

第13次中文文獻資源共建共享合作會議

13th Conference on Cooperative Development and
Sharing of Chinese Resources

跨域經緯・創新翻轉

—圖書館與數位人文的對話

會議論文集

Conference Proceedings of Dialogue between
Libraries and Digital Humanities

會議時間 | 2021年7月13日至7月14日

會議地點 | 視訊會議

主辦單位 | 中文文獻資源共建共享合作會議理事會(Council of the CCDSCR)

漢學研究中心(Center for Chinese Studies)

第13次中文文獻資源共建共享合作會議
13th Conference on Cooperative Development and
Sharing of Chinese Resources

跨域經緯・創新翻轉

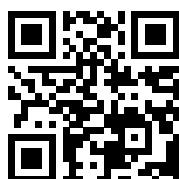
—圖書館與數位人文的對話

會議論文集

Conference Proceedings of Dialogue between
Libraries and Digital Humanities

第 13 次中文文獻資源共建共享合作會議

<https://pse.is/3e37pp>



目次 Table of Contents

- 3 序 Preface

- 5 會議簡介 Introduction

- 6 會議議程 Agenda

專題演講

Keynote Speeches

- 11 達成互惠互利：數位人文與大英圖書館

Working Towards Mutual Benefit: Digital Humanities and the British Library

米亞·里居博士 Dr. Mia Ridge

- 13 圖書館與數位人文合作之解密：數位人文與集中化問題

The Library's Secret Sharer: Digital Humanities and the Question of Centralization

亞列克斯·吉爾博士 Dr. Alex Gil

【子題一】圖書館發展數位人文學術服務

Session I: Developing Digital Humanities Academic Services in Libraries

- 20 交融式数字图书馆场景中的数字学术服务——以北京大学图书馆为例

Digital Scholarship Services in the Scene of Unified Digital Library -- Case of Peking University Library

陈建龙 Jianlong Chen/ 别立谦 Liqian Bie/ 童云海 Yunhai Tong

- 38 以旅券檔案數位加值資訊追尋日治時期臺灣人的跨境足跡

Tracing Overseas Footprints of Taiwanese with Digital Information from Passport Archives during Japanese Colonial Era

王麗蕉 Li-chiao Wang

【子題二】圖書館數位人文研究平台建置與應用

Session II: Creating and Using Digital Humanities Research Platforms in Libraries

- 60 汉学数字人文通用平台的需求与设计

The Functional Requirement and Designing of a Universal Digital Humanities Platform for Sinology

刘炜 Wei Liu / 王丽华 Lihua Wang

84 **歷代寶案脈絡分析系統之建置**

A Context Discovery System of LidaiBao'an

張毓哲 Yu-Che Chang/ 傅莉雯 Li-Wen Fu/ 項潔 Jieh Hsiang

122 **馬來西亞人物誌數位人文平台之建置與應用**

Development and Application of Digital Humanities Research Platform for Biographies of Malaysia Personalities

陳志銘 Chih-Ming Chen/ 張鐘 Chung Chang/ 徐志帆 Chih-Fan Hsu/ 林嘉培 Chia pei Lim

【子題三】圖書館之數位人文創新服務

Session III: Innovative Services for Digital Humanities in Libraries

150 **古文献數位化中的知識體系建設—兼論數位人文之應用**

The Construction of Knowledge System in the Digitization of Ancient Documentation – With a Study on the Application of Digital Humanities

肖琨 Long Xiao

154 **地方學數位共筆平台之資料度用與數位策展**

Data and Digital Curation for the Digital Collaboration Platform of Local Historical Studies

林信成 Sinn-Cheng Lin/ 范凱婷 Kai-Ting Fan

178 **開放資料服務於漢籍數位人文研究資源之發展與應用**

The Development and Application of Open Data Services in Classical Chinese Digital Humanities Resources

洪振洲 Jen-Jou Hung

【子題四】學術前瞻議題

Session IV: Academic Foresight Issues

202 **數位人文環境下 BIBFRAME 書目框架在古漢籍善本鏈結資料的再用與擴展**

BIBFRAME Reuse and Extension for Linked Data of Chinese Rare Books in the Context of Digital Humanities

陳淑君 Shu-Jiun Chen

序

漢學研究中心 (Center for Chinese Studies) 與全球東亞圖書館及漢學研究機構交流密切，今 (2021) 年很榮幸擔任第 13 次「中文文獻資源共建共享合作會議」(Conference on Cooperative Development and Sharing of Chinese Resources，以下簡稱 CCDSCR) 之主辦單位。回顧上一次本中心於 2010 年 11 月 3 至 4 日舉辦之第八次「中文文獻資源共建共享合作會議」，已事隔十多年，當時會議主題為「圖書館中文資源與數位典藏學術研討會」，介紹與討論各地區由圖書館或利用圖書館資源所建置的數位系統、相關技術、數位學習、應用推廣及產業加值等發展。

CCDSCR 係以兩岸四地的合作交流作為背景逐步醞釀而來，是一個專門針對圖書館中文資源的藏品發展及合作分享所規劃形成的團體，最初它以兩岸四地的圖書館合作型態來發展，至目前它已逐漸朝向國際化的團體邁進。

CCDSCR 源起於 2000 年，重要目的為聯繫世界各地典藏有中文文獻資源的圖書館或資訊機構，以促進中文文獻資源的共建與共用。歷年來已經分別在兩岸四地的多個城市，包括臺北、北京、澳門、南京、香港、敦煌、蘭州、山西等地，共舉行了 12 屆的大會。每次舉辦會議，除了臺灣、中國大陸香港、澳門、香港外，亦有來自全球各地的專家學者與會。而每一屆會議舉辦時，都設定特別的主題，專家學者就機構與教學研究上的研究與實務上的進展，進行演講分享之外，也得以利用機會討論可能的新合作方案，促進各館之間的分享與交流。

20 多年來，CCDSCR 主要以專案項目來帶動合作，經歷多年的研討與運作，已進行多項實質合作項目，至今已促成多達 14 個合作項目，包括：1. 中文石刻拓片資源庫、2. 中文名稱規範數據庫、3. 中國古代版印圖錄、4. 古籍聯合目錄資料庫、5. 中國科技史數字圖書館、6. 中國家譜總目、7. 圖書信息術語規範數據庫、8. 孫中山數字圖書館、9. 中文元數據標準研究及其示範數據庫、10. 中國近代文獻圖像數據庫、11. 西北地方文獻資源數據庫、12. 中華尋根網、13. 中文學術機構典藏聯合共享平台建構計劃、14. 華人音樂文獻集藏計劃。

受全球 Covid-19 疫情影響，漢學研究中心今 (2021) 年 7 月 13 日至 14 日間，以全線上視訊會議方式舉辦「第 13 次中文文獻資源共建共享合作會議」，會議主題訂為「跨域經緯 · 創新翻轉—圖書館與數位人文的對話」。圖書館與數位人文就如同跨域研究，希冀於會議中各圖書館的數位人文發展成果及案例，如同跨域經緯，透過兩者間的相互關係及對話，進而探索兩者間互相激發的創新發展形式。

今年我們很榮幸邀請來自英國、美國、中國、臺灣等地之圖書資訊與數位人文界的專家學者，透過線上會議，針對四大子題面向：圖書館發展數位人文學術服務、圖書館數位人文研究平台建置與應用、圖書館之數位人文創新服務及學術前瞻議題，進行相關專題與論文的發表，共計 2 場專題及 9 篇論文，分別探討「達成互惠互利：數位人文與大英圖書館」(Working towards mutual benefit: Digital Humanities and the British Library)、「圖書館與數位人文合作之解密：數位人文與集中化問題」(The Library's Secret Sharer: Digital Humanities and the Question of Centralization)、「交融式數位圖書館場景中的數位學術服務—以北京大學圖書館為例」、「以旅券檔案數位加值資訊追尋日治時期臺灣人的跨境足跡」、「漢學數位人文通用平台的需求與設計」、「歷代寶案脈絡分析系統之建置」、「馬來西亞人物誌數位人文平台之建置與應用」、「古文獻數位化中的知識體系建設—兼論數位人文之應用」、「地方學數位共筆平台之資料度用與數位策展」、「開放資料服務於漢籍數位人文研究資源之發展與應用」、「數位人文環境下 BIBFRAME 書目框架在古漢籍善本鏈結資料的再用與擴展」。我們特別將前述專家學者的論文、專題演講影音和簡報連結等集結成冊，以饗讀者。

中文文獻資源共建共享為一個長期發展目標，需各方合作單位投入更多的心力、財力和人力，秉持分享與合作精神，累積共同的經驗，持續向前邁進，希冀對全球文明作出更具意義及創新服務的貢獻，營造一個更自由、公平、開放的數位內容使用環境，深化知識的交流與分享，於人類文明發展中，再創圖書館事業世紀新價值。

漢學研究中心

館長 曾淑賢博士



2021 年 6 月 21 日

第 13 次中文文獻資源共建共享合作會議

跨域經緯・創新翻轉—圖書館與數位人文的對話

漢學研究中心 (Center for Chinese Studies) 與全球東亞圖書館及漢學研究機構交流密切，2021 年擔任「中文文獻資源共建共享合作會議」(Conference on Cooperative Development and Sharing of Chinese Resources，以下簡稱 CCDSCR) 之主辦單位。CCDSCR 源起於 2000 年，屬兩岸四地合作會議性質，旨在聯繫世界各地典藏中文文獻資源的圖書館或資訊機構，促進中文文獻資源的共建與共用，過往已分別於兩岸四地，臺北、北京、澳門、南京、香港、敦煌、蘭州、山西等地舉行了共 12 屆的大會。歷年會議與會者包括來自臺灣、香港、澳門、中國大陸、新加坡、美國、荷蘭、英國、日本和德國等國家和地區的代表。

20 年來促成了包括古籍目錄、拓片、家譜、文獻數位化、名稱規範等多領域之 14 個合作項目。每屆會議皆設定主題，討論中文文獻資源共建共用的相關問題，與會者除報告專案進展外，也洽談新的合作專案，促進各館間的分享與交流。

受 Covid-19 疫情影響，漢學研究中心於 2021 年 7 月 13 日至 14 日間，以線上視訊會議方式舉辦「第 13 次中文文獻資源共建共享合作會議」，會議主題為【跨域經緯・創新翻轉 --- 圖書館與數位人文的對話】。圖書館與數位人文就如同跨域研究，希冀於會議中各圖書館的數位人文發展成果及案例，如同跨域經緯，透過兩者間的相互關係及對話，進而探索兩者間互相激發的創新發展形式。

本次線上會議就下列四大子題進行深度探討：

- 【子題一】 圖書館發展數位人文學術服務
- 【子題二】 圖書館數位人文研究平台建置與應用
- 【子題三】 圖書館之數位人文創新服務
- 【子題四】 學術前瞻議題

議程

2021 年 7 月 13 日 (星期二)

時間	致詞者	
09:10-09:30	Webex 線上會議開啟	
09:30-10:00	大會開幕式 臺灣代表： 漢學研究中心 曾館長淑賢致詞 大陸代表： 中國國家圖書館 陳副館長櫻致詞 貴賓致詞： 國立臺灣大學圖書館 陳館長光華致詞	
【子題一】圖書館發展數位人文學術服務		
主持人 / 國立臺灣大學圖書館 陳光華館長		
時間	講題	論文主講者
10:00-10:20	交融式數位圖書館場景中的數位學術服務－以北京大學圖書館為例	陳建龍 北京大學圖書館館館長 別立謙 北京大學圖書館副館長 童雲海 北京大學圖書館副館長
10:20-10:40	以旅券檔案數位加值資訊追尋日治時期臺灣人的跨境足跡	王麗蕉 中央研究院臺灣史研究所研究副技師兼檔案館主任
10:40-10:50	主持人總結及討論時間	

議程

2021 年 7 月 13 日 (星期二)

【子題二】圖書館數位人文研究平台建置與應用		主持人 / 北京大學圖書館 別立謙副館長
時間	講題	論文主講者
11:10-11:30	漢學數位人文通用平台的需求與設計	劉煒 上海圖書館副館長 王麗華 上海大學圖書情報檔案系
11:30-11:50	歷代寶案脈絡分析系統之建置	項潔 國立臺灣大學資訊工程學系暨研究所特聘教授及國立臺灣大學數位人文研究中心主任 張毓哲 國立臺灣大學數位人文研究中心研究人員 傅莉雯 國立臺灣大學數位人文研究中心計畫研究專員
11:50-12:10	馬來西亞人物誌數位人文平台之建置與應用	陳志銘 國立政治大學圖書資訊與檔案學研究所教授兼圖書館館長 張鐘 國立政治大學華人文化主體性研究中心資訊工程師 徐志帆 國立政治大學圖書資訊與檔案學研究所專任助理 林嘉培 國立政治大學華人文化主體性研究中心專任助理
12:10-12:20	主持人總結及討論時間	

