料庫提供新的檢索方式

經數位化的淡新檔案資料庫（www.darc.ntu.edu.tw），首先發揮其「你是我（研究者）的眼」的特色，將所有檔案全部「看完」了，且就檔案內容做成全文電子檔（同時排印出版，惟刑事門尚未完全做成），不但解決原件字跡有時不易辨識的困擾，且使得資料庫的查詢功能更加強大。於是，可運用各個專門詞彙、人名、地名、年代等進行查詢，以檢索出各項研究議題所需要的案件。

例如，對於今天臺灣社會中關於「租賃」法律關係之演變有興趣，而欲了解清治時期官府或一般人民對相當於今之租賃有什麼樣的法律觀念與實踐。⁵ 研究者可能會先參考戴氏分類表，並直覺地認為「田房」類、「租借」款，會有所需的相關史料。不過，放在其他「款」內的案件仍可能與該議題相關。在過去，只能再一款一款地查，但有了資料庫後，就可採用較方便的檢索方式。

對淡新檔案已有一定認識的研究者，可至淡新檔案資料庫，先以「租借」作為關鍵字進行檢索，得出 180 筆文書在內容上有「租借」兩字者；所謂 1 筆即是 1 單張／圖檔，但不等於 1 件。從檢索後各筆所屬的案號，可發現以 22101 到 22107 占多數，亦即戴氏分類表中第 2 門（民事）、第 2 類（田房）、第 1 款（租借）的第 1 案到第 7 案；但重要的是，還可找出其他具有「租借」兩字的單張／圖檔，且顯示出所屬案號為 22415, 22604, 22605, 17305, 17333。這表示戴氏分類表第 2 門第 2 類（田房）第 4 款（霸占）第 15 案、第 2 門第 2 類（田房）第 6 款（爭財）第 4 案和第 5 案，以及屬於第 1 門（行政）的第 7 類（撫墾）第 3 款（隘務）的第 5 和第 33 案，也有載明「租借」兩字的文書，值得研究者查閱。

為求周延，可再輔以「租佃」作為關鍵字進行檢索。這次可得出 118 筆，各筆所屬案號，以 224xx 和 225xx 最常出現，意味著第 2 門（民事）第 2 類（田房）的第 4 款（霸占）和第 5 款（爭界），有不少文書中載有「租佃」的案子。且這 118 筆單張所屬案號之「類」或「款」相當多元，還有屬於第 1 門（行政）的案號，凡此皆可能與研究主題相關。甚至，再以另一關鍵字為檢索，或仍有其必要性。

從如上的舉例說明可知，今日被歸為「民事」的「租賃」法律關係相關史料，會出現在戴氏分類表中的「行政」門案件中。此再次證實淡新檔案內這些史料被書

⁵ 這類臺灣在清治時期法律觀念與行動的探問，是筆者在臺大法律系開課，要求修課者兩三人為一組撰寫淡新檔案報告時，經常被提出的問題意識。

寫時，亦即清治時期，並沒有行政、民事這樣的現代法上分類概念。也因此，前揭「明清檔案」若稱為「明清臺灣行政檔案」，則「行政」兩字，可能引發今之研究者的錯覺或誤會，以為該檔案中並無與今之「司法」相關的檔案，故不如刪除之，而僅稱為「明清臺灣檔案」。⁶

（三）運用此資料庫時須面對的局限性

淡新檔案內有眾多文件，性質上屬於訴訟文書，故有因此引發的局限性問題。目前本資料庫所檢索出的各筆，如上所述，僅僅是 1 個圖檔，其可能是一整份文件，也可能只是 1 份文件的部分。縱令是 1 份文件，其文字內容的意義，必須放在一整個案子，亦即前後各文件的脈絡來看。很重要的是，作為爭執犯罪行為是否存在或財產應如何分配時的文書（例如呈狀），其內容只是書寫者／表意者，包括原告、被告、證人、差役，乃至大老爺（知縣或同知）一方的說詞，其是否等於該事件發生的真實狀況，可能案發當時都沒人有把握，更何況今之研究者。我們能從淡新檔案內文書看到的「真實」是：這個人，在這時候，講這樣的話，官員採取這樣的作為。因此，研究者勿將圖檔上的文字，直接當成是針對當時社會的事實或現象所為的客觀描述。不過，發言者為了爭取別人相信其說詞，當然會盡量配合當時人們所存在的想法，故從文書的文本內容，詮釋出當時人們存在著某種觀念，則是可接受的研究取徑；另外，藉由眾多文書，描繪出當時官府審案的過程，也是可行的。⁷

