

# 數位時代的臺灣學

文·圖片提供／謝國興  
(中央研究院臺灣史研究所研究員兼所長)

## 以中研院臺史所數位系統建置與開放應用為例



▲臺灣史檔案資源系統，2008年上線。

數位科技的進展，改變歷史文化資產保存方式，更影響人文學術領域研究發展。因應數位時代的到來，中央研究院臺灣史研究所（以下簡稱臺史所）自籌備處設置以來，即參與中研院與國史館臺灣文獻館的日治檔案數位化計畫，完成臺灣總督府公文類纂、專賣局檔案，以及臺灣拓植株式會社檔案等三大日治時期檔案的數位典藏工作。並自2004年起，加入數位典藏國家型科技計畫一期計畫，分別與臺灣省諮議會合作進行臺灣省參議會、臨時省議會及臺灣省議會等三時期所有的議政檔案，以及挑選國立臺灣圖書館所接收之臺灣總督府圖書館、南方資料館和戰後山中樞館長向臺北帝國大學日籍教授購置之藏書等三大日治舊籍館藏，進行數位典藏計畫，以數位典藏方式取得，累積相當豐富的臺灣研究資源。

### 數位典藏系統建置與推廣多樣性應用

2007年起，配合數位典藏二期計畫，強調數位成果開放公眾取用與深化人文研

究應用之總體目標，臺史所檔案館以「先整合多樣性典藏、再加值獨特性史料」為數位系統發展策略。在整合多樣性典藏方面，先於2008年完成「臺灣史檔案資源系統」之建置與上線，以遵循國際檔案界通用之尊重全宗和原始順序等來源原則，和多元層次檔案描述標準，統整臺灣總督府專賣局檔案和臺灣省議會議政檔案等大宗官方檔案，和近年來陸續完成的陳澄波畫作與文書、顏水龍畫作與文書、楊肇嘉文書、臺灣銀行所藏日治時期文書，以及保密局二二八與白色恐怖相關史料等多宗私人文書與集藏。

2010年再完成「臺灣研究古籍資料庫」改版上線，整合舊系統，以書刊、期刊文獻、插圖資訊等使用介面，重新整理原有臺灣總督圖書館、南方資料館等三大館藏，陳述各圖書館成立宗旨與沿革，和展現各館獨特分類標準及藏書特色。

臺史所以二大數位典藏系統正式上網開放為基礎，陸續舉辦多項檔案展覽，以推廣數位典藏成果，特別是2013年的「時空旅行：臺史所20週年檔案特展」與2014年「光影旅行者：陳澄波百二互動展」，皆移展至北中南各地的圖書館、博物館及新興的文創園區展出，吸引眾多參觀人潮。同時策畫建置「從北緯23.5°出發 陳澄波」主題網站，臺灣Google在2015年2月2日，推出陳澄波120歲冥誕的紀念畫作，這應該是臺灣籍畫家作品登上Google首頁的



▲「光影旅行者：陳澄波百二互動展」。(攝影／綠桃子)

第一次。(謝國興，2015)並和許多文創團隊合作，包括《百年芳華：臺灣女性百年風貌》(ccc創作集第6集)、《臺北歷史地圖散步》等，使臺灣學的數位典藏成果普及至大眾日常生活。

### 數位人文系統與深化臺灣學研究

完成數位典藏系統建置與正式上線後，為深化人文研究應用，隨即在加值獨特性史料方面，陸續選擇私人日記和職員錄等一手史料，開發「臺灣日記知識庫」和「臺灣總督府職員錄系統」，呈現數位人文學的系統資料特色。這兩大數位人文系統的建置與開放應用，概述如下：

#### 臺灣日記知識庫

臺史所檔案館典藏之個人與家族文書中，日記是最珍貴與難得的一手紀錄。史學界相當重視日記史料，在於日記是當時人記當時事，可真實反映史實；長年記載周遭發生之事，可觀察社會變遷的軌跡；是個人私密紀錄，對事或人的記載較為直接無隱。(許雪姬，2000)因此，由臺史所研究人員組織日記解讀班，投入日記判讀、註解與出版工作。

