

國史館數位檔案 檢索系統介紹

黃汎如 國史館審編處辦事員



圖 1 數位檔案檢索系統首頁

壹、前言——合作起源

國史館自民國 91 年（2002）加入國科會「數位典藏國家型科技計畫」以來，即積極進行史料數位化的工作。為了使史料應用更趨普及、有效率，除了參與數位計畫之標的外，本館也加速的將其他史料進行數位化，並透過資料庫的改善及功能建置，滿足使用者的需求，便利各界運用。98 年 4 月，恰逢國立臺灣大學數位典藏研究發展中心項主任潔來館拜訪，彼此意見交流中意外激起火花，為雙方合作埋下了契機。

99 年 3 月，雙方正式洽談「數位檔案檢索系統」建置案，期望在臺大成熟的 IT 技術及本館豐碩的史料檔案結合之下，建置強調系統易用性，並以後分類查詢為特色的新使用系統。5 月 4 日，臺大李校長嗣洵與本館林前館長滿紅共同簽署合作協議書，之後並舉辦多場系統說明會、工作坊，向外界推廣系統雛型，並針對與會者之回饋不斷進行系統修正，期望能更貼近使用者需求。

99年11月11日，系統建置完竣，於本館臺北辦公室舉辦「國史館數位檔案檢索系統發表會」，並邀請學者專家介紹專藏史料，獲得廣大迴響。(註1)100年1月27日，正式以本館網址(<http://ahdas.drnh.gov.tw>)上線供各界使用(圖1)。

貳、檔案內容

本系統由臺大數典中心與本館合作開發建置，收錄本館已數位化之總統副總統文物及一般史料檔案，內容如下(依全宗號排序)(註2)：

一、總統副總統文物

蔣中正總統文物、蔣經國總統文物、嚴家淦總統文物、李登輝總統文物、陳誠副總統文物、謝東閔副總統文物。

二、一般史料檔案

國民政府檔案、資源委員會檔案、臺灣省政府地政處檔案、外交部檔案、閻錫山史料、汪兆銘檔案。

類型包含文件、照片、底片、視聽等，共計12個全宗檔案、54餘萬筆目錄資料、590餘萬個數位檔案。

參、系統特色

如何幫助使用者在豐富的檔案中快速獲

得想要的資訊，一直是資料庫建置者努力的方向，尤其本館特殊的檔案性質及豐碩可觀的檔案數量，更有賴資料庫功能的精進來提供史料查詢應用的效率與便利。

本系統除了提供跨全宗的資料整合查詢功能，最大的特點在於加入「後分類」檢索技術。有別於傳統資料庫提供關鍵字查詢僅能獲得單一面向的檔案資訊，例如：符合關鍵字的檔案有哪些？或分布於哪些全宗檔案？「後分類」檢索技術進一步將檢索結果統計分析，提供檔案相關年代分布、相關人員、地名等多維度的檔案面貌，幫助使用者更迅速掌握檔案概況，對於提昇研究效率，甚至啟發研究問題等皆有相當大的助益。有關「後分類」詳細功能將於下節介紹。

肆、系統功能

本系統以使用者思維出發，建置許多功能，以下分別介紹之。

一、檢索方式

本系統提供兩種檢索方式：「瀏覽查詢」與「關鍵字查詢」，介面如圖2所示。

(一) 瀏覽查詢：想要了解史料的整體面貌，可以透過全宗瀏覽來瀏覽資料。瀏覽查詢完整呈現各全宗的檔案層級架構，只要點選介面中的「+」符號，便可展開全宗、系列、副系列、副副系列等檔案階層。

