

中央研究院近代史研究所近代史數位資料庫 (MHDB) 介紹

張哲嘉 中央研究院近代史研究所副研究員

在資訊科技持續發展的激勵下，如何開發數位資源與研究工具，成為近年來各研究機構致力發展的重點工作。中央研究院近代史研究所自 1997 年起，即開始投入史料的數位化，以本院推動漢籍電子文獻資料庫為契機，逐漸發展成「近代史全文資料庫」。此外，近史所的檔案館多年來持續進行外交及經濟檔案之數位化，建置了「館藏檢索系統」及「人名權威檢索系統」。胡適紀念館製作有「胡適檔案資料庫」，郭廷以圖書館則開發「報刊資料檢索系統」，此外尚有「婦女雜誌資料庫」、「清代糧價資料庫」等小型資料庫。近兩年來本所資料庫小組更推展各類主題式資料庫，如「英華字典資料庫」、「近代春秋 TIS 系統」、「近現代人物資訊整合系統」等，以期將前述各個資料庫鏈接起來，建立不管是從人、時、事、地、物、還是以概念為出發點，均可追蹤史料間關聯性的歷史數位網路。

2014 年我們整合前述資料庫，組成近代史數位資料庫群組 (MHDB, Modern History Databases)，並建置 MHDB 入口網，提供使用者明確的申請說明及登入口，未來本所陸續增置的各類數位資源皆將於此平台展示。以下就分別簡介各個主要資料庫：

近代史全文資料庫

「近代史全文資料庫」為中文全文



MHDB 入口網 mhdb.mh.sinica.edu.tw

(Full-text) 檢索資料庫，免費開放提供世界各地學者，申請帳號即可使用。目前共收錄 13 套資料庫，約 7,500 萬餘字，資料類型涵蓋工具書、文集、日記手稿、檔案史料，為研究中國近代史的重要利器。

此外，值得一提的：部分史料提供原稿與文字參照，讓使用者得一窺近代名人筆墨，也是學習辨識行草的珍稀教材。

檔案館館藏檢索系統及人名權威檢索系統

館藏檢索系統內容包含外交檔案、經濟部門檔案、館藏地圖三大類；外交檔案包括清季、北洋政府以及國民政府以後之外交檔案。涵蓋時間自 1861 年至 1980 年代，為研究近、現代外交最完整之史料。經濟部門方面，本館所藏的經濟檔機構全宗種類繁多，來源除中央主管單位，亦有如經濟安定委員

會、美援運用委員會、國際經濟合作發展委員會等部門。涵蓋時間自 1903 年至 1980 年代。館藏的另一特色是數量可觀的各式地圖，包括清末至民國水利行政部門所存的河道、水文圖，及軍事機構所繪的地形圖與要塞圖。這 1 萬 5 千餘幅的地圖，且有部分手繪原稿，極其珍貴。

此外，檔案館另建有館藏之人名權威檢索系統，內容包含外交與海關人物。目前外交人物資料約有 1,800 人，海關人物約 3,000 人。

胡適紀念館胡適檔案檢索系統

胡適是近代中國最具影響力的人物之一，本館所藏的文件檔案多是 1949 年後胡適在美國，以及 1958 年回臺灣以後的文稿信函。「胡適檔案 Metadata 資料庫」自 2000 年起建置。2009 年中國社會科學院近代史研究所捐贈本館一批 1949 年之前的胡適文件檔案影像資料，整編後共建檔 30,789 筆目錄，命名為「北京檔」系列，於 2011 年納入「胡適檔案檢索系統」，合計筆數達 51,965 筆。

郭廷以圖書館報刊資料檢索系統

本館報刊資料檢索暨圖書館服務系統內容主要包含本所圖書館已數位化之近代報刊影像檔與 metadata 資料，資料類型有報

紙、期刊、公報等。此外，本系統也介紹原有圖書館網站，整合提供為一站式報刊資料檢索暨圖書館服務系統。資料庫開放線上書目檢索，欲瀏覽全文影像或列印等服務的讀者，可依使用要點申請利用。