議程

2021 年 7 月 13 日 (星期二)

【專題演講一】		主持人 / 國立政治大學資訊科學系 劉吉軒教授
時間	講題	主講者
14:00–15:00	Working Towards Mutual Benefit: Digital Humanities and the British Library	Dr. Mia Ridge British Library's Digital Curator for Western Heritage Collections
【子題三】圖書館之數位人文創新服務		主持人 / 國立臺灣師範大學圖書館 柯皓仁館長
時間	講題	論文主講者
15:00–15:20	古文獻數位化中的知識體系建設—兼論 數位人文之應用	肖瓏 北京大學圖書館研究館員 山西大學圖書館館長
15:20–15:40	地方學數位共筆平台之資料度用與數位 策展	林信成 淡江大學資訊與圖書館學系教授 范凱婷 淡江大學資訊與圖書館學系博士 研究員
15:40–16:00	開放資料服務於漢籍數位人文研究資源 之發展與應用	洪振洲 法鼓文理學院佛教學系副教授兼 圖書資訊館館長
16:00–16:10	主持人總結及討論時間	

議程

2021 年 7 月 14 日 (星期三)

【專題演講二】		主持人 / 中原大學 陳昭珍講座教授
時間	講題	主講者
9:30–10:30	The Library's Secret Sharer: Digital Humanities and the Question of Centralization	Dr. Alex Gil Digital Scholarship Librarian at Columbia University Libraries
【子題四】學術前瞻議題		主持人 / 中國科學院文獻情報中心 主任助理 趙豔女士
時間	講題	論文主講者
10:30–10:50	數位人文環境下 BIBFRAME 書目框架 在古漢籍善本鏈結資料的再用與擴展	陳淑君 中央研究院歷史語言研究所助研 研究員兼數位文化中心執行秘書
10:50–11:00	主持人總結及討論時間	
11:00–11:20	綜合座談及閉幕致詞 漢學研究中心 曾館長淑賢主持	

專題演講

Keynote Speeches

專題演講一 1st Keynote Speech

達成互惠互利：數位人文與大英圖書館

Working Towards Mutual Benefit: Digital Humanities and the British Library



米亞·里居博士
Dr. Mia Ridge

米亞·里居博士為大英圖書館西方文化資產館藏的數位資源管理專家，是該館數位學術研究團隊的一員，協助推動以大英圖書館數位典藏為基礎的創新研究；對於將電腦資訊運算的研究方法應用於歷史性的典藏之此類創新研究，提供支援、訓練、與指導。

其目前專責的計畫項目，包括以「群眾外包」(crowdsourcing) 方式，騰打具有歷史的舊海報、戲劇節目單等，以及嘗試實驗性地將機器學習方法，應用於圖書館之各類專藏。2020 年 1 月，米亞博士獲得了藝術與人文研究委員會的「集體智慧計畫」(Collective Wisdom) 之獎助，此獎助計畫的主要目的是冀期於 2020-2021 年計畫執行期間，充分掌握應用「群眾外包」與「數位參與」(digital participation) 的資訊技術於文化資產上。

米亞博士也是「與機器共存計畫」(Living with Machines) 共同研究人員，她藉由「群眾外包」方式，帶領公眾參與數位學術與文化資產典藏之整理。「與機器共存計畫」是一個重要的跨學科領域計畫，係有關歷史性資料與資料整理科學的研究計畫；該計畫大規模分析經過數位化的資源，為當今的「工業革命」，提供有關機械化處理之洞見。

Dr. Mia Ridge is the British Library's Digital Curator for Western Heritage Collections. As part of the Library's Digital Scholarship team, she helps enable innovative research based on the British Library's digital collections, offering support, training and guidance on applying computational research methods to historical collections.

Current projects involve crowdsourcing the transcription of historical playbills, and experimenting with machine learning-based methods with library collections. In January 2020 she was awarded funding by the Arts and Humanities Research Council for 'Collective Wisdom', a project that will capture the state of the art in crowdsourcing and digital participation in cultural heritage in 2020-21.

She is a Co-Investigator on the Living with Machines project, where she leads public engagement with digital scholarship and heritage collections through crowdsourcing. Living with Machines is a major inter-disciplinary historical and data science research project analysing digitised sources at scale to provide new insights into mechanisation in the industrial revolution.

演講摘要

本專題演講將介紹大英圖書館如何與數位人文領域的研究人員進行合作之概要，涵蓋最新的資料處理科學，也廣泛介紹所謂的數位化學術研究；講者將探討此研究可能帶來的機會與挑戰，也剖析合作的成功範例之重要因素，而以引領促成可同時嘉惠數位人文學者與研究圖書館之計畫為最高目的。

Abstract

This talk will provide an overview of how the British Library has worked with researchers in the digital humanities (and more recently, in data science and digital scholarship more broadly). It will discuss the opportunities and challenges of this work, and draw out themes from successful collaborations, with the ultimate aim of devising projects that benefit both digital humanities scholars and research libraries.

預錄及簡報檔

Pre-recorded Video:

<https://reurl.cc/3aYEe9>



專題演講二 2nd Keynote Speech

圖書館與數位人文合作之解密：數位人文與集中化問題

The Library's Secret Sharer: Digital Humanities and the Question of Centralization



亞列克斯·吉爾博士
Dr. Alex Gil

亞列克斯·吉爾博士任職於美國哥倫比亞大學圖書館，主要職務是數位化學術研究；透過與教師、學生、以及圖書館專業人員合作，增進人文學研究之運算與網絡技術、教學內涵、以及知識的創造。他是幾項進行中而深受肯定的計畫創始人之一，目前仍為計畫的領導人，包括：哥倫比亞大學 Studio@Butler 的協同計畫主持人，此計畫重點在為圖書館規劃建置具有科技亮點的數位學術與增進教學法之創新空間；他也是「哥倫比亞人文實驗方法小組」（Columbia's Group for Experimental Methods in the Humanities）的共同創始人和主持人，該小組積極推動跨學科領域的研究群組，專注於實驗性之人文學研究。他也是加勒比之數位研究期刊“sx archipelagos”的資深編輯，同時也是共同經營加勒比一系列數位化相關會議的經理人之一；也是“Global Outlook::Digital Humanities” (<http://www.globaloutlookdh.org/>) 的創辦人和前任主席。

亞列克斯·吉爾博士目前參與的數位相關計畫包括“Ed”（為一種僅需要最少修訂的文獻文字檔數位平台）、“wax”（為數位文化文物的微型展覽平台 <https://minicomp.github.io/wax/>）、“In The Same Boats”（內容為以視覺化呈現跨大西洋交會的 20 世紀黑人知識份子；

<https://sameboats.org/>)，以及最近的“nimble tent”(為智慧型數位工具網站 <https://nimbletents.github.io/>)、“Torn Apart/Separado”(目的在透過數位工具，動員人文學者、圖書館、學生等，以語言支援、數位典藏、技術專業等，關注全球的人權問題)、以及“Covid Maker Response”(美國哥倫比亞大學因應 COVID-19 的研究，開發供醫護使用的物資；<https://research.columbia.edu/covid/devices/covidmaker>)等。

Dr. Alex Gil is the Digital Scholarship Librarian at Columbia University Libraries. He collaborates with faculty, students and library professionals leveraging computational and network technologies in humanities research, pedagogy and knowledge production. He is among the founders of several ongoing, warmly received initiatives where he currently plays leadership roles: Co-director of the Studio@Butler at Columbia University, a tech-light library innovation space focused on digital scholarship and pedagogy; co-founder and moderator of Columbia’s Group for Experimental Methods in the Humanities, a vibrant trans-disciplinary research cluster focused on experimental humanities; senior editor of *sx archipelagos*, a journal of Caribbean Digital Studies, and co-wrangler of The Caribbean Digital conference series. He is also founder and former chair of Global Outlook::Digital Humanities.

Active digital projects include Ed, a digital platform for minimal editions of literary texts, and wax for minimal exhibits of cultural artifacts; In The Same Boats, a visualization of trans-Atlantic intersections of black intellectuals in the 20th century; and most recently, the nimble tent interventions Torn Apart/Separados and Covid Maker Response.

演講摘要

所謂「數位人文」，可被解釋為利用資訊運算法整理耙梳文化素材、為知識的創造建構新的架構之數位人文科學。此類數位人文科學的各種研究活動，其最高期待與最早的原型，是學者強烈意識到要在各種研究範疇中，耙梳媒介的、社會的、物質的等脈絡。全球各地的現代研究型圖書館，大約可為數位人文研究解決大概一半的脈絡問題，而現在大多數的研究型圖書館，已經開始聘請工作人員和設置研究單位，從事數位人文相關的工作。然而這些數位人文研究活動卻與現代的研究型圖書館之既有業務格格不入，造成不安。其主要原因是，數位人文團隊和相關計畫，相當於是未盡完全的特定圖書館功能、甚至於是整體圖書館功能的翻版，其作業內容包括詮釋資料 (metadata) 建置、數位化作業 (digitization)、協助查詢 (discovery)、資料管理運用 (curation)、自動化處理 (automatization)、資料呈現的規劃 (presentation) 等；這些作業同樣都是數位人文與一般圖書館之業務。

大型圖書館本來就有精確的人力分工，以維運圖書館的整體功能，包括提供使用者得以查詢利用典藏的文化文獻。而圖書館內的數位人文工作小團隊的實務工作者，相當於是小規模地重複了部分圖書館原本的業務，造成人力分配上的模糊地帶。由此引發一個值得思考的問題：我們應該將數位人文的相關業務，建立為一套集中化管理營運的模式，如經營一個現代圖書館一般，使之具有精確的人力分工？還是鼓勵圖書館內多元部門共同參與數位人文相關工作，而不使之成為一個集中化的單位？如此則參與數位人文的個人與相關團隊，其在所屬的圖書館這個大家庭中的關係與定位又是如何？在人力分配上，其又如何精確對應到圖書館既有的分工？

講者亞列克斯·吉爾博士將以自身的經驗，回答上述問題：首先將從技術不對稱的觀點，綜觀世界上一些例子的現況，而這也是其本身親歷世界各地，在有圖書館資源及無圖書館資源之下研究不同的運算實務之體驗。亞列克斯·吉爾博士也將以其在美國哥倫比亞大學圖書館以及「哥倫比亞人文實驗方法小組」(Columbia's Group for Experimental Methods in the Humanities) 的研究為例，說明其觀點與立場：當有一定的自由度可以進行實驗，則相較於一個想像中係以預擬的服務類別套裝組合、提供相關工具等之「特定服務提供者」(specialized service provider) 的集中化數位人文單位，毋寧是以非集中化的模式，較可裨益學術研究、以及對原本集中化管理的圖書館較為有利。但亞列克斯·吉爾博士也強力主張，非集中化的模式，必須允許自由與其他機構的其他團隊合作，尤其如果是在缺乏經費的情形下，則此種凝聚智慧的工作團隊的組合，將可發揮強大力量，達到高度效益，對社會整體都是極為有利的。

Abstract

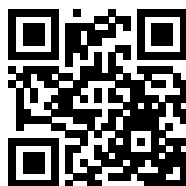
Digital Humanities can be understood to be the productive union of algorithmic approaches to cultural corpora and the architecture of new genres of knowledge production in the humanistic sciences. This set of activities finds its highest and most original forms when it is intensely aware of the mediatic, social and material contexts in which it goes about its business. Modern research libraries around the world provide half of that context, and most of them have already begun the process of hiring staff and institute units to work on digital humanities. And yet, these digital humanities are an uncomfortable travel companion to modern research libraries. The main reason for this disquiet is the fact that digital humanities teams and projects make imperfect copies of specific library functions, and a copy of the library as a whole: metadata, digitization, discovery, curation, automatization, presentation are all part of the domain of digital humanities and of libraries. While large libraries have very

precise divisions of labor to serve the overall function of maintaining and providing access to a collection of cultural artifacts, the small teams of digital humanities practitioners replicate these efforts at a smaller scale, and with a blurring of the divisions of labor. This begs the questions: should we have a centralized digital humanities that works as modern libraries do, with a precise division of labor, or should we encourage a diversity of practices within libraries that can never be centralized? What should be the relationship of these individuals and teams to the centralized home where they live? How can a mirror image of the library count as a division of labor of the library? I will try to answer these questions by first providing an account of the state of the field around the world in the context of asymmetries in access to technology, as I've experienced it during my travels studying different practices of computing within and without libraries. I will then use case studies from our work at Columbia University Libraries and Columbia's Group for Experimental Methods in Humanistic Research, I will defend the position that a de-centralized approach, if allowed relative freedom to experiment, can produce better scholarship, and benefit the centralized library more than a unit imagined as a specialized service provider with a pre-determined portfolio of genres and tools. This de-centralized approach, I will contend, must also be allowed to collaborate with other teams in other institutions freely and in the absence of grants. If this is the case, the enormous talent of this generalist work-force can even become a beneficent and powerful actor in civic society at large.