「臺灣日記知識庫」之研發與建置，是奠基在私人日記典藏與解讀出版的雙重

優勢，在評估史學研究者的資訊需求和當時web 2.0社會軟體工具的進展後，擬定日記知識庫的發展策略。首先，確認知識庫收錄的原始資料為日記紀錄，包括日記內文(原文或譯文)、註解資訊和手稿影像；接著，選擇wiki作為知識庫的系統平臺，強調允許使用者社群協作之共建共享平臺；第三，進行專業資訊加值程序，深入分析日記手稿紀錄與解讀成果。每一天紀錄為最基本單元，每一頁面包括標題、內文、註解和手稿影像，以及解析日記逐日紀錄的時序特性，參考檔案尊重來源原則與階層編排方式，同一部日記為一整體，以年、月、日依序編排，並以西元日曆一致性的分類標準，既能整合多元來源日記，又能維護各部日記獨特的脈絡。

「臺灣日記知識庫」於2009年正式上線，目前已收錄及開放的有林獻堂(灌園先生)、張麗俊(水竹居主人)、簡吉、黃旺成、楊基振、田健治郎、呂赫若、吳新榮、三好德三郎、和楊水心等十部日記，約計3萬2千多篇(天)、1千2百萬多字(參見表1)。目前收錄的日記記主之間都有些往來社會關係或共同生命經驗，如林獻堂夫人的《楊水心日記》是目前收錄的第一部女性日記，不僅可補足林獻堂缺漏的1928年日記，亦可相互參照與驗證。



▲臺灣日記知識庫，2009年上線。



表1. 臺灣日記知識庫收錄日記和上網使用、碩博士論文引用統計

日記名稱	日記年代	篇／天數	註解(則)	字數(萬)	自開放以來累計瀏覽頁/天數	臺灣碩博士論文引用(截至9月底)
灌園先生日記	1927-1955	9,523	17,234	519	198,317	184
水竹居主人日記	1906-1937	10,222	3,418	177	67,235	150
簡吉獄中日記	1929-1930	139	84	9	9,889	24
黃旺成先生日記	1912-1923	3,655	5,750	127	48,037	34
楊基振日記	1944-1950	1,817	5	36	17,846	10
臺灣總督田健治郎日記	1919-1923	1,461	2,691	61	22,889	64
呂赫若日記	1942-1944	689	242	12	16,937	63
吳新榮日記	1933-1967	3,988	6,971	209	71,292	112
三好德三郎回憶錄	1888-1938	314	136	34	5,312	8
楊水心女士日記	1928-1942	1,046	1,038	26	1,873	2
總計		32,854	37,569	1,210	459,627	650

而兩部日本人日記，一位是臺灣第一位文官總督田健治郎，一位是人稱民間總督茶商三好德三郎，也是研究日人來臺官僚、政商等社群互動之比較研究的重要史料。

「臺灣日記知識庫」自2015年3月起全面開放國內外各界人士申請帳號即可連線使用，使系統使用自去年起大幅成長；至2015年11月止，連線人數已超過1萬1千人，瀏覽頁面已突破20萬頁次。為進一步探究日記使用情形，利用國家圖書館的「臺灣碩博士論文知識加值系統」查詢各部日記被引用情形，並將引用書目整理於日記知識庫，提供使用者研究參考。目前被引用最多的是《灌園先生日記》，有184篇，分析碩博士論文的學科，以文史相關學科最多，但也有社會、建築、法律、教育、藝術、音樂、中醫、體育等科系，足