此外，與前揭明清檔案一樣，淡新檔案只記載進入官府的犯罪事件或紛爭及其處理的狀況，不宜直接認為當時社會上凡是發生這類事情，皆是如此處理。將屬地方政府層級審案文書的淡新檔案，與大多數屬中央政府審案文書的明清檔案相互比較，並加以詮釋，應該是令人感興趣的議題。

再者，資料庫本身的檢索方式，也有其一定的局限性。承前所述，淡新檔案內文書宜以「案」為單位來閱讀，但目前資料庫檢索結果都是呈現 1 個單張／圖檔，而未呈現構成 1 案的數個單張／圖檔。只要本資料庫增加以「案」為單位，甚至可以「款」為單位之「瀏覽」功能，應即可彌補這項缺憾，讓使用者可以更多元的方式「看」淡新檔案。不然，有需要的使用者，仍可搭配以戴氏分類表為準之刊出全文的出版品，或從微捲複製得的紙本，來閱讀整個「案」或「款」的內容。

⁶ 遠流出版公司已於 2009 年 10 月 28 日舉辦《明清臺灣檔案彙編》110 鉅冊的新書發表會。

⁷ 對上述研究方法有興趣者，可參見陳韻如〈帝國的盡頭：淡新檔案的姦拐故事與申冤者〉（國立臺灣大學法律系碩士論文，2004）、堯嘉寧〈官府中的紛爭解決——以淡新檔案觀察相當於今日新竹市之區域之案件〉（國立臺灣大學法律系碩士論文，2005）。

四、日治法院檔案資料庫：內容精彩但有使用上的門檻

（一）日治法院檔案的整編及資料庫的建置

相對於淡新檔案，「日治法院檔案」在臺灣尚未廣為人知。經由臺大法律學院與臺灣大學共同進行數位化而完成的「日治法院檔案資料庫」，甫於 2008 年 9 月間，始向國內外學術研究者公開。日治法院檔案係自 2000 年後，始陸續為包括筆者在內的幾位臺灣法律史研究者所發現，並予以命名。其主要內容為臺北、新竹、臺中、嘉義等 4 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。不過，日治時期臺中地院刑事判決原本係保存於司法官訓練所。此外尚有少量由臺灣高等法院、高雄和花蓮地方法院、澎湖地政事務所等收藏的日治時期司法文書，至今尚未完成整編工作，故還未被納入日治法院檔案資料庫。

完成所剩整編工作後的日治法院檔案，所涉時間始自 1895 年迄於 1945 年，所涉地域為日治時期臺南地方法院轄區以外幾乎所有的法院管轄區域。⁸ 其相當完整地記錄日治時期現代型法院在臺灣的運作實況，可用於探究臺灣日治時期法律、政治、經濟、社會等各方面的發展，故重要性足以和記載日治時期行政權運作實況的臺灣總督府檔案中「臺灣總督府公文類纂」相提並論（王泰升，2009b：171-180、194-196）；同時也是日本帝國史、東亞法律現代化、世界殖民地統治史等研究所不可或缺的參考資料，堪稱世界級的珍貴史料。在開放申請使用的 1 年內，已有來自臺灣及日本、韓國、美國等國內外 70 餘名學術研究者（至 2009 年 9 月為止），獲准使用該資料庫。

日治法院檔案所收納的文書，數量上非常龐大，姑且不算有待後續補充者，即已多達 5,645 冊，以數位相機拍攝得 2,295,000 張圖檔，若以一般的 DVD 收藏來估算約需 287 張。日治法院檔案資料庫因此須依一份「架構表」，將每一冊歸入特定的位置。由於所收錄的日治時期司法文書是現代法施行後的產物，故這份具有三個層次的架構表，係依照當時與現今司法人員所熟悉的分類標準，亦即現代民事訴訟法和刑事訴訟法的用語所做成的，並無不屬現代法產物的淡新檔案需自創分類標準的問題。