(二) 關鍵字查詢：可利用關鍵字、卷

圖 2 檢索介面

名、檔案系列、內容描述、日期、地名、人員、照片日期、拍攝地點、拍攝者等欄位進行單一條件查詢，或利用「AND」、「OR」、「NOT」等檢索條件組合來進行進階檢索。

二、檢索結果後分類

所謂「後分類」係指針對瀏覽查詢或關

鍵字查詢的結果再進行分析整理，以呈現檔案多維度資訊的一項技術，亦是本系統最大的特色。

本系統提供五種後分類條件，包含：檔案階層、檔案年代、相關人員、相關地區、相關主題等（註3），使用者可依據需求進行條件組合，最多三層，以快速掌握檢索結果

圖 3 檢索結果後分類介面

的概貌與分布情形。例如：以「中日和約」為關鍵字檢索，並進一步設定後分類條件第一層為「檔案年代」，第二層為「相關人員」，如圖 3。

首先觀察「檔案年代」，發現與中日和約相關的檔案時間以民國 41 年出現次數最多，恰好與中日和約簽訂的時間相吻合；

進一步展開該年度的「相關人員」，同樣由次數分布可以觀察人物在此事件中的重要性，排名前三位分別為蔣中正、河田烈及葉公超，皆為簽訂中日和約的要角。透過「排除選擇」或「縮小範圍」，可以快速進行條件篩選，節省使用者在眾多檔案中大海撈針的時間。例如想要查找簽訂中日和約的相關

檔案，可以設定：縮小範圍至檔案年代「41年」、相關人物包含「河田烈」或「葉公超」。

三、人名權威查詢

除了上述應用外，本系統已完成和「國史館人名權威查詢系統」（註4）的連結，只要在後分類條件「相關人員」人名之後有名片圖案之標示，即代表可以連結至該系統參閱相關人物資料。此一設計，除了達到資源共享的目的，也方便使用者只要利用一個系統即能整合所有查詢功能。

四、檢索結果檢視

輸入檢索條件後，畫面中央會顯示檢索

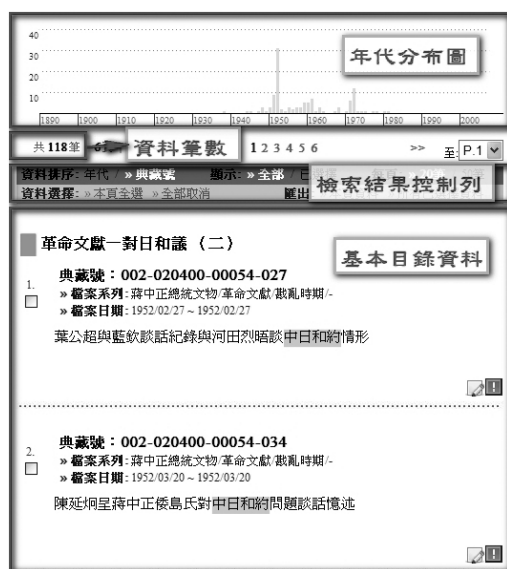


圖 4 檢索結果畫面

結果，包含檔案年代分布圖、資料筆數、檢索結果控制列，以及基本的目錄資料（參圖 4）。

（一）年代分布圖：（註5）將檢索結果的檔案年代次數分配以圖像化呈現，讓使用者可以藉由長條圖的觀察，迅速掌握檔案年代分布情形，並可藉由檔案年代分配趨勢啟發新思維。

（二）資料筆數：指符合檢索條件的目錄資料數量。

（三）檢索結果控制列：提供使用者根據個人需求調整檢索結果顯示方式的自由性，包含：

1. 資料排序方式：依年代或典藏號排序。
2. 資料顯示：顯示全部符合檢索條件的資料，或是只顯示使用者勾選的資料。
3. 每頁顯示數量：提供兩種選擇，一頁顯示 20 筆或 50 筆。
4. 目錄資料選擇：使用者可以針對檢索結果中，符合個人需要的目錄資料進行勾選（參圖 5），並搭配「資料顯示」或「匯出」