預錄及簡報檔

Pre-recorded Video:

<https://reurl.cc/3aYEe9>



【子題一】

圖書館發展數位人文學術服務

Session I

**Developing Digital Humanities Academic Services
in Libraries**



陈建龙 Jianlong Chen

陈建龙，北京大学信息管理系教授、博士生导师，北京大学图书馆馆长，教育部高校图工委主任、中国图书馆学会副理事长，曾赴美国加州大学欧文分校 (UC Irvine) 访问，主要研究方向：信息服务、信息资源管理、信息化与信息行为等。代表作有《信息服务学导论》《信息市场经营与信息用户》等专著和《信息服务模式研究》《大学图书馆现代化转型发展刍议》等论文。

Jianlong Chen is Professor and PHD supervisor of Information Management Department, Library Director, Peking University. He also serves as Director of Steering Committee for Academic Libraries of China, Vice President of Library Society of China, and visiting scholar of UC Irvine. Research directions include information service, information resource management, informatization and information behavior. Academic achievements include monographs such as "Introduction to Information Service Studies", "Information Market Management and Information Users" and articles such as "A Study of Information Service Models" and "Discussion on the Transformation of Academic Libraries Modernization".



别立谦
Liqian Bie

别立谦，北京大学图书馆副馆长，研究馆员。图书馆学硕士。分管信息资源保障与服务、古籍、特藏以及分馆建设等业务工作。兼任中国图书馆学会学术研究委员会资源建设与共享专业委员会委员、北京高等教育学会图书馆工作研究会副理事长、高校图书馆数字资源采购联盟（DRAA）理事。

Liqian Bie is the Deputy Director and Research Librarian of Peking University Library. She has a master's degree in Library Science. Currently she is responsible for resources development organization, preservation and services on books, periodicals, ancient books, special collections and electronic databases, and branch library development. She is a member of the Academic Research Committee on the Resource Development and Sharing, Library Society of China; vice chairman of Beijing Academic Library Association of Beijing Association of Higher Education; council member of the Digital Resources Acquisition Alliance (DRAA) of Chinese Academic Libraries.



童云海
Yunhai Tong

童云海，博士，北京大学信息科学技术学院教授，博士生导师。现任北京大学图书馆副馆长，负责数据资源建设与服务、在线图书馆建设、信息技术应用与保障等业务工作，兼任移动数字医院系统教育部工程研究中心副主任，中国通信协会智慧医疗专家委员会委员。目前主要从事大数据分析建模和数据隐私保护的研究和教学工作。

Yunhai Tong, PHD is professor in the School of Electronics Engineering and Computer Science (EECS), Peking University. He is currently the Deputy Director of the Peking University Library, responsible for the development and service of data resources, digital library, and the application and support of information technology. He is also the Deputy Director of Engineering Research Center for Mobile Digital Hospital System Ministry of Education, a member of Smart City Committee of Experts China Institute of Communication. His research interests include big data analysis modeling and data privacy protection.

交融式数字图书馆场景中的数字学术服务 ——以北京大学图书馆为例

Digital Scholarship Services in the Scene of Unified Digital Library -- Case of Peking University Library

陈建龙

Jianlong Chen

北京大学图书馆馆长

Director, Peking University Library

别立谦

Liqian Bie

北京大学图书馆副馆长

Deputy Director, Peking University Library

童云海

Yunhai Tong

北京大学图书馆副馆长

Deputy Director, Peking University Library

摘要

数据的爆炸性增长，带来了科学研究方法的变革，催生了科学研究的第四范式，即“数据密集型科学发现”，为科学发展与学术交流提供了新的机遇和思路，同时也对学术图书馆的资源建设与服务设计提出了新的挑战。数字学术作为 21 世纪出现的一种新的研究形式，越来越受到科研人员的青睐，也带来了高等教育的文化变革和学术交流方式变革，越来越多的学术图书馆将数字学术服务作为未来转型的重要方向和前行动力。本文从交融式数字图书馆的概念和特征入手，分析数字学术的概念范畴和需求，在分享北京大学图书馆数字学术服务实践的基础上，探讨学术图书馆提供数字学术服务的关键问题和战略构想

关键词：交融式数字图书馆、数字学术、用户导向、人机交融

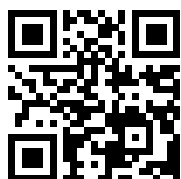
Abstract

The explosive growth of data has brought about the change of scientific research methods and gave birth to the eScience -- A Transformed Scientific Method, provided new opportunities and ideas for scientific development and academic exchanges, and also posed new challenges to the resource construction and service design of academic libraries. As a new form of research in the 21st century, digital scholarship is becoming more and more popular among scientific researchers. It has also brought about the cultural change and the change of academic exchange mode of higher education. More and more academic libraries regard digital scholarship service as an important direction and driving force of future transformation. Starting with the concept and characteristics of the Unified Digital Library, this article analyzes the conceptual category and needs of digital scholarship, and discusses the key issues and strategic ideas for academic libraries to provide digital scholarship services on the basis of sharing the practice of digital scholarship services in Peking University Library.

Keywords: Unified Digital Library, Digital Scholarship, User Oriented, Human-Computer Unified

預錄及簡報檔

<https://pse.is/3e37pp>



壹、引言

数据的爆炸性增长，带来了科学研究方法的变革，催生了科学研究的第四范式，即图灵奖获得者吉姆·格雷（Jim Gray）所称的“数据密集型科学发现”¹。格雷呼吁资助开发数据采集、管理和分析的工具，以及交流与发布的基础设施，还强调要建立起与传统图书馆一样普及和强大的现代化数据与文件存储体系。²

数据密集型科学发现的提出，为科学发展与学术交流提供了新的机遇和思路，同时对学术图书馆的资源建设与服务设计提出了新的挑战。数据的采集、管理、分析、可视化和长期保存等相关工作，需要一个能把所有的科学数据和文献资源联系在一起，并提供数据和文献适时交互操作的平台，在线科学数据平台应运而生。基于在线科学数据平台，科研人员既可以在阅读某篇论文时查看其原始实验数据，甚至可以重做分析，也可以在查看某些数据时查出其相关的文献，这是一个全新的数字学术（digital scholarship）环境。数字学术作为 21 世纪出现的一种新的研究形式，越来越受到科研人员的青睐，也带来了高等教育的文化变革和学术交流方式变革³。面对这一变革，越来越多的学术图书馆也将数字学术服务作为未来转型的重要方向和前行动力。

本文从交融式数字图书馆的概念和特征入手，分析数字学术的概念范畴和需求，在分享北京大学图书馆数字学术服务实践的基础上，探讨学术图书馆提供数字学术服务的关键问题和战略构想。

貳、交融式数字图书馆与数字学术概念辨析

交融式数字图书馆或将成为数字学术的重要场景，理清两者的概念内涵和演变发展至关重要。

一、交融式数字图书馆的概念和特征

自上个世纪九十年代开始，数字图书馆这一概念逐渐引起图书馆学、情报学和计算机科学技术等领域的重视，产出大量的理论研究成果与最佳实践案例。美国、英国、加拿大、

-
1. Jim Gray, Alex Szalay. eScience-A Transformed Scientific Method [J]. presentation to the Computer Science and Technology Board of the National Research Council, Mountain View, CA, 2007. 或 [EB/OL]. http://jimgray.azurewebsites.net/talks/NRC-CSTB_eScience.ppt.
 2. 潘教峰, 张晓林. 第四范式：数据密集型科学发现. 北京：科学出版社, 2012: v.
 3. 刘玲玲. 美国大学图书馆数字学术服务调查研究 [J]. 图书馆工作与研究, 2021, (01): 19-28.

法国、日本等国家纷纷投入巨大的人力物力开展数字图书馆建设⁴。1994年,美国国家科学基金会(NSF)等机构联合资助了DLI项目(Digital Library Initiatives),开展数字化资源生成、收集、组织、存储等方面关键技术的研发,并于1998年开展DLI二期项目,探索数字图书馆相关基础性理论前沿研究⁵。1995年,英国eLib项目(Electronic Libraries Programme)资助了六十多个数字图书馆项目,分三期进行,致力于推动英国高等教育信息资源与技术的开发及利用⁶。我国数字图书馆发展也较早,国家图书馆在1995年成立了调研数字图书馆的专项小组,1996年在北京召开的国际图联(IFLA)大会设立了“数字图书馆”专题,此后数字图书馆研究论文与专著增长迅速⁷。数字图书馆二十余年的发展历程,逐步经过了早期以图书馆集成管理系统(ILS)为主的摸索阶段,后来针对“复合图书馆”纸电资源管理与服务的“资源发现”平台为特色的深化阶段,当前利用云服务和共享知识库连接图书馆、数据资源和用户的服务应用平台为特征的转型阶段。新型的数字图书馆,将实现人(用户和馆员等)、机器(计算机和智能终端等)、信息(数据和知识等)三者充分交互,深度融合,一体化发展,本文称其为交融式数字图书馆。

交融式数字图书馆具有多方面的特征:一是用户导向,二是多源融合,三是数字孪生,四是人机交融,五是人信共济。

用户导向。用户真正成为图书馆服务工作的出发点和归宿。这是指“以用户有待解决的问题及其目的为依据的服务方向、以用户利益为前提的价值取向和以用户的期望、体验和效用为核心的行为意向。”⁸具体包括真正帮助个人和机构用户解决问题而不仅仅是解答需求的用户待解问题导向,同步提升用户信息素质而不仅仅是检索技能的用户素质导向,注重用户使用效果而不仅仅是服务过程的用户效用导向,关注用户满意度和价值实现而不仅仅是服务业绩的用户价值导向。

多源融合。图书馆各种类型的文献和数据集以及承载于文献和数据集中的主题知识等多源信息融合,数据资源和知识资源将成为图书馆的核心资源,空间资源、馆员资源也将以数据的形态出现。数据资源既包括图书馆的文献资源,也包括实验室的大规模、多模态的实验数据和可能来自个人生活中的各种数据。多模态的海量数据涉及众多学科,未来的科学研究将融合使用这些数据,图书馆需要创建工具或平台,以支持从数据采集、验证、加工、管

4. 郑淼磊,梁晨.近10年国外数字图书馆重大项目进展研究[J].科技情报开发与经济,2010,20(21):8-10+13.

5. Stephen M. Griffin. NSF/DARPA/NASA Digital Libraries Initiative: A Program Manager's Perspective [M]. D-Lib Magazine. July/August 1998: 4-11.

6. eLib: The Electronic Libraries Programme 1995-2001 [EB/OL]. <http://www.ukoln.ac.uk/services/elib/>.

7. 孙承鉴,申晓娟,刘刚.我国数字图书馆发展十年回顾——综述[J].数字图书馆论坛,2006,(01):1-13.

8. 陈建龙,申静.信息服务学导论[M].北京:北京大学出版社,2017,40.