日治法院檔案資料庫即提供依照此項架構表，「瀏覽」這些文書圖檔的功能。由於收藏於法院的原檔內文書，不乏因年代已久而破損或無法掀開者，若欲修復即

需鉅額經費與相當長的時間，故置於資料庫者係僅就現狀拍攝所得圖檔。且日治法院檔案內文書所使用的文字，乃是連現今一般日本人都不得看得懂的戰前日文，有時還以難以辨識的草書來書寫，因此如同淡新檔案般做成全文電子檔，實際上並不可行。就好比同性質的臺灣總督府公文類纂，在建置資料庫時亦未做全文輸入。不過，日治法院檔案資料庫原本即將使用者設定為學術研究者，而非一般民眾，故宜自我評估為具備閱讀日文史料的能力，再向資料庫管理機構申請，以取得進入資料庫的專用帳號。只要能進入本資料庫，透過其瀏覽功能，全世界各地的學術研究者可舒服地在電腦前，逐冊、逐案、逐張閱覽原檔文書，不必再經繁瑣的申請、到位於各處的法院庫房裡，蹲著看這些難讀的司法文書。

同時，本資料庫也提供一個相當便利的「查詢」系統，以檢索所需史料。就日治法院檔案中最重要，且在全檔中占約 64% 的民事、刑事判決原本和公證書原本，可查詢到每一案，而不僅僅是每一冊。按民刑事判決和公證書的每一案，已以結案年代、事件號、裁判或公證事由，以及原告、被告、訴訟代理人、檢察官、法官、書記官等人名作為編目資料（metadata），故在查詢系統中以這些文字為關鍵字，即可檢索出具有這些編目資料的案件。例如以某甲這個人名為關鍵字，即檢索出編目資料中有某甲這個人的每一案。就查詢的結果，還可用「後分類」功能，依「所屬類別」、「相關人員」、「結案年代」、「相關事由」、「承辦法官」等做進一步的分類。當檢索出的筆數過多時，後分類功能尤其好用。藉此查詢系統，研究者應可順利檢索出與研究議題相關的判決與公證書。

（二）可運用本檔案或其資料庫突破研究瓶頸

善用日治法院檔案內司法文書的特性，以及本資料庫所提供的強大搜尋力，就算舊議題也可能做出創新的研究結果。例如，日治時期國家法／司法當局如何處置政治犯？從過去抗日史觀當道，到後來強調臺灣主體性的年代，都已經有人研究過這項議題。但是能夠談的，就是一些用以處罰政治犯的法律條文內容，頂多是增加幾則法院相關的判決要旨，或是像 1923 年臺灣議會事件（俗稱「治警事件」）等極少數重大政治案件的審理過程與結果。但是，運用包含了地方法院所有刑事判決的日治法院檔案，或進一步運用該資料庫查詢系統來進行研究，將可能有令人耳目一新的發現。

在向來的臺灣史中最常被提出來、用以批判日本殖民政權的是 1898 年制定的《匪徒刑罰令》，甚至抗日史觀者可能就把日本官方文獻中所稱的「匪徒」，直接改稱為「抗日義士」（劉彥君，2006：5）。筆者曾以《匪徒刑罰令》中的匪徒罪，係以「不問其目的為何，凡以暴行或脅迫為達成其目的而為多眾結合」作為構成要件，故肆行結夥搶劫的非政治性的盜匪（集團）也可能該當此一匪徒罪，並舉出一

8 於日治前期屬於臺灣三大地方法院之一的臺南地方法院，由於整個日治時期的檔案在戰後已佚失，無從納入日治法院檔案中，此外日治時期宜蘭亦曾設地方法院分院，然今之宜蘭地方法院亦找不到日治時期的檔案。

則當時覆審法院的判決要旨，而認為不宜將所有違反《匪徒刑罰令》者皆視為「抗日英雄」（王泰升，1999：272）。在日治法院檔案被發現並進行整編後，臺大法律學院劉彥君碩士（2006：52-82）檢視了該檔案內臺中地方法院與《匪徒刑罰令》相關的眾多刑事判決原本（其寫作當時臺北地方法院刑事判決原本尚未整理完畢），而在論文中提出以下的重大發現：

1. 匪徒罪與強盜罪不易區分，但法院對匪徒罪構成要件採限縮性的解釋

在日治法院檔案臺中地方法院刑事判決中，所謂「訴盜判匪」，亦即檢察官以強盜罪起訴、法官以匪徒罪定罪者，有 12 件。其中有 6 件係匪徒刑罰令制定後約一個月時為判決，故可能是起訴時尚無此令；有 5 件是在此令施行約一年間為判決，則可能是司法界對於構成要件模糊的匪徒罪如何認定尚在摸索中。