概念主題資料庫：英華字典資料庫

「英華字典資料庫」由數位資料庫小組製作，目前已收錄 1822 年馬禮遜 (Robert Morrison)、1847 年麥都思 (Walter Henry Medhurst)、1866 年羅存德 (Wilhelm Lobscheid)、1884 年井上哲次郎、1887 年鄭其照、1908 年顏惠慶、1916 年赫美玲 (Karl Ernst Georg Hemeling) 等人所編纂的早期英華字典，包含近 7 萬筆英文詞目、31 萬筆中文詞目，是研究近代中國區域方言辭彙及中西文化交流軌跡的重要工具書。

本資料庫為「英華」與「華英」雙向檢索系統，可跨辭典搜尋。除檢索外，亦提供原書瀏覽，使用者可在瀏覽介面自行選擇欲瀏覽的字典。

將原書建置成資料庫後，更可發揮僅靠紙本無法企及的強大數位功能。以赫美玲的《官話》為例，該書在詞目後標示〔俗〕、〔文〕、〔新〕、〔部定〕等類別，但並未明列這些詞彙的數量與詳目，透過檢索，我們可直接查出全書中標為〔俗〕的字彙有 609 筆，且清楚得到這些詞彙的列表。

時間主題資料庫：近代春秋 TIS 系統

「近代春秋 TIS (Timelines Information System) 系統」的概念是以時間軸串接史料，將本所資料庫中帶有時間質素的《中華民國史事日誌》、《蔣公大事長編初稿》與諸名人日記等串聯起來，使得以時考史更形便利，這是整合 MHDB 資料庫網路的大動脈。

資料庫查詢可用「關鍵字」，也可以「某時間段」或「特定某日」搜尋。搜尋結果列表後，點選「時間軸」按鈕，將開啟時間軸頁面，以便觀察歷史事件發生前後的史料記載。

本資料庫同時具備簡易的「詞頻」與「綴詞」分析工具，如以「**革命」做綴詞分析，可看出本資料庫中共有 513 個詞彙含「革命」一詞，並統計詞彙出現的頻率，點選「匯出」可另存詞彙列表。

得到檢索結果後，可在列表頁面再點擊按鈕，連結至其他資料庫。這個功能一方面讓保有各資料庫的獨立地位，同時也使得本所各個資料庫共同形成一個如常山之蛇般首尾靈動的史料網路。

人物主題資料庫：近現代人物資訊整合系統

建立「近現代人物資訊整合系統」的初

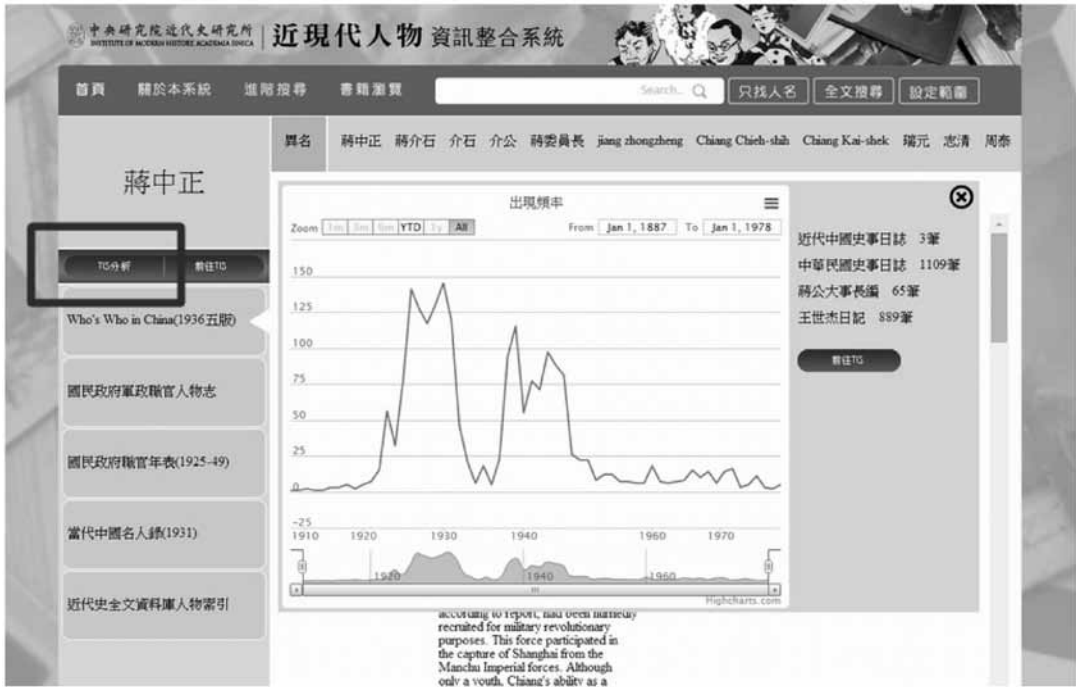
衷，是在「時」之外開通另一條以「人」貫串近代史料的幹線。目前收錄近代史全文資料庫人物索引、上海地區工商界人物錄、中國人物傳記資料、職官表、楊建成「日治時期臺灣人士紳資料庫(1915-42)」等多項傳記資料，時間自清中葉迄至現代，現有資料 4 萬 7 千多人。