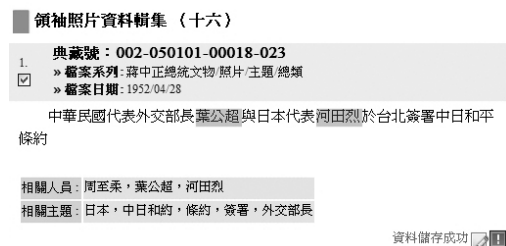


圖 5 目錄資料勾選圖示



圖 6 目錄資料匯出畫面

目錄資料」功能使用。除了一筆一筆勾選欲選擇或欲取消選擇的目錄資料外，也可以透過「本頁全選」或「全部取消」以節省時間。

5. 匯出目錄資料：除了在线上瀏覽目錄資料，亦可選擇將目錄資料匯出儲存。系統提供兩種匯出方式，除了匯出符合檢索條件的目錄資料（全部匯出），也可以選擇只匯出使用者勾選的目錄資料（所有已選擇資料）。

目錄資料匯出畫面如圖 6 所示，上方為控制列，可以選擇下載儲存或直接列印。

「下載儲存」為 CSV 格式，可以 Excel 程

式開啟，為表單形式；（註 6）「列印」則是直接列印此匯出畫面。畫面中間會顯示匯出此份目錄資料的相關資訊，包含：使用人員、輸出時間、使用條件、輸出數量、典藏地點及典藏時間等，提供使用者參考。

（四）基本目錄資料：為幫助使用者對檢索結果可以一目了然，檢索結果畫面僅會顯示基本目錄資料，包含：卷名、典藏號、檔案階層、檔案起迄日期及內容描述。人名資訊、地名資訊及相關主題預設為隱藏，僅在符合欄位檢索條件時才顯示。如果使用者想要瞭解檔案更詳細的內容，則可點選「典藏號」，進入詮釋資料與影像閱覽畫面。

五、詮釋資料與影像閱覽

詮釋資料與影像閱覽可分為三個部分（參圖 7）。畫面左上方呈現每一筆檔案的 metadata（詮釋資料），包含：入藏登錄號、卷名、典藏號、檔案階層、內容描述、檔案

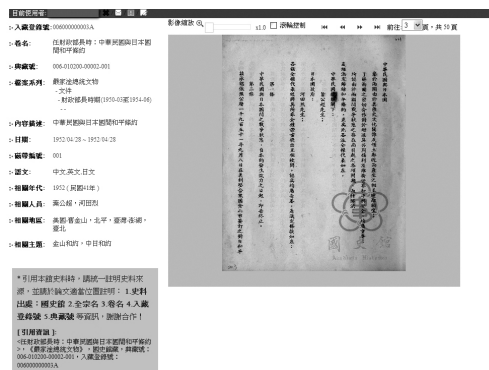


圖 7 詮釋資料與影像閱覽畫面

起迄日期、光碟 / 微捲 / 磁帶編號、檔案語文、相關年代、相關人員、相關地區、相關主題（關鍵字）等。畫面左下方為檔案引用資訊，提供使用者於研究論文中引用格式的參考。畫面右方為檔案影像圖檔，（註7）透過上方控制列，可以任意縮放影像大小。每一張影像圖檔右下方皆印本館浮水印。

六、檢索組合與儲存

在傳統的檢索系統中，檢索的歷程是一條無法回頭的直線，在檢索過程中如果使用的檢索條件無法滿足需求或是有了新的發現時，無法中途更改檢索策略，或是替換檢索條件，而必須重新開始一個檢索流程，不僅麻煩也費時。（註8）而「檢索組合」、「檢索儲存」與「檢索歷史」的功能設計，便是強調將檢索過程彈性化，讓使用者隨時可以更改檢索策略，而且一點也不麻煩！以下將分別介紹：

（一）檢索組合

系統依序記錄當次檢索輸入的所有條件，使用者可以點擊任何一個條件返回該步驟，再配合新的檢索條件繼續檢索；也可以刪除或選擇暫時隱藏遮蔽某個條件，來觀察檢索結果的變化。