同一批資料內，所謂「訴匪判盜」，亦即檢察官以匪徒罪起訴、法官以強盜罪定罪者，多達 94 件。依警察機關之移送而由檢察官認定為係「匪徒」的強盜犯，由於法院認為所謂的「多眾結合」，須是一個緊密的團體，而非臨時、一時衝動的犯罪結合，故數名強盜犯若不足以稱為「緊密的團體」即不構成「匪徒罪」。由此亦可知，當時司法審判權之運作並非從屬於行政部門。

2. 遭匪徒罪定罪者其犯行大多數與抗日無關，且大多數的被害人係臺灣人

從 1898 年至 1910 年，臺中地方法院判處匪徒罪之案件計 1,089 件。依所認定之犯罪事實，屬於「單純武裝反抗」，即指「直接與殖民地官方相關連」者有 167 件（15.33%）；屬於「非武裝反抗」，即指「純粹侵害個人法益而與日本官方無直接關連」者有 822 件（75.48%）；屬「複合犯行」，即指「包括單純武裝反抗與非武裝反抗」者有 100 件（9.18%）。故遭匪徒罪定罪者，約有四分之三可能是與政治無關的盜賊。

依《臺灣總督府統計書》就 1900 年至 1902 年「土匪被害案件」所為統計，於臺中地區，被害人計 766 人，其中臺灣人（「本島人」）有 708 人（92.42%），日本人（「內地人」）有 58 人（7.57%）。雖然這份統計可能是以警察部門所認定的匪徒罪為準，其部分案件最終可能不為法院認定為係觸犯匪徒罪，但考量這項可能的誤差，還是可說大多數的匪徒罪受害人乃是臺灣人。

基於以上的分析，於今仍應譴責日本殖民統治當局以《匪徒刑罰令》「陷民於匪」，並科以極刑，實為不當。但是，將被法院判處匪徒罪者一律視為「抗日英雄」，亦屬扭曲歷史事實，同樣應受批判。

由上例可知，國家法的內涵，須靠法院（而不是行政機關的警察或者為國家訴追犯罪的檢察官）在眾多司法個案中進行法律解釋適用來具體化。故日治後期國家

法究竟如何處置許多臺灣人政治異議者？也有待檢視法院的判決始知。在發現了法院裁判史料後，接著須面對的是：如何在數量龐大的刑事判決原本中，找到關於某位政治異議者的判決？而此正是本資料庫「查詢系統」發揮作用之時。臺大法律學院王祥豪碩士，即先運用史學界既有研究成果，以向山寬夫所著、1 千多頁的《日本統治下的臺灣民族運動史》一書中人名索引裡，所列出的數百位政治異議者的姓名，當作「關鍵字」，在 2009 年公開的日治法院檔案資料庫中，檢索出與這些人名相關的刑事判決原本，以進行多角度的分析（王祥豪，2009：7-8；向山寬夫，1987：86-92）。

內容豐富的日治法院檔案，絕不僅僅運用在探究日治時期臺灣法律史而已。還可結合前述淡新檔案、戰後臺灣司法機關檔案等，做跨時期的法律史研究，亦可與同時期的日本、韓國、中國的法院史料做跨地域的比較研究。同時也可被運用於社會史、經濟史、家族史等，或諸如法律現代化或殖民地法制等全球性議題的探討（王泰升，2009b：180-185；王泰升編，2009）。

（三）運用此資料庫時須面對的局限性

日治法院檔案所收錄者，乃是依現代國家司法權之運作而產生的文書。尤其民、刑事判決原本，更是依現代民、刑事訴訟程序所作成的文書。在現代法治國依法裁判的要求下，當時的司法者是將一般人所熟知的「社會事實」，配合法律構成要件用語，而轉化成用法律語言來描述的「法律事實」。且不論是法律事實或社會事實，均是法院依據訴訟雙方的主張及舉證等程序而認定的「事實」，尤以強調當事人進行主義的民事訴訟程序為然。是以研究者可能需另從其他管道或史料，探求非國家法面向的事實；由民間書寫甚或當事人自己書寫的史料，例如回憶錄或日記，或可用來與官方史料交叉比對。上舉碩士論文〈日治後期國家法對政治異議者的處置〉，即同時參考政治犯張深切與簡吉的日記（王祥豪，2009：133-142）。