見日記史料的應用面向。

在「臺灣日記知識庫」永續發展方面，除了因應資訊科技進展，即時更新資訊工具功能與改善服務介面外，系統收錄資料持續累積才能擴大與深化資料庫的價值，目前開放的10部日記，書寫年代自1888年起至1967年止，跨越清領、日治到戰後三個統治政權，主要集中於日治時期。為追溯清領及更早期的歷史和延續戰後相關紀錄，臺史所檔案館正在處理馬偕牧師日記、熱蘭遮城日誌，以及黃金運臺關鍵人物吳嵩慶、駐韓大使邵毓麟等人的日記資料庫建置工作。有大量的數據後，未來可透過資料

探勘等數位技術，挖掘日記紀錄間的關聯性，特別是社群網絡關係，如以《灌園先生日記》為例，運用研究人員所註解8千位人物，初步可擷取出日記紀錄出現有14萬餘筆人名資訊，就可藉此資料集，分析林獻堂在各領域的社群互動關係。（王麗蕉，2015）

### 臺灣總督府職員錄系統

臺灣總督府在統治臺灣期間，比照日本內閣官報局彙編發行官廳職員名錄的制度，每年印行《臺灣總督府職員錄》，收錄任職於公務體系的官職員資料，上至總督、民政長官等高級官員，下至地方單位囑託、雇員等基層人員。歷年職員錄傳承至今，不僅是研究日本殖民統治臺灣官僚體系運作的第一手史料，亦是解讀日治時期檔案不可或缺的參考工具書。



▲臺灣總督府職員錄系統，2011年上線。

臺史所檔案館蒐集了1896年至1944年，51本職員錄，其中有兩本因地方改制而新增臨時版，有七個年度未能尋得而以日本內閣職員錄替補之，著手規畫「臺灣總督府職員錄系統」的建置，根據各年度職員錄編輯體例，確認系統主要收錄的資料內容，總計有95萬筆名錄資料，訂定編碼與結構標準；另職員錄全書編輯體例和官制官規等組織章程，則以電子書影像呈現。

「臺灣總督府職員錄系統」於2011年正式上線，提供名錄查詢、組織瀏覽、電子書等三大功能。隨系統全面開放（毋需申請帳號），連線數量逐年成長，不僅成為學術界研究過程中重要的參考工具，亦是社會大眾家族尋根的資訊查詢來源，其中包括日籍「灣生」追尋臺灣記憶的故事；並在2014年11月臺北市長選舉爆發的爭議事件當月，系統連線多達六千餘人次、瀏覽頁面有五萬多頁之多。顯見此數位人文系統的建置與開放，不僅承續原有職員錄真實記載常民生命經歷的一手史料價值，透過系統全面開放各界連線使用，更使珍貴的歷史紀錄從學術殿堂走入更多元的日常生活應用。

「臺灣總督府職員錄系統」收錄資料

幾乎完整涵蓋了日治時期臺灣所有公職名錄資料，具備獨特性、連續性與完整性等特色，尤其是在名錄資料中刊載本籍（出生地），是研究日治時期人流與跨境等議題的重要史料來源。如日治時期初等學校師資人力分布與日臺籍人士比例及其流動情形（王麗蕉、謝明如2016），以及臺灣女性在日治時期的公職參與人力分析（王麗蕉，2016）等，均為例證。

### 結語與展望

因應數位時代資訊科技進展，臺灣在新世紀伊始，全面推動數位典藏國家型科技計畫，由人文領域主導，結合圖書館、檔案館與資訊技術專業，數位系統的發展，已由線上目錄查詢系統、數位典藏系統、整合查詢平臺與知識庫，到運用資訊探勘與視覺化工具等，深入分析與呈現大數據中的關聯性資料。不論是數位典藏系統，或是數位人文系統，皆從目錄查詢與資源取用等資訊服務，朝向深化人文研究的發展領域。臺史所目前正在建置臺灣民間信仰的資料庫，包含文獻影音圖片等資料，以資料庫的豐富詳盡資料為基礎，結合地理資訊系統的應用，將對臺灣傳統社會文化的理解呈現新的觀點與風貌。☒



▲清咸豐西港刈香路關。