理、分析和长期保存的整个流程。需要仔细考虑哪些数据应该长期可用，并且必须采集或创建配套的元数据而使这些数据的长期可用成为可能。

数字孪生。以多源融合为基础，实体图书馆和虚拟图书馆将成为一个不可分割的整体。充分利用物理模型、实时传感、运行历史等数据，集成多学科、多物理量、多尺度、多概率的仿真过程，在虚拟空间中完成映射，从而反映相对应的实体装备的全生命周期过程。实体图书馆和虚拟图书馆，将通过一个或多个重要的、彼此依赖的映射系统，实现两者的和谐共生。

人机交融。图书馆在用户面前，不仅仅是一个庞杂的海量数据库，数据分析将覆盖整个工作流的所有活动，数据库的建设不再是只能读取数据的文件集，更需要具有建模、分析和数据可视化等功能。如何为用户提供其关心和需要的数据资源，并利用机器的强大计算功能，为用户提供加工后的信息和知识，并提供沉浸式的交互手段，构建起人与机器之间的密切关系。让用户“畅游”图书馆。

人信共济。人和信息之间是一种互动共生关系，这里的“人”既包括信息的使用者（科研人员），也包括数据资源的提供者（图书馆员）。图书馆要分析和研究信息流的特征及其用户、流程和源头，通过有关概念的关联和信息的流动特征，设计学术服务场景，做到以人为本，趋利避害。

可见，交融式数字图书馆是图书馆数字化、网络化、智能化发展到一定阶段的产物，是更加便捷、灵活、智慧的随身图书馆，将为数字学术活动提供广阔的舞台和优质的服务。

二、数字学术的概念及其研究进展

数字学术环境下学术研究的需求和特点将会带来哪些变化，在新时期图书馆该如何定位，如何应对这些需求的变化，是全球学术图书馆面对的重大挑战。

（一）数字学术的缘起和概念

有研究表明，数字学术的概念最早出现于 20 世纪 90 年代末的英国，是一种将新技术、数据分析工具等应用于学术研究的理念，是在数字科学、电子科学、电子学术、网络基础设施等概念的基础上发展演进而来的，涉及数据库技术、高性能网络、高性能

计算、可视化技术等一系列的信息技术⁹。数字学术利用数字证据、调查方法、研究、出版和保存的方法来开展学术活动和实现研究目标¹⁰。华盛顿大学图书馆提出数字学术可理解为广泛利用数字媒体的独特功能为教学和科研开辟一种或多种新可能性，包括但不限于新形式的协作、新形式的出版以及可视化和分析数据的新方法，并且认为数字学术通常包括非正式出版的原生数字、多媒体、数据库作品，以及对其他原生数字材料、数字文本和图像、数字音乐或艺术、数据集的分析，这种形式的学术数据、展示和传播意味着出版和传统的图书馆收藏发生了转变，是数字技术应用于学术的自然演变和适应¹¹。

《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》描绘了美好的数字化战略图景。在数字化转型发展的新时代新阶段，数字经济、数字政府、数字文化、数字社会、数字生态等迅猛发展的今天，数字学术的概念也将取得更大发展。数字学术将成为人们以数据资源为基础、以数据资产和数据要素为对象的新的知识生产和文化生活方式。图书馆也将迎来包括数据资源建设、数据资产管理、数字技术应用、数字素养培育等多方面内容的数字学术服务新的发展空间。

（二）国外图书馆提供数字学术服务的现状

早在 21 世纪初，北美地区的研究图书馆协会 (Association of Research Libraries, 简称 ARL) 便已开始将数字学术运用到学术研究领域，开展了图书馆数字学术研究的实践和技术应用。ARL 成员图书馆通过创新计划来支持数字学术，致力于数字图像、文本和声音的收集和管理，并创建了与这些材料一起使用的工具，作为该机构使命的核心部分。¹² 2016 年 1 月，研究图书馆协会面向成员图书馆发放了针对数字学术主题的问卷，并根据问卷调查结果，于同年 5 月发布了《SPEC Kit 350：支持数字学术》调查报告¹³。2016 年 6 月，美国大学与研究图书馆协会 (Association of College and Research Libraries, 简称 ACRL) 发布的《2016 高校图书馆发展趋势》将数字学

9. CONSORTIUM T N M. NMC horizon report 2017 library edition [EB/OL]. <https://www.issuelab.org/resources/27498/27498.pdf>: 40.

10. Abby Smith Rumsey. New-model scholarly communication: Road map for change; proceedings of the Scholarly Communication Workshop, F, 2011 [C]. <http://uvasci.org/institutes-2003-2011/SCI-9-Road-Map-for-Change.pdf>.

11. University of Washington Libraries. About Digital Scholarship [EB/OL]. <https://www.lib.washington.edu/digitalscholarship/about>.

12. Digital scholarship support in ARL Member Libraries: An overview [EB/OL]. <https://www.arl.org/digital-scholarship-support-in-arl-member-libraries-an-overview/>.

13. Rikk Mulligan. SPEC Kit 350: Supporting Digital Scholarship [EB/OL]. <https://publications.arl.org/Supporting-Digital-Scholarship-SPEC-Kit-350/>.

术列为未来图书馆领域发展的十二大趋势之一¹⁴。欧洲研究型图书馆联盟《研究型图书馆为数字时代的可持续知识提供动力：2018-2022 战略》中提出 2022 年愿景是“研究型图书馆的作用将在于为数字时代的可持续知识提供动力”，明确了三个战略方向，包括：图书馆成为创新学术交流的平台、数字技能和服务的枢纽、研究基础设施的合作伙伴。¹⁵

2021 年 3 月，欧盟委员会发布《欧洲数字十年：2030 年实现数字驱动欧洲》，提出欧洲数字化指南针，包括：培养具有数字技能的公民和高技能的数字专业人员，建立安全、高性能和可持续的数字基础架构，企业数字化转型，公共服务数字化。¹⁶ 图书馆的数字化转型以及为数字学术服务的前景令人期待。

(三) 数字学术服务研究进展

国外高校图书馆或研究型图书馆在数字学术服务的理论与实践方面已有较为深厚的积淀。根据涂志芳和徐慧芳等人对 15 所高校图书馆的调研整理结果可见，各校图书馆对数字学术服务的命名有所不同，体现出各校图书馆对数字学术服务的概念理解与定位的差异，分别侧重于提供服务、提供研究性服务、提供物理空间、提供交流与共享环境、提供资源与技术等不同方面，开辟了物理空间、数字科研工具、研究数据服务、学术交流与数字出版服务、数字人文服务、数字学术研讨与培训服务、数字技术支持服务等服务项目。¹⁷ 针对数字学术服务多样化的特点，还有研究将高校图书馆相关服务项目总结为九类的¹⁸、七类的¹⁹或五类的²⁰。虽然表述有所不同，但归根结底，高校图书馆提供数字学术服务的目的是为教师、学生和员工提供技术支持、设备和协作机会，开创新的知识创造模式。尤其是作为数字学术的重要组成部分，数字人文的研究受到广泛关注。主要集中在以下 5 个方面：培养新一代数字人文学者，将数字方法和工具纳入到他

14. ACRL Research Planning and Review Committee. 2016 top trends in academic libraries: a review of the trends and issues affecting academic libraries in higher education. College and research libraries news. v.77,no.6, 2016 [EB/OL]. <https://crln.acrl.org/index.php/crlnews/article/view/9505/10798>.

15. Research Libraries Powering Sustainable Knowledge in the Digital Age: LIBER Europe Strategy 2018-2022. [EB/OL]. <https://libereurope.eu/wp-content/uploads/2020/10/LIBER-Strategy-2018-2022.pdf>.

16. Europe's Digital Decade: Commission sets the course towards a digitally empowered Europe by 2030. [EB/OL]. <https://digital-strategy.ec.europa.eu/en/news/europes-digital-decade-commission-sets-course-towards-digitally-empowered-europe-2030>.

17. 涂志芳, 徐慧芳. 国内外 15 所高校图书馆数字学术服务的内容及特点 [J]. 大学图书馆学报, 2018, 36(04): 29-36.

18. 曾粤亮. 美国高校图书馆数字学术中心服务调查与分析 [J]. 图书与情报, 2017, (04): 25-33.

19. 胥文彬. 国外高校图书馆数字学术服务调查分析 [J/OL]. 情报杂志. <https://kns.cnki.net/kcms/detail/61.1167.G3.20210415.1406.024.html>.

20. 鄂丽君. 加拿大高校图书馆数字学术支持调查分析 [J]. 图书馆论坛, 2019, 39(05): 162-169.

们的研究、教学、学习中去；为数字人文项目建设的整个生命周期提供咨询与技术支持服务；致力于数字人文软件工具及关键方法生态系统的研发；支持、鼓励并孵化创新的数字人文项目，创造跨学科协作的机会；为数字学术服务提供空间服务、资源与资金支持²¹。

概括而言，数字学术服务包括数字人文服务、数字学术空间服务、科学数据服务（含开放存取）、数字技术支持服务（数字化服务、可视化服务）、数字出版与版权服务、数字科研工具、数字项目支持、数字学术教育（研讨与培训）、教学与课程支持、合作与互动式研究等众多可行的方式。相关研究的进一步深化，必将促进数字学术的更大发展。

参、北京大学图书馆数字学术服务的积极尝试

数字学术的目标还是在于为学术研究服务，学术研究因数字学术环境的变化而呈现出更多的动态特征，图书馆在其中更多地要起到中介或者桥梁的作用，发挥信息专家的角色和作用，有效建立起图书馆、学者、技术专家三方面的链接，促成合作、协同作战，在文献资源、电子资源和数据资源融合的基础上，借助一定的技术或工具提供学术服务，这也是促成数字学术发展的核心所在。

一、设立研究机构持续探索数字图书馆发展

北京大学图书馆在探索数字学术发展方面起步早。为求新的发展，必须借助信息技术的力量，加强合作与共享。早在 1999 年，图书馆、信息科学技术学院和 CALIS 管理中心联合成立了北京大学数字图书馆研究所，并挂靠在北京大学图书馆。研究所以数字图书馆建设的理论、方法和实践的研究为先导，通过与国内外同行的广泛交流与合作，针对馆藏各种类型的数据资源进行建设实践和服务，积累知识，培养人才，建立系统化的北京大学数字图书馆，以期为我国的数字化工程作贡献。研究所成立最初几年，围绕数字图书馆研究这一主题，开展了有关数字图书馆模式、标准规范（元数据、数字加工标准等）、关键技术、互操作层与互操作标准等的研究，并进行了小规模应用实践，为中国高等教育数字图书馆和北京大学数字图书馆的建设奠定了技术基础。

近年，随着数据科学的科研基础设施的不断进步，科学研究第四范式的广泛应用，学术

21. 刘玲玲. 美国大学图书馆数字学术服务调查研究 [J]. 图书馆工作与研究, 2021, (01): 19-28.

交流移动互联网化，虚拟现实与人工智能快速发展，学术环境不断变化，学术交流系统也随之不断变革。数字图书馆研究所加强了在大数据、人工智能、区块链等技术上的合作探索，力求在图书馆发展模式模型上、在探索学术生态转型上取得突破和创新。

二、着力搭建学术生态系统

面对学术环境的巨大变化，图书馆的角色和价值也发生了变化，图书馆员需要学习与掌握新技术、新技能，并重新找到角色定位，才能够为高校的学术交流提供相适应的服务。从 2011 年开始，北京大学图书馆以机构知识库建设为契机，设计、构建与推动北京大学新一代学术交流生态系统的建设与发展，构建了包括科研准备、搜索发现、研究分析、成果写作、成果发表、成果推广、科研评价在内的学术交流生态系统。

（一）北京大学开放研究数据平台

北京大学开放研究数据平台 (Peking University Open Research Data)²²，是由北京大学图书馆、国家自然科学基金 - 北京大学管理科学数据中心 (Data Center for Management Science, NSFC-PKU)、北京大学科研部、社科部联合主办和推出的开放数据平台，平台以“规范产权保护”为基础，以“倡导开放科学” (Open Science) 为宗旨，定位于功能完备、服务优质、具有国际影响力的开放数据平台，吸引国内外研究人员发布与使用数据，为基于数据的研究、决策提供国际一流的支撑平台。鼓励研究数据的发布 (Publish)、发现 (Discover)、再利用 (Reuse) 和再生产 (Reproduce)，促进研究数据引用 (Citation) 的实践和计量，并探索数据长期保存 (Preservation)，培育和实现跨学科的协同创新。

（二）北京大学机构知识库

北京大学机构知识库²³，是支撑北京大学学术研究的基础设施，致力于收集并保存北京大学教师和科研人员的学术与智力成果；为北京大学教师、科研人员和学生的学术研究和学术交流提供系列服务，包括存档、管理、发布、检索和开放共享。北京大学机构知识库为嵌入北京大学科研信息管理系统做好整合准备，面向教师学者，突出科研成果的学术价值和影响力，全面回溯北大历年学术成果，引入替代计量学 (altmetrics) 理念，全面揭示成果文献的社会影响力。

22. 北京大学开放研究数据平台 [EB/OL]. <https://opendata.pku.edu.cn/>.

23. 北京大学机构知识库 [EB/OL]. <http://ir.pku.edu.cn/>.