例如，上舉法院判處匪徒罪之判決原本，僅表達出法院認為某人有構成匪徒罪要件的行為，此人是否真的有集結成團體以行搶的行為是另一回事。所以前述論述的重點在於，倘若法院所認定的事實係某一群人並未集結成團體，就不認為構成匪徒罪；至於真實是什麼？可能真的是集結成團體，但法官為了放被告一馬而認定為沒有此情形，也可能真的沒此情形。又例如，公證書表現出雙方成立某項法律關係，也只不過是在國家法上是如此，雙方是否有其他的「相互瞭解」，仍不得而知。

文書本身所引發的另一個使用上門檻，就是必須具備閱讀戰前日文的能力，以及對於臺灣日治時期法制具有一定程度的了解。此與資料庫的使用者設定為學術研究者，有互為因果的關係。若研究上有需要使用日治法院檔案資料庫，則必須克服

這雙重的障礙，包括與具有這項專業能力者共同研究。

資料庫的功能本身也有局限性。目前日治法院檔案資料庫所使用的是華文，但是為遷就文書本身的文字，在編目資料中常出現戰前日文的繁體漢字，例如「調停」、「會社」等。因此，將來宜本於「調停」為今之「調解」、「會社」為今之「公司」等對照關係，讓輸入其中一個詞即可檢索所需的史料。甚至可考慮同時建置日文或英文的使用環境，以方便不諳華文、但熟悉戰前日文的使用者，或在編目資料上再加上例如「適用法條」等，以增加更多的檢索方法。

五、結論

上述明清檔案、淡新檔案及日治法院檔案，皆透過數位化技術而成為研究者的電子「眼」，使研究者可以掌握數量龐大的史料，並可藉由編目資料所提供的各種資訊，檢索出特定研究議題所需的相關史料。前兩個資料庫已就史料所載做成全文電子檔，尤其有利於史料之搜尋。但話說回來，針對某項議題進行檢索時，哪些是恰當的「關鍵字」呢？這個問題本身就是個關鍵點，需要對文書本身的書寫模式或知識體系有一定的了解方可。

運用明清檔案，可以初步地發現清朝官員在討論應如何審斷案件時，雖在多數情形會引用大清律例之名稱或進一步述及條文內容，但不少仍是不完全依照大清律例的成文規定而為審斷。仍宜留意的是，由於被納入該資料庫的文書多數屬中央政府檔案，故前述觀察尚不足以代表臺灣清治時期的所有官府承辦案件，更不能推論及所有當時所發生的犯罪事件或紛爭皆如此處理。

不過，淡新檔案資料庫透過全文電子檔及功能良善的檢索系統，恰可彌補使用者可能對戴氏分類表中分類基準不熟悉之憾。不過此資料庫所檢索出的僅是文件的 1 單張，顧及淡新檔案之為訴訟文書的特性，研究者宜檢視 1 整案，而非 1 件或僅 1 單張，此時戴氏分類表上的案號，實有助於研究者檢視全案，甚至 1 整款的案件。

就日治法院檔案資料庫，可依類似當今法院所用的分類，來瀏覽日治時期各類司法文書，亦可使用包括「後分類」等設計在內的查詢系統，來檢索所需史料，進而突破過去研究上的瓶頸。不過，由於文書本身的特性，使用者須跨越語文及法制知識的雙重門檻，且勿忘其與前述兩檔案資料庫的文書同樣是屬於官方的表達。