舉例來說，圖8顯示這次的檢索歷程使用了三個條件，分別為：1. 關鍵字有中日和約；2. 檔案年代為41年；3. 相關人員出現河田烈或葉公超。若使用者觀察檢索結果

The screenshot displays two main sections: '檢索組合' (Search Combination) and '檢索儲存' (Search Storage). The '檢索組合' section lists three search criteria: 'x1. 關鍵字 = 中日和約', 'x2. 相關年代 = 1952民國41年', and 'x3. 相關人員 = 河田烈' and 'x3. 相關人員 = 葉公超'. Below these is a search box containing '中日和約' and a '儲存' (Save) button. The '檢索儲存' section shows a list of saved search conditions, including '1. 中日和約' with a date of '2012/02/03,16:01'. It also lists five '檢索歷史' (Search History) items with their respective keywords: '關聯字 = 中日和約', '關聯字 = 廬山會議', '關聯字 = 廬山會議', '關聯字 = 租界', and '關聯字 = 北伐'. A '更多檢索歷史' (More Search History) link is visible next to the second history item.

圖8 檢索組合與儲存之畫面

後，突然想更改條件3為：「相關人員出現蔣中正或吉田茂」，此時，不需要重新進行檢索，只要點選第二步驟「相關年代=1952民國41年」，則系統會直接回到檢索條件為：1. 關鍵字有中日和約；2. 檔案年代為

41 年，使用者只要再加入檢索條件「相關人員 = 蔣中正 or 吉田茂」即可。

若使用者僅是想取消「2. 檔案年代為 41 年」的條件，其他檢索條件不動，可以直接點選條件左方的「X」，移除該條件。若是不想移除條件，只想觀察少了某個條件，檢索結果會產生什麼變化，則可以選擇右方的「v」，暫時遮蔽條件，欲取消遮蔽時，再次點選「v」即可還原。

（二）檢索儲存與檢索歷史（註9）

使用資料庫時，有沒有一種經驗，當你查詢多筆資料後，才發現前幾次的檢索結果可能有參考價值，於是要趕快回想當時的檢索條件組合，重新進行檢索。為了解決這樣的問題，系統提供「檢索儲存」與「檢索歷史」兩種功能。使用者可以將自己認為可能重要的檢索歷程命名儲存，或者利用系統自動儲存最近 20 次的檢索歷史，快速的回到某次檢索歷程，以減少重新檢索的時間。

七、使用者筆記與錯誤回報

除了基本檢索功能外，本系統另提供「使用者筆記」及「錯誤回報」功能，以下分別介紹。

（一）使用者筆記

為方便使用者瀏覽檢索結果時，可以即時記錄所有想法，在每一筆目錄資料右下方都有一個「使用者筆記」的按鈕，使用者

可以隨心記錄所有想法，並於目錄資料匯出時，一併將使用者筆記匯出（參圖 9）。此功能的好處在於讓許多乍現的靈光得以保存，尤其蒐集大量研究資料時，可以更快的掌握每一筆目錄資料對研究的參考價值。

（二）錯誤回報

本館一直致力於提供質量俱佳的檔案供讀者參考，然而在大量的檔案編目過程中，難免有疏失，除了在審查檔案時予以修正之外，讀者的回饋也是幫助檔案品質提升的關鍵。由於本系統的使用者以專業人士居多，對於檔案皆有一定程度的了解，因此往往可以在應用檔案發現不合理的錯誤時，向館方提出反應。為了使這樣的回饋管道更加容易，系統設置「錯誤回報」功能（參圖 10），一方面讓使用者可以在發現錯誤的當下直接利用系統進行回報，一方面也可以讓系統管理者第一時間掌握訊息，即時查證，