(三) 北京大学期刊网

北京大学期刊网 (PKU-OAJ)²⁴，由北京大学社科部、科研部牵头，各期刊编辑部配合，北京大学图书馆负责承建。平台以北京大学 e-Research 的学术环境为依托，整合北京大学的学术期刊资源，为各期刊编辑部提供信息发布和期刊数字出版等方面的支撑，提供具有国际化水平的文章内容展示功能，实现期刊的集成检索功能，以及编者、读者、作者互动的学习交流功能，是图书馆推动开放获取运动在北京大学的最佳实践，在繁荣学术交流和研究方面发挥了积极的作用。

(四) 北大学者主页

北京大学学者主页 (Scholars @ Peking University)²⁵，旨在为本校学者自主设立学术型个人主页提供便捷服务。平台建站工具简单易用，从最初的设计到后期的管理，均可由学者用户自行掌控，个性化定制，随时更新，动态而全方位地展示北大学者的个人学术风采。

三、引领探索数字人文研究

在数字人文研究方面，北京大学图书馆起步也较早。在 2016 年至 2018 年间连续举办三届数字人文论坛^{26. 27. 28}，开设了数字人文工作坊和讲座活动。图书馆在国内率先发起数字人文的交流与研究，搭建了数字人文的国际化交流平台，为人文学科的学者学习数字技能提供了帮助，在国内掀起数字人文研究之风气。

2018 年开始，北京大学信息管理系在校内开设了“数字人文”研究生课程，讲授相关理念与技术；成立了数字人文开放实验室，与哈佛大学等校建立了国际合作关系，共同攻关国家级科研课题。2021 年 1 月，北京大学正式成立数字人文研究中心，致力于整合北京大学各院系相关学科的研究力量，共同构筑跨学科、多层次、协同创新的数字人文平台²⁹。

总体而言，北京大学图书馆关注数字学术研究起步早，不仅面对人文学科的发展，而是

24. 北京大学期刊网 [EB/OL]. <http://www.oaj.pku.edu.cn>.

25. 北京大学学者主页 [EB/OL]. <http://scholar.pku.edu.cn/>.

26. 朱本军, 聂华. 跨界与融合: 全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述 [J]. 大学图书馆学报, 2016, 34(05): 16-21.

27. 朱本军, 聂华. 互动与共生: 数字人文与史学研究——第二届“北京大学数字人文论坛”综述 [J]. 大学图书馆学报, 2017, 35(04): 18-22.

28. 第三届北京大学数字人文论坛“孵化与实践: 需求驱动下的数字人文项目”日程安排 [EB/OL]. <https://www.lib.pku.edu.cn/portal/cn/news/0000001780>.

29. 北京大学数字人文研究中心 [EB/OL]. <http://digitalhumanities.pku.edu.cn/>.

放眼于全部学科，考虑更大范畴的数字学术支撑服务，不仅是数字人文，还有包括社会科学和自然科学领域在内的数字学术的发展，从构建新一代学术交流生态系统的角度全面规划，努力搭建与北京大学教学科研发展的相适应相匹配的学术生态和服务环境，这些积极有益的实践，为进一步开展数字学术服务打好了前站和基础。

肆、交融式数字图书馆场景中的数字学术服务架构

除了在实际行动上开展了上述一系列的研究与平台建设外，北京大学图书馆放眼未来二十年，于 2018 年讨论制定了《北京大学图书馆 2035 愿景和 2019-2022 行动纲领》，2020 年开始在全校范围内颁布实施全新修订的《北京大学文献保障与信息服务体系管理办法》。在 2020-2021 年制定图书馆十四五规划时，图书馆更多地关注与考虑信息化和数字化发展战略，从用户及其待解决问题出发，通过顶层设计、层层部署、加强管理和落地实施，先后进行了机构重组和岗位聘任，设立数据资源服务中心，加强算力赋能，资源与服务一体化发展转型，以及文献典藏与空间服务的整体规划等，加强为学校教学与科研数据资源的服务支撑以及做好数字学术服务，实施战略转向。

一、“十四五”规划将建设交融式数字图书馆作为信息化建设的重点

“十四五”时期，北京大学图书馆信息化建设将紧密围绕图书馆 2035 年愿景中明确的“使命”和 2019-2022 年行动纲领中的“行动”，增强机遇和危机意识，集中全馆智慧，将建设交融式数字图书馆作为支撑学校数字学术发展的重要战略规划。这个五年中，图书馆要建设数字学术服务体系、系统驱动的知识产权信息服务体系、开放获取资源组织和服务系统、算力驱动的计算服务体系，以若干工程的建设为抓手，建设交融式数字图书馆。

二、以“人机交融”“人信共济”为理念，充分发挥智能化图书馆的价值

在建设交融式数字图书馆过程中，强调以“人机交融”“人信共济”为理念，力争发挥智能化图书馆的拟人、赋能、机智、高效等价值，具体包括：

构建以人为本的图书馆信息化体系，坚持“流程优化，人机交融”重要理念，以信息化主导人与机器的共生关系，实现图书馆的拟人价值。

提升“馆员服务 - 机器算力 - 信息资源”交融为核心的治理体系和治理能力，坚持现代化方向和新发展理念，实现图书馆的赋能价值。

拓展基于信息流的服务场景布局，坚持“概念关联，人信共济”重要理念，以场景化增进人与信息的互动关系，实现图书馆的机智价值。

完善用户效用和文化繁荣为指针的服务体系和服务能力，培育“图书馆”的信息力，实现图书馆的高效价值。

三、开展支撑数字学术服务的三项重点工程

北京大学图书馆将以图书馆运营效能评估指标体系和评价系统建设为牵引，以信息化安全体系、标准体系和运维体系等“三大体系”为保障，立足当前，放眼未来，构筑起以信息基础设施建设为支撑、以业务创新工程为核心、以大数据中心（数据治理体系）建设为抓手的信息化发展能力，在建设交融式数字图书馆的进程中，重点通过以下三个工程为数字学术提供支撑。

首先，实施资源治理创新工程。加快馆藏珍贵文献的数字化进程，建成基于标准化流程的数据化加工管理系统，提升数字化加工处理的智能化程度；在数字版权保护系统的支撑下，数据资源面向全校师生开放；探索和建立馆藏各类文献信息资源的智能揭示、组织方法，推进异质文献信息资源的一体化检索和发现，面向读者需求提供文献信息资源的主动服务、个性化推荐及特定主题的服务；建立馆藏数据资源统一标识符体系，促进馆藏数据资源的使用与引用，提升馆藏数据资产的可管理度与可见度；加快推进馆藏数据资源长期保存系统的研发，实现数据库、图片、音视频、文档、研究数据等多种媒体的馆藏文献资源和数据资源的长期可靠保存及面向个人和机构的长期保存服务。

第二，实施协同服务创新工程。完善以机构知识库为代表的北京大学学术资产体系，逐步打造覆盖师生学术研究全流程、收藏科研机构学术产出全品类的机构知识库；构建北京大学知识产权信息服务平台，与学校有关单位协同，为北京大学相关团队的知识产权布局、潜在高价值专利发现、知识产权申请等服务；进一步完善数字教学参考资料系统，支撑学生自主学习和线上教育教学，加强图书馆支撑学科建设、人才培养的力度。

第三，开展大数据中心建设工程。强化数据资源规划建设，形成关键数据资源目录体系，建立统一高效的图书馆数据治理体系，构建安全可靠的大数据平台；以制度和标准强化数据资源管理，建立数据采集、管理、交换、处理、安全、隐私保护相关的数据制度、标准与规范。推动数据资源分类分级管理，提高数据的有效性、安全性、可用性；推动数据资源开发与利用，支撑图书馆运营管理与领导决策，提供面向特定用户群体的信息资源服务；建成面向数据开放共享的数据资源开放服务平台，加强数字出版，推进开放获取。

伍、结语

在师生学习方式、教学模式和科研范式的深刻变革中，数字学术的发展和繁荣，给数字图书馆的建设和发展带来了新的机遇和挑战，以用户导向、多源融合、数字孪生、人机交融、人信共济为主要特征并具有强大功能的交融式数字图书馆似将应运而生。

交融式数字图书馆场景中为数字学术活动提供服务，是图书馆服务的重要生长点，或将成为图书馆超越现有的发展格局、创造独特的新价值的有益探索。

参考文献

1. Jim Gray, Alex Szalay. eScience-A Transformed Scientific Method [J]. presentation to the Computer Science and Technology Board of the National Research Council, Mountain View, CA, 2007. 或 [EB/OL]. http://jimgray.azurewebsites.net/talks/NRC-CSTB_eScience.ppt.
2. 潘教峰, 张晓林. 第四范式: 数据密集型科学发现. 北京: 科学出版社, 2012: v.
3. 刘玲玲. 美国大学图书馆数字学术服务调查研究 [J]. 图书馆工作与研究, 2021, (01): 19-28.
4. 郑淼磊, 梁晨. 近 10 年国外数字图书馆重大项目进展研究 [J]. 科技情报开发与经济, 2010, 20(21): 8-10+13.
5. Stephen M. Griffin. NSF/DARPA/NASA Digital Libraries Initiative: A Program Manager's Perspective [M]. D-Lib Magazine. July/August 1998: 4-11.
6. eLib: The Electronic Libraries Programme 1995-2001 [EB/OL]. <http://www.ukoln.ac.uk/services/elib/>.
7. 孙承鉴, 申晓娟, 刘刚. 我国数字图书馆发展十年回顾——综述 [J]. 数字图书馆论坛, 2006, (01): 1-13.
8. 陈建龙, 申静. 信息服务学导论 [M]. 北京: 北京大学出版社, 2017, 40.
9. CONSORTIUM T N M. NMC horizon report 2017 library edition [EB/OL]. <https://www.issuelab.org/resources/27498/27498.pdf>: 40.
10. Abby Smith Rumsey. New-model scholarly communication: Road map for change; proceedings of the Scholarly Communication Workshop, F, 2011 [C]. <http://uvasci.org/institutes-2003-2011/SCI-9-Road-Map-for-Change.pdf>.
11. University of Washington Libraries. About Digital Scholarship [EB/OL]. <https://www.lib.washington.edu/digitalscholarship/about>.
12. Digital scholarship support in ARL Member Libraries: An overview [EB/OL]. <https://www.arl.org/digital-scholarship-support-in-arl-member-libraries-an-overview/>.

13. Rikk Mulligan. SPEC Kit 350: Supporting Digital Scholarship [EB/OL]. <https://publications.arl.org/Supporting-Digital-Scholarship-SPEC-Kit-350/>.
14. ACRL Research Planning and Review Committee. 2016 top trends in academic libraries: a review of the trends and issues affecting academic libraries in higher education. College and research libraries news. v.77,no.6, 2016 [EB/OL]. <https://crln.acrl.org/index.php/crlnews/article/view/9505/10798>.
15. Research Libraries Powering Sustainable Knowledge in the Digital Age: LIBER Europe Strategy 2018-2022. [EB/OL]. <https://libereurope.eu/wp-content/uploads/2020/10/LIBER-Strategy-2018-2022.pdf>.
16. Europe's Digital Decade: Commission sets the course towards a digitally empowered Europe by 2030. [EB/OL]. <https://digital-strategy.ec.europa.eu/en/news/europes-digital-decade-commission-sets-course-towards-digitally-empowered-europe-2030>.
17. 涂志芳, 徐慧芳. 国内外 15 所高校图书馆数字学术服务的内容及特点 [J]. 大学图书馆学报, 2018, 36(04): 29-36.
18. 曾粤亮. 美国高校图书馆数字学术中心服务调查与分析 [J]. 图书与情报, 2017, (04): 25-33.
19. 胥文彬. 国外高校图书馆数字学术服务调查分析 [J/OL]. 情报杂志. <https://kns.cnki.net/kcms/detail/61.1167.G3.20210415.1406.024.html>.
20. 鄂丽君. 加拿大高校图书馆数字学术支持调查分析 [J]. 图书馆论坛, 2019, 39(05): 162-169.
21. 同 3.
22. 北京大学开放研究数据平台 [EB/OL]. <https://opendata.pku.edu.cn/>.
23. 北京大学机构知识库 [EB/OL]. <http://ir.pku.edu.cn/>.
24. 北京大学期刊网 [EB/OL]. <http://www.oaj.pku.edu.cn>.
25. 北京大学学者主页 [EB/OL]. <http://scholar.pku.edu.cn/>.
26. 朱本军, 聂华. 跨界与融合: 全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述 [J]. 大学图书馆学报, 2016, 34(05): 16-21.
27. 朱本军, 聂华. 互动与共生: 数字人文与史学研究——第二届“北京大学数字人文论坛”综述 [J]. 大学图书馆学报, 2017, 35(04): 18-22.
28. 第三届北京大学数字人文论坛“孵化与实践: 需求驱动下的数字人文项目”日程安排 [EB/OL]. <https://www.lib.pku.edu.cn/portal/cn/news/0000001780>.
29. 北京大学数字人文研究中心 [EB/OL]. <http://digitalhumanities.pku.edu.cn/>



王麗蕉
Li-chiao Wang

王麗蕉，臺灣大學圖書資訊博士，現任中央研究院臺灣史研究所研究副技師，兼檔案館主任。其研究領域為檔案學研究及數位典藏。

Li-chiao Wang, PHD in Library and Information Science, National Taiwan University, she serves as Associate Research Specialist and Director of Archives of Institute of Taiwan History at Academia Sinica. Her research interests focus on Archives study and Digital archives.