總之，運用足以廣納史料並得有效搜尋的數位化資料庫，對歷史研究是個突破既有成果的契機，但是所引發的研究上陷阱也不少，值得研究者引以為戒。

參考文獻

- 中央研究院歷史語言研究所編，1985a，〈奏為續獲天地會三犯一律斬決餘犯從重議處事〉，《明清史料》，戊編第 1 冊，北京：中華出版社。
- 中央研究院歷史語言研究所編，1985 b，〈為遵旨議奏事〉，《明清史料》，己編第 2 冊，北京：中華出版社。
- 中央研究院歷史語言研究所，2009a，〈移會為福建巡撫奏拿獲械鬥首犯事〉，內閣大庫線上資料庫（編號 0000100004-0141436），臺北：中央研究院。
- 中央研究院歷史語言研究所，2009b，〈福建巡撫為黃佛毆死人命擬絞監候〉，內閣大庫線上資料庫（編號 0000100017-0054309）。臺北：中央研究院。
- 中國人民大學清史研究所、中國第一歷史檔案館合編，1980-1988，〈為審明續獲天地會餘匪並究出復犯劫竊各情分別正法定擬恭摺奏聞事〉，《天地會》，第 5 冊，北京：中國人民大學。
- 王泰升，1999，《臺灣日治時期的法律改革》，臺北：聯經。
- 王泰升，2009a，《臺灣法律史概論》，臺北：元照。
- 王泰升，2009b，〈日治法院檔案的整編與研究〉，《臺灣史研究》，16（1），頁 169-201。
- 王泰升編，2009，《跨界的日治法院檔案研究》，臺北：元照。
- 王泰升、堯嘉寧、陳韻如，2004，〈「淡新檔案」在法律史研究上的運用〉，《臺灣史料研究》，22，頁 30-71。
- 王祥豪，2009，《日治後期國家法對政治異議者的處置》，國立臺灣大學法律學研究所碩士論文。
- 向山寬夫，1987，《日本統治下における台湾民族運動史》，東京：中央經濟研究所。
- 汪輝祖，1971，《病榻夢痕錄》，臺北：廣文書局。
- 國立故宮博物院圖書文獻處文獻科編輯，1982a，〈為遵旨嚴審從重定擬恭摺會奏事〉，《宮中檔乾隆朝奏摺》，第 67 冊，臺北：國立故宮博物院。
- 國立故宮博物院圖書文獻處文獻科編輯，1982b，〈為拏獲匪犯審明正法恭摺具奏事〉，《宮中檔乾隆朝奏摺》，第 71 冊，臺北：國立故宮博物院。
- 國立臺灣大學圖書館，2009，淡新檔案，檢索日期：2009 年 10 月 20 日。
網址：http://www.lib.ntu.edu.tw/CG/resources/Taiwan/taiwan_ds1.htm

戚其章主編，1989-1996，〈閩浙總督邊寶泉奏陳步梯危城力守兵單城陷請量予未減片〉，《中日戰爭》，第3冊，北京：中華出版。

張偉仁，1989，〈良幕循吏汪輝祖：一個法制工作者的典範（上）〉，臺大法學論叢，19（1），頁1-49。

劉彥君，2006，《強盜或抗日——以日治法院判決中的「匪徒」為核心》，國立臺灣大學法律學研究所碩士論文。

臺灣銀行經濟研究室編，1963a，〈大學士穆彰阿等覆奏審擬郭沆侯一案供情摺〉，《臺案彙錄》，甲集第2冊，臺北：臺灣銀行。

臺灣銀行經濟研究室編，1963b，〈訊明購獲謀財害命兇犯擬照軍法從重就近懲辦及賞格銀兩可否入冊報銷稟（呈、咨、移）請示遵（核覆）辦理由〉，《臺案彙錄》，壬集，臺北：臺灣銀行。

臺灣銀行經濟研究室編，1969，〈雍正元年癸卯（一七二三）春二月壬戌（十二日），召董永艾來京，以秦國龍為福建按察使（由御史遷）。……〉，《東華錄選輯》，第2冊，臺北：臺灣銀行。

史料整體分析工具之幕後—— 介紹「臺灣歷史數位圖書館」的 資料前置處理程序

陳詩沛*、杜協昌**、項潔***

摘要

「臺灣歷史數位圖書館」（THDL）是臺灣大學近年建置的一個大型全文史料數位圖書館，共含近八萬件的第一手臺灣史文獻，累積的全文字數已達約一億五千萬字。本文會對 THDL 納入史料時進行的資料前置處理方式做說明，並指出其獨特之處。一般數位圖書館在納入資料時，往往不會假設這些資料之間有顯著的關聯，所以在納入分批建置的新資料時，除了重新製作檢索索引（re-index）之外，並不需太多的前置作業。然而，為了讓臺灣史研究者能更有效地利用與分析 THDL 中的大量史料，我們發展了一系列的史料分析工具，這些工具需要對史料作整體性的分析，所以 THDL 對資料的處理方式遠比一般的數位圖書館複雜，尤其在輸入一批新資料的時候，需要與現有資料作整體、精細的前置處理，預先建立全體資料的分析資訊與資料之間的關聯性，讓研究者有能力深入大量的史料，進行分析與研究。這個前置作業程序也反映我們打造 THDL 的原則，即是一個好的數位圖書館不應僅是一個資料倉儲，而應是資料與分析工具之間無縫的結合。

* 臺灣大學資訊工程系博士候選人。

** 臺灣大學資訊工程系博士後研究員。

*** 臺灣大學資訊工程系特聘教授。