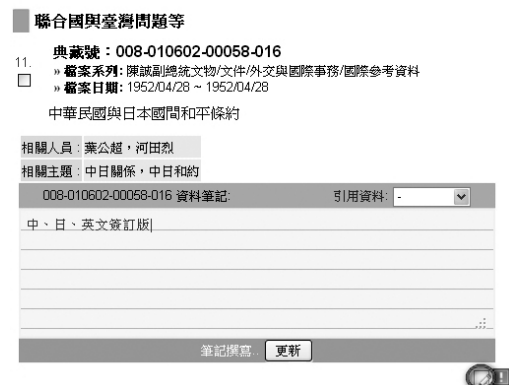


圖 9 使用者筆記畫面

中共外交情報

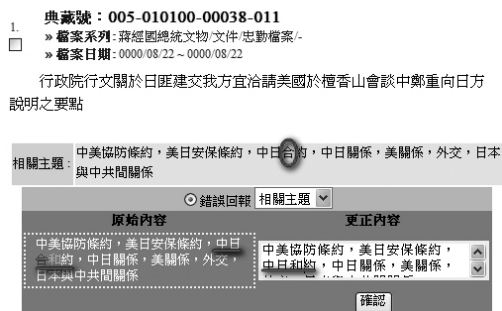


圖 10 錯誤回報畫面

並快速修正，俾利未來使用者。

八、訊息公告

有關系統的相關消息公告，置於畫面的右下角供使用者參考（參圖 11）。除了系統更新等訊息，讀者提供的錯誤回報，經審核確認後，系統管理者會進行修正，並公告修正檔案之典藏號，俾利讀者參考使用。

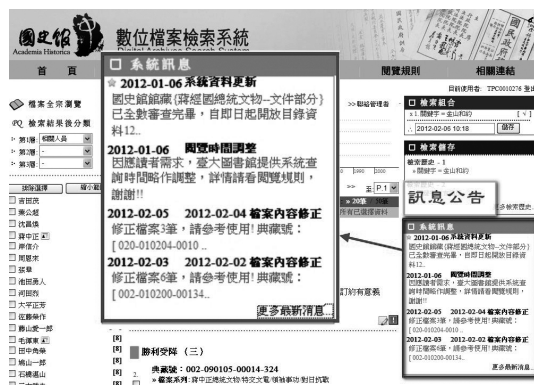


圖 11 訊息公告介面

伍、結語

本館保存國家許多重要文化資產，記錄國家歷史發展重要歷程，一直以來為各界研究的重要參考。為了能夠讓此批珍貴史料廣為大眾應用，並保留檔案完整狀態，本館一直致力於檔案數位化的工作，並不斷精進資料庫的好用度。數位檔案檢索系統的誕生，承中研院計算中心為我們數位化工作打下的基礎，及臺大數典中心團隊於本合作案中提供豐富的經驗與技術分享，才能運作得如此順利。但儘管我們接收各界的回饋，不斷更新系統功能，希望能幫助本系統更臻完善，但仍有改進空間，為避免使用者研究時解讀資料的誤解，特別說明如下：

一、Metadata 的限制

本系統係一個跨全宗整合的查詢系統，目前已收錄 12 個全宗資料供使用者查詢利用。在資料庫追求越來越好用的同時，我們也一直期望對研究者而言，最重要的檔案內容也能更加的完整，而這部分有賴 metadata 的充實，因此一直不斷檢討修正 metadata 編目建檔著錄規範。早期建置之檔案，部分欄位並不完整，例如蔣檔籌筆系列，並無人名權威之連結，也無相關地區的註記，因此若透過後分類篩選人名、地名資料時，此部分檔案可能產生遺漏。

而在資料建檔的格式，除了因應檔案性質而有差異之外，（註 10）同樣也受到不同

時期著錄規範的影響，而有些微不同。因此可能發現同樣地名，有些加著省級，有些只寫城市。

基於上述原因造成使用者利用後分類檢索結果可能產生缺漏的情形，我們虛心接受，並努力改進。然而由於檔案數量相當龐大，審查作業需要耗費不少時間與人力，目前檔案內容都持續審查修正中，但仍有待時間完成，尚請讀者見諒。

另外，在檔案年代的註記上，檔案的起迄時間為一個區間範圍，將時間區間於後分類上呈現有實行上的困難，因此本系統以檔案起始日期作為後分類「檔案年代」的時間計算，而在目錄資料中則呈現檔案起迄日期供使用者參考。因此當使用者發現某個事件發生的時間點與後分類顯示的檔案年代不符合時，可以觀察檔案起迄日期再作解讀。