以旅券檔案數位加值資訊追尋日治時期臺灣人的跨境足跡

Tracing Overseas Footprints of Taiwanese with Digital Information from Passport Archives during Japanese Colonial Era

王麗蕉

Li-chiao Wang

中央研究院臺灣史研究所研究副技師兼檔案館主任

Associate Research Specialist & Director of Archives,
Institute of Taiwan History, Academia Sinica

摘要

日文的旅券，意指護照，即現今國際通用的 passport，是國民身處海外時的身份證明。日本於 1895 年依據馬關條約領有臺灣，臺灣住民在選擇去留兩年期滿後，未離臺者自然成為日本國民。故自 1897 年起，在臺住民出國需向地方官廳申請旅券。各官廳定期彙整旅券申請紀錄呈繳總督府，再由總督府彙送日本外務省，現今留存有 1897 年至 1944 年間旅券檔案，是記錄日治時期臺灣人海外移動的一手史料。旅券紀錄登載申請官廳、姓名、生日 / 年齡、本籍、旅行目的、旅行地點等項目，經由內容加值與標準化資訊建置，完成 20 萬餘筆數位資料集，可全面觀察臺灣人海外跨境活動的長期性變化與整體性圖像，深入分析旅行地點與旅行目的，亦可深化臺灣人藝術創作、海外留學與觀光旅遊等主題研究。

關鍵字：數位典藏、旅券檔案、人文資訊

Abstract

Using standardized and value-added information from 200,000 archived passport records, this study traced the cross-border and overseas travel activities of Taiwanese between 1897 and 1944. From 1897, those who chose to remain in Taiwan naturally became Japanese nationals and had to apply for passports (旅券) as a proof of identity when traveling overseas. Applications submitted to local government offices were first sent to the Government-General of Taiwan and then forwarded to the Ministry of Foreign Affairs of Japan. These archived passport documents from 1897 to 1944 detailed the government office handling applications as well as applicants' name, birthday/age, domicile, purpose and destinations of travel, thus preserving first-hand historical records of overseas travels of Taiwanese during the Japanese

colonial era. Analyzing these records would shed light on long-term changes and offer a comprehensive perspective of cross-border activities, which may inspire historical research on cross-border themes including artistic creations, overseas studies and tourism.

Keywords: Digital Archives, Passport Archives, Humanity Informatics

壹、前言

日文的「旅券」即一般所謂的護照 (Passport)。「臺灣總督府旅券下付及返納表」¹ 係日本領臺期間，在日本帝國的旅券管理制度下，由總督府轄下機關及地方官廳，依規定將所核發及回收之旅券登錄造冊而產生之官方檔案，為紀錄日治時期臺灣人海外移動的一手史料，亦為戰前日本帝國及其殖民地人民海外移動總體記錄之一環。

日本旅券制度之起源，與幕府末年由鎖國被迫走向開放的體制密切相關，是明治維新追求西化與近代化過程的一部份，1878 年日本外務省正式公布旅券制度，而因日清甲午戰爭，清朝戰敗簽署馬關條約將臺灣割讓日本，自 1895 年起臺灣島民被迫納入日本近代帝國體制，臺灣總督府於 1896 年頒佈〈清國人臺灣上陸條例〉，管理臺島港口之入出境，繼而於翌年 (1897) 發布〈外國行旅券規則〉，將日本的旅券制度施行於臺灣，爾後島民出入境皆需向總督府所屬官廳申請旅券，據此出境管制而所留存的旅券檔案，成為追尋日治時期臺灣人海外跨境足跡之關鍵史料。

「臺灣總督府旅券下付及返納表」，因其文書歸檔與呈繳制度，主要典藏於日本外務省外交史料館，另有少部分檔案留存於〈臺灣總督府公文類纂〉之中。有鑑於此宗旅券檔案的史料價值，中央研究院臺灣史研究所 (以下簡稱中研院臺史所) 自 2010 年起系統性地徵集與整理，隨著日本外交史料館旅券典藏微捲開放時程，歷經 10 年的海外徵集計畫，完成〈臺灣總督府旅券下付及返納表〉檔案全宗的蒐藏，² 再以 2 年的數位典藏與加值計畫，建置「臺灣總督府旅券系統」，收錄臺灣住民自 1897 年至 1944 年，約計 20 萬餘筆海外旅券申請紀錄，為研究者提供具完整性與結構化的數位內容資訊，提高獨特性史料在人文研究之應用性。

本文聚焦於旅券檔案內容中臺灣島民申請海外足跡與從事的跨境活動，首先闡述由旅券典藏轉向以人為中心的數位資訊建置工程，接著解釋運用資訊擷取與分析技術，進行旅行地點與旅行目的再加值之程序，藉以展示日治臺灣島民海外跨境的整體圖像，進而比較前往廈門與上海之跨境活動，期能拓展旅券檔案數位資訊多樣性應用與深化研究內涵。

1. 旅券下付及返納表，即為旅券核發及回收紀錄表。本文為行為方便，簡稱「旅券檔案」或「旅券紀錄」。

2. 有關〈臺灣總督府旅券下付及返納表〉檔案全宗之數位典藏經緯，參見拙著：〈臺灣總督府檔案中旅券紀錄及其數位人文加值〉，在《第十一屆臺灣總督府檔案學術研討會論文集》，南投：國史館臺灣文獻館，2021。頁 39-64。

貳、由旅券轉向以「人」為中心之數位資訊建置

「臺灣總督府旅券下付及返納表」是紀錄日治臺灣住民海外移動的一手史料，然而旅券檔案主要典藏於日本外交史料館，且該館多以微捲形式提供閱覽，對臺灣的研究者而言，使用上極為不便。因此，中研院臺史所自 2010 年起，以執行海外史料徵集與數位典藏計畫，配合日本外交史料館旅券檔案開放進程，歷經 10 年，完成所有旅券檔案的海外徵集工作；同時，深入瞭解旅券紀錄之呈繳與歸檔歷史脈絡，進而全面清查「臺灣總督府公文類纂」中旅券相關檔案，完成「臺灣總督府旅券下付及返納表」之檔案全宗典藏，涵蓋時間自 1897 年至 1944 年間，相當完整保存日治期間所有臺灣人出境申請的檔案紀錄。並自 2014 年起陸續開放，「臺灣史檔案資源系統」，迄今累積調閱量達 2,500 件、複印 3 萬多餘頁。時值新冠肺炎疫情爆發，尚於世界各國猛烈延燒之際，旅券紀錄的徵集回臺與開放應用，更顯其史料價值與典藏效益。

一、由旅券轉向以「人」為中心

為擴展與深化臺灣人跨境與移動的研究，將旅券檔案數位加值，從以「券」為中心轉向以「人」為中心。旅券檔案係日本外務省管理國民出入境之行政作為所衍生之產物，旅券登錄資訊隨旅券規則及其處理手續的修訂有所調整，初期各地方政府旅券申請紀錄的表格規範不一，至 1900 年 11 月發布〈外國旅行券規則取扱手續〉，首度明訂旅券登錄表之格式，翌年起旅券紀錄表格方為一致，依性質區分為下付表和返納表兩種，下付表內容依序為旅券號碼、姓名、族稱、本籍地、現住所、年齡、旅行目的、旅行地名、下付月日、備考；而返納表少了旅行目的，多了返納月日。1907 年 11 月起配合旅券規則重新頒佈，取消「族稱」欄，年齡改為「生年月日」。1921 年起實施「往復旅券制」，故增加「旅券種別」欄位。

更重要的是依當時旅券申請規則，同行家族得併記於同一張旅券，亦即同一旅券號碼之下。³ 如 1904 年 7 至 9 月臺南廳下付及返納表中，連雅堂時年 27 歲申請因商業目的前往廈門，妻子連肅雲 31 歲、長女連順治 7 歲、次女連招治 4 歲和長男連震東 1 歲，皆因未滿 12 歲而併記妻子名下旅券中，申請隨行前往廈門，參見圖 1。

3. 依據 1900 年頒佈之外國旅行券規則第三條，「與戶主同行之家族，或與丈夫同行之妻子，或與父母同行之子女，申請旅券時，得將其姓名、身分及年齡與戶主、丈夫或父母併記，但除與丈夫同行之妻子外，以未成年者為限。」復依據 1907 年頒佈之外國旅券規則第三條，「與戶主同行之家族、與丈夫同行之妻子，或與父母同行之子女，申請旅券下付時，得併記其姓名、身分及年齡於戶主、丈夫或父母之旅券，但除與丈夫同行之妻子外，以未滿 12 歲者為限。」（參見《府報》第 833 號，1900.10.06，頁 8；第 2304 號，1907.10.30，頁 75）

[illegible]

圖 1 1904 年臺南連雅堂及其妻小申請前往廈門的旅券下付紀錄。資料出處：〈1904 年 7-9 月外國旅行券下付表〉(T1011_030)，《臺灣總督府旅券下付及返納表》，中研院臺史所檔案館數位典藏。

又如在 1936 年 1 至 3 月臺中州彙報的旅券下付表中，林獻堂、林猶龍父子，和霧峰林家成員，以及清水楊肇嘉等人，申請前往上海、廈門、汕頭、廣東、天津和香港等地視察，由此下付表，亦可觀察表格上已取消族籍、年齡改為出生年月日、以及往復旅券等資訊欄位的變化。參見圖 2。

[illegible]

圖 2 1936 年臺中州林獻堂等人申請旅券下付之紀錄。資料出處：〈1936 年 1-3 月外國旅券下付表〉(T1011_191)，《臺灣總督府旅券下付及返納表》，中研院臺史所檔案館數位典藏。

經由前述旅券紀錄產生歷史背景與登錄規則變革分析，⁴ 旅券內容數位加值，為兼顧使用者需求和數位加值成本效益，建置資訊項目從一開始的申請年代、申請官廳、姓名、旅行目的、旅行地名等主要欄位，後續為提供辨別申請人資訊，陸續增加年齡、出生年月日、以及本籍等個人資料，以及單件識別號、影像號、資料出處等來源資訊。

二、旅券資訊建置與著錄規範修調

由旅券紀錄確定轉向以「人」為中心和定義加值資訊欄位後，再隨數位加值進展，修調各欄位判讀與著錄規範，基於尊重史料來源及其原有脈絡，而日治時期臺灣旅券紀錄係以日文為官方書寫語文、且為手寫留存的手稿紀錄，故因各人書寫習慣與字體不一，時有辨識或解讀之困難，亦有抄錄時錯漏字、漏抄再補寫、異體字、書寫簡稱，以及旅行地名與現今用語不同等問題，為避免主觀判斷困難或不一，以照錄史料紀錄為原則，唯當時登錄者多以自身所習慣的日本漢字、異體字或簡字書寫，為利現今讀者查詢，故以正體字建檔為原則，如前述圖 2 旅券紀錄中書寫為「林獻堂」，姓名欄位中即以正體字「林獻堂」著錄之。簡言之，旅券各欄位資訊加值，以「史料紀錄照錄、正體字建檔」為基本原則。

再根據其資訊特性逐一訂定各欄位著錄規範，首先，「申請官廳」一欄，係旅券紀錄之作成及彙整呈繳單位，包含臺灣總督府所屬單位及地方官廳。日治時期臺灣地方行政區劃幾經改變，由六縣三廳縮編為三縣三廳，再歷經擴展二十廳和縮減為十二廳，最後改制五州二廳和增為五州三廳，且同一地方官廳名稱也有多種。因此，申請官廳之欄位著錄規範如下：