本資料庫特別整理出人物的字號、別名等，列為「異名欄」，對同人異名或異人同名的現象，使用者可利用「消歧義」功能或「模糊查詢」功能搜尋。

本資料庫與前述資「近代春秋 TIS 系統」有緊密的連結，「近現代人物整合系統」內搜尋到的人物可直接連結到後者，預設成「近代春秋 TIS 系統」的檢索詞，直接進行內容檢索。反之先用時間系統查詢資料後，也可連回本系統，查詢是否有相關人物資料可資參考。

「人物整合系統」另設有統計功能，目前統計母體僅限於「近代春秋 TIS 系統」。以「蔣中正」詞目為例，點選「TIS 分析」，就會出現統計圖，顯示在「近代春秋 TIS 系統」中的每本書各出現幾次「蔣中正」，還有「蔣中正」分別在不同年份出現的次數折線圖。

本所 MHDB 資料庫群組中，尚有其他主題資料庫，如「婦女雜誌資料庫」、「清代糧價資料庫」以及 2014 年底即將完工的「清季史料全文資料庫」等，礙於篇幅所限，無法一一介紹，讀者可直接登入



MHDB 入口網一覽全豹。

結語

根據本所十幾年來的資料庫建置經驗，我們認為現有文史資料庫建置的進程中存在著極大的數位落差，從紙本史料保存到資料數位化的知識載體轉型的過程中，一方面仍然還有許多報刊、檔案、照片、影音等史料尚待數位化，這些需耗費龐大的資金與人力的基礎數位化工作仍任重道遠；另一方

面，則已有學者開始提倡思考如何以數位研究工具進行人文研究。文史資料與資訊技術的融合是跨科際的體驗，需要不同學科領域學者的共同合作。

我們希望本所的 MHDB 群組能兼顧這種數位落差中的各個面向，本所典藏的大批檔案資料，基礎 metadata 建置工作仍將持續進行，並且試圖在不同資料庫平台中取得平衡，建立各資料庫的關連，而已經累積大批資料量的全文資料庫則開始提供數位研究工具，開放學者使用。從「近代史全文資

料庫」、「英華字典資料庫」至「人物資訊整合系統」、「近代春秋 TIS 系統」等，我們強調文史資訊的整體關聯與資料重組後可能呈現的不同意義結構，也提供詞目比對、時間軸日期比對、綴詞分析、圖表統計等各種數位分析方法。同時，MHDB 資料庫的內容是對全世界開放的，我們希望世界各地的學者在 MHDB 資料庫中取得這些史料後，能自行做各種加值運用，共同促進文史研究的發展。本所也一貫保持開放的態度，期望與各地研究單位，不論在史料分享或資料庫技術上，能有各種合作模式。

新書介紹



中華民國政府遷臺初期重要史料彙編： 中美協防（一）～（三） 臺海危機（一）、（二）

2013 年 12 月～2014 年 11 月出版
精裝各 530～600 元

政府遷臺初期是一個重要的關鍵年代，中華民國從此進入不同軌道的國家發展進程。在這段期間，臺灣遭遇中共的武力威脅，中華民國政府透過外交獲得美國軍援，力圖生聚教訓，轉危為安。國史館蒐集相關史料，已出版 5 本重要史料彙編，藉由史料呈現當年史實。

請洽國史館訂購

地址：臺北市中正區長沙街一段 2 號

電話：02-23161067

網址：<http://www.drnh.gov.tw>