二、檔案數量

觀察檔案數量多寡是許多使用者利用本系統發現研究問題的一個方法，然而檔案數量除了某種程度反映歷史事實之外，也受到史料徵集情形及編件規範的影響。例如：蔣中正總統照片，一張編為一件；嚴家淦總統照片則以一個事件編為一件，因此蔣檔的照片數量，是其他全宗的數倍之多。故使用者在應用檔案數量進行資料解讀或推論時，應特別留意數字的意義。

本系統建置的目的，在於提供使用者更有效率的檢索經驗，並期待以多維度的資料

呈現，刺激使用者不同的發想。本系統自建置階段至正式上線以來，由於眾多使用者的支持，並持續給予我們指教，才能使系統有今天的模樣。提供讀者更好用的資料庫一直是本館努力的目標，也期盼各界繼續給予任何使用經驗上的回饋，我們當竭盡所能，讓系統越來越好。

【註釋】

1. 有關系統發表會相關內容，可參考：黃汎如，〈「國史館數位檔案檢索系統發表會」紀實〉，《國史館館訊》，第6期（2011年6月），頁125-138。
2. 「嚴家淦總統文物」與「閻錫山史料」為100年新增之全宗檔案，「閻錫山史料」目前只提供微捲閱覽，因與蔣檔相關性高，頗有參考價值，因應讀者需求，亦納入本系統中供參考使用，惟欲閱覽該全宗案卷內容，請至國史館申請微捲閱覽。
3. 「相關主題」因編目建檔規範之不同分為兩種類型，(1)主題詞：「國民政府檔案」、「蔣中正總統文物」、「資源委員會檔案」等3個全宗檔案，依照檔案主題建置「主題詞表」，例如：人事－任免、財政－預算，有興趣者可參考「國史館數位典藏計畫網站」(<http://dft.drn.gov.tw/resources-1-1.htm>)－主題詞表；(2)關鍵詞：上開全宗以外之9個全宗檔案，「相關主題」係指檔案中出現的關鍵詞，特此說明。
4. 國史館於民國91年參加人名權威合作計畫，由中央研究院計算中心、數位典藏國家型科技計畫後設資料工作組提供技術援助，開發完成「國史館人名權威查詢系統」，網址為：<http://dft.drn.gov.tw/textdb/people/>。
5. 檔案日期有起迄之分，在此處係以「檔案起始日期」做為統計年代分布的依據。
6. 系統提供 UTF-8 及 Big5 兩種編碼供下載，使

用戶可以依據個人資訊軟體版本選擇，建議以 UTF-8 編碼為主，若開啟時無法顯示正確編碼時，再改以 Big5 編碼下載，惟 Big5 不支援部分難字，特此說明。

7. 本系統所有目錄資料皆開放於網路供各界查詢，欲閱覽影像圖檔，請至本館、臺大圖書館特藏資料閱覽區、及臺大數典中心申請。惟因檔案特殊性，蔣中正總統照片、特交檔案、汪兆銘史料等影像圖檔，僅於本館提供閱覽。
8. 參考蕭屹靈，〈日治法院檔案系統及其後分類呈現〉，（臺北：國立臺灣大學資訊工程學研究所碩士論文，2008年6月），頁3。
9. 系統以登入帳號為基礎儲存檢索紀錄，使用者每次登入系統給予一次性帳號，所有儲存紀錄隨登出後結束，再次登入，所有檢索紀錄將重新開始。
10. 以嚴家淦總統文物為例，照片檔案的「相關地名」，著錄照片拍攝地點，著錄至鄉鎮層級，且地點明確，如：三軍軍官俱樂部（臺北）；而文件檔案著錄檔案內容中重要地名，著錄至縣市層級，如：臺中（臺灣）。相關著錄規範可參考：〈國史館編目建檔著錄規範〉，收錄於「國史館數位典藏計畫網站」：<http://dft.dnh.gov.tw/resources-1-1.htm>（2012/2/6 點閱）。