- (一) 主要著錄依據來源以呈繳公文或下付表格為主，但不著錄「縣」、「廳」、「州」等行政區劃層級文字。
- (二) 同一地方官廳不同名稱，如桃仔園後改桃園、阿猴後改阿緱，則以改名後的寫法分別著錄為「桃園」、「阿緱」。
- (三) 臺灣總督府轄下各單位，如外事課、民政部、總督官房外事課、總督官房文書課、總督官房外務部、外事部，一律標準化著錄為「臺灣總督府」。
- (四) 日本駐中國公使館或領事館，如在中華民國日本公使館、在上海日本總領事館、在廈門日本領事館等，亦一律標準化著錄為「日本駐中國使館」。

在「本籍」欄位著錄規範，旅券紀錄中本籍書寫詳、簡不一，原評估加值成本與專業判

4. 日本外務省自 1912 年起，簡化返納表之繳交程序及其欄位，僅保留旅券號碼、姓名、旅行地、下付年月日、返納月日等 5 個欄位，因返納表留存數量少，且重要性未及下付表，故旅券建置欄位考量以下付表為主。參見〈外國旅券規則取扱手續改正〉，《府報》第 3506 號，1912 年 5 月 1 日，頁 1。

讀人力，並未著錄，後考量此資訊為判別申請人國族身份，甚至臺灣各地方人士跨境與流動研究的重要資訊，故再新增列入加值項目。為利後續分析應用，著錄當時行政區劃第一層，⁵並比照前述「申請官廳」一欄，取消行政縣、州、廳之別，促使本籍分類架構及後續分析呈現更為簡明清晰。

接著，在「旅行地點」與「旅行目的」兩欄位資訊，為分析臺灣島民海外跨境最重要資訊，依當時規定需詳實填寫出國地點與目的，故會有一個以上的地名與目的，但又未有一致性書寫規範，因此書寫方式相當多元。在「旅行地點」一欄，如南洋中最多臺人前往的爪哇，即有蘭領爪哇、蘭領東印度爪哇、東印度爪哇、蘭領ジャバ、和蘭領爪哇、蘭領爪哇島、蘭印爪哇、爪哇島、ジャバ、蘭領ジャワ、蘭領ジャワ、和蘭爪哇、ジャ、ジャワ、蘭領ジャバ、爪哇、和蘭國瓜亞地、蘭領印度爪哇、和蘭領爪哇、ジャバ、ジャハ、蘭領爪哇、蘭領ジャバ島、蘭領東印度ジャバ、和蘭瓜亞地、蘭領ジャハ、東印度群島爪哇、爪哇諸島、爪哇國、蘭領東印度ジャバ、蘭領東印度ジャワ等 30 餘種寫法。此外，旅行地名常有書寫單字簡稱之現象，如西班牙簡寫成「西」、澳門的「澳」、土耳其（今土耳其）的「土」等單字地名，在確認簡寫地名所指地點之前提下，將單字地名調整全稱或雙字詞著錄，如「西」著錄為「西班牙」，「米」著錄為「米國」等。⁶換言之，在資料數位加值過程中，雖依循「照錄史料」之基本原則，惟仍需考量使用者需求與系統程式運算，因「地」制宜，即時修調著錄規則，實有其必要。

而在「旅行目的」一欄中，同樣因書寫者習慣用字，敘述繁、簡差異，同一目的往往呈現多樣表述，如商業，即有商用、商取引、商業ノ為、商、商業經營、商業ノ為メ、商用ノ為、營業所二歸ル、商品取引、商用ノタメ、商品販賣、商業經營ノ為メ、商務打合、商品仕入ノ為メ、營業、商打合、貿易、商品賣買、商取引ノ為、物品賣買、商取引ノ為メ、貿易商用、商品賣込、商業取引ノ為メ、貿易商經營ノ為、商業二從事、商業取引ノタメ、商業從事、商品仕入ノ為、商品買入ノ為メ、商業上ノ為メ、商業經營ノ為、商談、商業幫助、商品仕入ノタメ、商品買入ノタメ、商業打合、商店整理、商取引決算、商業上ノタメ等，更有多達 100 餘種相似的寫法。然為避免主觀判斷和建檔者專業能力考量，旅行目的一欄，以照錄史料為原則，包括以片假名的書寫紀錄，不作例外規範。

5. 鑑於澎湖廳在地方改制過程中，僅自 1921 年至 1925 年短暫時間被併入高雄州降級為澎湖郡，且原有行政區劃未更動，故而此段期間本籍屬「高雄州澎湖郡」之旅券，本籍欄著錄為澎湖，而非高雄，惟申請官廳欄則著錄為高雄。

6. 旅行地名單字約計有 30 種，係在資訊加值與系統功能分析過程中陸續發現與討論。肇因於歐美和南洋等地名，常有不同漢字或片假名譯名，系統中建置與提供地名對照表，有利於使用者快速查全所需資料，但若西班牙對照中有「西」，反而會檢索出廣西、巴西、西伯利亞等地名中有「西」的資料。此外，若無法確認單字地名者，則不做調整，如「瑞」，因無法確認為瑞士或瑞典，仍照錄史料，不予更動。

綜合上述，臺灣總督府旅券檔案自海外徵集入藏、編排描述、分析解讀、到數位化加值，經歷繁瑣的檔案作業程序，再從旅券轉向以旅人為中心，終於完成自 1987 年至 1944 年間臺灣島民海外旅券紀錄數位內容建置。「臺灣總督府旅券下付及返納表」，雖與「臺灣總督府職員錄」同屬日治時期人物相關檔案紀錄，然職員錄係為正式發行之印刷物，⁷ 而旅券則為手寫紀錄，由不同人抄錄謄寫；又因呈繳和歸檔規定，旅券檔案現今主要典藏於日本外交史料館，然臺灣總督府公文類纂中亦留存部分年代版本，因此，從原始檔案紀錄海外徵集與清查比對，旅券紀錄判讀與建檔、標準化規範修調與校正，遠較職員錄更具挑戰性。

參、旅券數位資訊再「加值」與日治臺灣島民海外跨境圖像

旅券紀錄數位資訊建置成果，主要涵蓋臺灣總督府彙報自 1897 年至 1942 年間各官廳呈繳之旅券下付及返納表，以及日本駐中國領事館彙整 1926 年至 1944 年間旅券申請紀錄，相當完整呈現日治時期臺灣島民海外跨境申請紀錄。茲就此 20.8 萬餘筆旅券申請紀錄之數位資料集，分別從申請官廳、旅行地點、以及旅行目的，分析日治時期臺灣島民海外跨境申請紀錄實態，概述如下。

一、申請官廳旅券紀錄整體分析

日治時期臺灣地方區劃更動頻繁，地方申請官廳名稱多達 34 種；此外，總督府及轄下行政單位有 7 種官廳名稱，日本駐中國使館亦有 18 種使館名稱。為利數位資訊分析，如同前述在申請官廳資訊建置過程，加以標準化規範著錄規則，申請官廳由近 60 種精簡為 22 類。

分析各申請官廳旅券申請數量，最多是臺北，自 1897 至 1942 年間，因歷經縣、廳、州各時間，因此多達 11 萬多次，超過總旅券申請紀錄的一半 (55%)；再者依序是臺南、臺中、高雄、新竹，介於 1 萬至 3 萬人次之間，亦即後來劃分為五州的地方官廳；接續是澎湖有 5 千多人，自 1897 至 1942 年間，期間僅自 1921 年至 1925 年被劃入高雄州。而臺灣總督府主要是辦理因公務出國的旅券紀錄，自 1899 至 1942 年間，不到 3 千人次；日本駐中國使館，自 1926 至 1944 年間，亦只有近 4 千人次，其比例皆不到總申請數量的 2%。各申請官廳旅券下付及返納紀錄數量，詳見表 1。

7. 關於《臺灣總督府及其所屬官署職員錄》之發行沿革與臺史所檔案館「臺灣總督府職員錄系統」之數位加值應用，可參見拙著〈臺灣總督府職員錄系統在人文研究之應用：以日治臺灣初等學校教師及其跨境為中心〉，《臺灣史研究》27：3 (2020.12)，頁 167-212。

表 1 申請官廳旅券紀錄統計 (自 1897 年至 1944 年)

申請官廳	旅券申請紀錄年代	總數	臺籍	日籍
臺北縣 / 廳 / 州	1897-1942	113,881	102,251	8,128
宜蘭廳	1897-1920	1,697	1,622	54
基隆廳	1901-1909	2,418	2,009	380
深坑廳	1901-1909	266	254	12
新竹縣 / 廳 / 州	1897-1898, 1901-1942	10,634	10,346	108
桃園廳	1901-1920	1,504	1,426	27
苗栗廳	1901-1909	756	730	16
臺中縣 / 廳 / 州	1897-1942	16,305	15,681	339
南投廳	1901-1907, 1910-1913, 1915-1920	70	52	15
彰化廳	1901-1909	2,154	2,072	81
嘉義縣 / 廳	1897-1898, 1901-1920	1,869	1,732	137
臺南縣 / 廳 / 州	1897-1942	29,114	27,469	943
斗六廳	1901-1909	389	386	3
鹽水港廳	1902-1909	199	181	18
蕃薯藔廳	1901-1909	52	46	6
高雄州	1920-1942	11,797	11,167	626
鳳山縣 / 廳	1897-1898, 1901-1909	1,064	1,017	47
阿猴廳	1901-1920	1,162	1,088	33
恆春廳	1901-1909	259	253	4
臺東廳	1897-1926, 1938, 1942	611	568	35
花蓮港廳	1910-1939	601	547	37
澎湖廳	1897-1920, 1926-1940	5,067	4,898	102
臺灣總督府	1899-1942	2,712	沒有本籍資訊	
日本駐中國使館	1926-1944	3,940	本籍未列入分析	

再以臺灣各地方官廳旅券申請人的本籍分析，多數地方官廳旅券申請有超過九成是臺灣人；而各地日籍人士的旅券申請平均不到一成，最多是在臺北地方官廳，有 8 千多人，其餘各地方官廳皆不到 1 千人，其中較特別是基隆廳，自 1901 至 1909 年間，就有 380 人，約

是當地申請量的 16%，此或與日人在臺短期停留或來臺過境，就近居住於港口有關。參見表 1。

二、旅行地名分類與跨境空間分區展示

「旅行地名」經由資訊擷取分析，因可申請多地跨境，故在 20 萬餘人次旅券紀錄中，旅行地名資訊高達 30 萬餘筆，有 1,600 多種書寫地名，若未將地名資訊再標準化的分類歸納，難以跳脫史料零雜之侷限，為鳥瞰綜觀日治臺人跨境整體圖像，在數位資訊加值再新增「地名分類」一欄，先區劃為「東亞」、「南洋」、「歐美非」三大區域，再綜合考量日治臺灣所處時空背景，與各地旅券申請數量，⁸ 加以次分類。茲就各區分類規則及其跨境分析，說明如下。

（一）東亞

日治期間臺灣島民或因原鄉與地緣關係，申請旅券數量最多是前往廈門、福州等沿海港口城市，因此，此區的分類兼採「熱點」城市與「區塊」省份或國家，細分為 32 類。參見表 2。綜觀東亞各地旅行申請人次，依序是廈門、福州、香港、上海、汕頭等五大城市，皆超過 1 萬人；另，前往由日本扶植的滿州國有 2 千多人，另一殖民地韓國僅不到 2 百人，申請時間集中於 1905 年成為日本保護國之前。

表 2 東亞各地旅行地點分類與數量統計（超過百人次）

排序	地名分類 ⁹	旅行人次	分類說明
1.	廈門	154,809	
2.	福州	36,318	
3.	香港	26,373	
4.	上海	22,724	
5.	汕頭	18,561	
6.	廣東	7,936	不含汕頭
7.	滿州國	2,265	1932 年成立，新京（奉天）、長春歸入滿州國。1931 年以前赴該地者，則歸類為東北三省。

8. 旅券地名分類，採計臺灣總督府及所屬官廳彙報旅券下付及返納表，和日本駐中國使館彙整旅券下付中臺灣人，合計為 20.5 萬餘筆，各地數量統計會隨資料勘誤，而略有些變動差異。

9. 尚有湖南、關東州、江西、雲南、廣西、安徽、山西、日本、四川、蒙古、陝西等 11 類，數量介於 1~120 之間。

排序	地名分類 ^{9.}	旅行人次	分類說明
8.	泉州	2,053	不含廈門
9.	天津	1,782	
10.	南京	973	
11.	清 / 中華民國	863	書寫為清或中國，無法確認所在省分者，歸入此類。
12.	浙江	773	
13.	福建	642	不含廈門、福州、泉州
14.	河北	611	不含天津
15.	澳門	387	
16.	江蘇	324	不含上海、南京
17.	山東	268	
18.	湖北	231	
19.	東北三省	185	1931 年前歸此類，1932 年後歸入滿州國
20.	韓國	182	1905 年日韓簽訂《乙巳條約》，韓國成為日本的保護國後，就不須再申請旅券。
21.	海南島	116	

(二) 南洋

第二大區域—南洋，因西方帝國擴張和殖民政權更迭，此大區次分類亦兼「熱點」城市與「區塊」政權，如新嘉坡等，單獨作為一類，而旅行人次較少之地點，則歸入所屬政權區劃，如「孟買」、「コロンボ」（科倫坡）歸入英領印度等，而澳洲和紐西蘭較接近當時所謂的「裡南洋」、「表南洋」，故歸入此區域，共計細分為 12 地，參見表 3。概觀日治時期下南洋各地旅券申請人次，最多是爪哇，不到 4 千人；接著是比律賓、新嘉坡、和婆羅洲等三地區，皆為 2 千多人；而暹羅和印度支那皆為 1 千多人，其餘各地區則不到千人，澳、紐僅 40 多人。

表 3 南洋各地旅行地點分類與數量統計

排序	地名分類	旅行人次	分類說明
1.	爪哇	3,944	
2.	比律賓	2,310	今通稱之菲律賓、

排序	地名分類	旅行人次	分類說明
3.	新嘉坡	2,294	今通稱之新加坡
4.	婆羅洲	2,292	分屬英、荷兩國，數量較多，單獨一類
5.	暹羅	1,292	今通稱之泰國
6.	印度支那	1,095	今越南、柬埔寨、寮國
7.	蘭領東印度	811	不含爪哇、婆羅洲兩島
8.	英領馬來亞	614	不含新嘉坡
9.	英領印度	193	
10.	緬甸	174	
11.	濠洲、新西蘭	45	今澳洲、紐西蘭
12.	南洋	20	書寫為南支南洋者，歸入此類。

(三) 歐美非洲

第三大區域—歐、美、非洲，前往旅行者大多數是申請多地跨境，且旅行人次較少，故以「國家」為次分類原則，如美、英等國，未明確書寫者或數量極少者，則以區域劃分，如書寫「歐米」者亦歸入歐洲，埃及歸入非洲等，共細分為 19 類，參見表 4。申請前往歐美非各國相較是非常少數，合計僅 1 千多人次，最多是前往美國，有 1 千人左右；接著英、德、法等三國，超過百人、但不到 2 百；其餘各國皆不到百人。

表 4 歐美非各地旅行地點分類與數量統計

排序	地名分類 ^{10.}	數量	分類說明
1.	米國	1,006	美國；書寫為「歐米」者，亦計入。
2.	英國	182	
3.	獨逸	155	德國
4.	佛國	143	法國
5.	伊太利	89	義大利
6.	加奈陀	72	加拿大
7.	瑞西	62	瑞士

10. 歐美非的旅行地點分類，尚有白耳義（比利時）、奧國（奧地利）、非洲、北歐、南美、中歐（不含奧地利）、歐洲（書寫為「歐米」者，亦計入）、南歐（不含義大利）、中東等 9 類，旅行人次更未及 50 人。

排序	地名分類 ^{10.}	數量	分類說明
8.	和蘭	55	荷蘭
9.	露國	53	俄國
10.	中美西印度群島	51	

綜合上述旅行地點分類與跨境分析結果，旅行地點分類之「再加值」，先考量地理空間分佈，區劃為東亞、南洋、歐美非等三大區域；再兼顧時空背景與旅券數量細分各區域次分類，東亞之下再細分為 32 類，南洋地區細分為 12 類，歐美非細分為 19 類。藉此，全面統計分析三大區域旅券申請數量，最多申請前往是東亞，僅廈門一地就有 15 萬人次，接著是福州、香港、上海、汕頭等沿海港市，介於 1.8 萬至 3.6 萬之間；由日本扶植成立的滿州國則有 2 千多人。第二多是南洋，最多是前往爪哇，有近 4 千人次；比律賓、新嘉坡、婆羅洲亦是 2 千多人。最少是前往歐美與非洲，主要前往美國，但僅有 1 千人左右，依序是英、德、法等三個歐洲國家，亦都不到 2 百人次。由此，旅行地點分區與分類，可全面鳥瞰日治臺灣島民海外跨境之整體圖像，再由旅券申請前往廈門一地，就超過 75% 的數據，亦可再次驗證日治期間臺人往返兩岸的頻繁關係。

三、旅行目的分類與海外跨境活動概覽

旅券紀錄中「旅行目的」經資訊擷取與分析，有近 21 萬筆，雖不如旅行地點數量，卻有多達 1.7 萬種書寫值，唯其中 1.1 萬種僅出現 1 次；如同前述舉例，因商業出國，其目的書寫即有百餘種之多。為維護史料原有歷史脈絡，於資訊著錄時以照錄史料為原則，然如同旅行地點再分類，在保存史料原貌之同時，透過數位資訊「再加值」方能從隱藏於浩瀚資料的關連資訊，推展更具宏觀性分析與全面性觀察。

基於此故，再新增「目的分類」一欄。主要以行業類別，並兼顧旅券紀錄中所登載目的之數量與特色。首先是商業活動，蓋清領以降臺灣即與中國沿海有繁盛的貿易往來，自臺灣開港後，臺灣雖由兩岸貿易擴大至世界貿易，惟與中國之間的區域分工，迄日本領臺後，不少臺灣商家為節省關稅而透過日本商號進行轉運，形成臺、日、中三地的國際貿易。從旅券紀錄可見商號業主為採購藥材、雜貨，或視察臺灣商品之販路，或在中國開店經商，而多次往返兩岸。商業活動之外，再續分為農林業、工礦製造業、航運漁業、醫藥業、公務等行業與職業類別。

其次是家族親友活動，由於福建、廣東乃多數島民之原鄉，自清領時即遠渡重洋來臺拓墾者，有不少父母及親友仍留在家鄉。總督府外事課於 1912 年發現當年度旅券申請張數下降，推測與「本島人之父母（中國人）漸次死亡，為省親而渡清之必要減少」有關。¹¹ 因此，將探親、探病、結婚、與家人同行等歸類為家族親友活動，另將掃墓、修墳、喪禮、祭祀等，歸為喪葬掃墓一類。

最後綜合考量旅券數量和特定研究主題，再分為宗教活動，教育研究、就職受雇、觀光旅遊、演藝娛樂、法律訴訟等，合計為 14 類，參見下表 5。

表 5 旅行目的分類定義與數量統計

序號	目的類別	定義	數量 (約計)	比例
1.	商業活動	凡與商業相關之活動均屬之，包括金融業、服務業（如零售、理髮業等）及食品製造業（如製菓、醬油等）從業及受雇者，皆歸於此類。	84,000	41%
2.	家族親友活動	凡與家族內部或親友之間互動等均屬之，如探病、結婚、親友探訪、會面、返鄉、與家人同行等。	50,000	24%
3.	宗教活動	從事或參與佛、道、神道、基督教、天主教等信仰活動，如進香，或進行宗教相關研究者，均屬之。	11,000	5%
4.	就職受雇	為特定組織、人士雇用效力者。但受雇商業、農林、工礦製造、航運漁業、醫藥、演藝娛樂等工作，則分別歸入各類別之下。	7,900	4%
5.	喪葬掃墓	參與或籌辦喪葬祭儀活動，包括喪禮、祖先祭祀、祖祠參拜、掃墓、修墳、洗骨等。	7,100	4%
6.	工礦製造業	利用各種原料所從事的製造業及相關事業，如製糖、手工業、營造、採礦等。製冰等製造業，歸屬此類。	4,900	2%
7.	教育研究	為學習或教授知識、升學考試、或從事專業研究者，如醫學研究、繪畫研究、學事視察、和留學等。	4,200	2%
8.	農林業	第一級產業中的農業、林業及畜牧業，包括以利用和生產農林產品，如製材、製茶、精米等。	3,300	2%
9.	航運漁業	從事船舶運送、捕魚、養魚業者等，如船員、漁夫、漁業視察、魚種買賣、船隻救助等。	2,900	1%

11. 臺灣總督府，《臺灣總督府事務成績提要 第十八編上》（臺北：臺灣總督府，大正元年度分）頁 64。

序號	目的類別	定義	數量 (約計)	比例
10.	醫藥業	從事中西醫療、公衛、獸醫、藥業相關工作，包含醫院藥局經營、衛生醫療等視察活動。	2,400	1%
11.	公務	以官吏身份前往奉職，或持有總督府官房文書課發給之公務旅券，為執行公務而出境者。	2,400	1%
12.	觀光旅遊	以遊覽觀光為旅行目的者，如漫遊、觀光、和遊學等歸入此類。	2,400	1%
13.	演藝娛樂	從事演戲、歌舞、游藝、表演、運動競技等活動以提供顧客娛樂及休閒服務者屬之。	890	0.4%
14.	法律訴訟	擔任辯護士（律師）及其受雇者，或參與出席法律訟案者屬之。	380	0.2%

由前述旅行目的分類再加值之統計結果，可發現日治時期臺灣島民申請海外旅行目的，因商業出國最多，超過 8 萬人次，其次是家族親友活動有 5 萬多人，第三是宗教活動亦有 1 萬多人；第四、五是就職受雇、喪葬掃墓，分列有 7 千多人；其後各類至觀光旅遊則介於 2 千至 5 千人次，最後兩類是較為獨特跨境活動，演藝娛樂有 8 百多人、法律訴訟有 3 百多人。

四、旅行地點與目的交叉分析：以跨境廈門、上海為例

經由前述申請官廳、旅行地點與旅行目的等再加值資訊之整體性分析，本研究進一步比較旅行地點及其跨境活動的差異，選定由臺灣各官廳核發之旅券，合計為 179,408 人次，以廈門和上海兩地進行比較，前往廈門有 133,731 人，近 75%，至上海有 21,900 人，約有 12%，參見表 6。再觀察歷年旅券數量變化，往廈門者，第一波高峰自 1898 年至 1910 年，其次是自 1931 年至 1935 年間，皆超過八成。反觀至上海者，起初未及一成，直自 1921 年至 1931 年間才漸次增加超過二成，而在 1932 年遽降至一成以下，最大變化是在 1938 年，增加至 36%，相較至廈門僅 31%，1932 年與 1938 年皆是日本侵華之際，箇中緣由，值得深究。

表 6 臺灣總督府自 1897 至 1942 年旅券下付數量統計

登載 年代	旅券 數量	前往 廈門	所佔 比例	前往 上海	所佔 比例	登載 年代	旅券 數量	前往 廈門	所佔 比例	前往 上海	所佔 比例
1897	2,634	1,882	71%	31	1%	1920	2,677	1,640	61%	193	7%
1898	2,912	2,458	84%	62	2%	1921	4,358	3,007	69%	609	14%
1899	2,967	2,510	85%	83	3%	1922	3,763	2,514	67%	694	18%
1900	2,828	2,466	87%	113	4%	1923	3,357	2,247	67%	731	22%
1901	3,179	2,589	81%	117	4%	1924	3,218	2,011	62%	774	24%
1902	3,895	3,158	81%	108	3%	1925	3,268	2,311	71%	702	21%
1903	3,824	3,134	82%	133	3%	1926	4,069	2,909	71%	957	24%
1904	3,057	2,638	86%	79	3%	1927	4,283	3,110	73%	896	21%
1905	3,659	3,180	87%	115	3%	1928	4,764	3,384	71%	1,174	25%
1906	4,221	3,489	83%	121	3%	1929	6,043	4,495	74%	1,458	24%
1907	3,724	3,075	83%	125	3%	1930	7,423	5,626	76%	1,703	23%
1908	2,883	2,541	88%	68	2%	1931	8,578	6,872	80%	1,954	23%
1909	2,750	2,403	87%	64	2%	1932	6,077	4,921	81%	576	9%
1910	2,305	1,892	82%	64	3%	1933	9,399	7,877	84%	894	10%
1911	2,365	1,832	77%	99	4%	1934	8,923	7,130	80%	1,038	12%
1912	2,261	1,673	74%	117	5%	1935	9,295	7,422	80%	1,165	13%
1913	2,116	1,531	72%	122	6%	1936	7,469	5,562	74%	1,423	19%
1914	2,500	1,861	74%	118	5%	1937	4,171	2,502	60%	785	19%
1915	3,238	2,595	80%	138	4%	1938	3,899	1,226	31%	1,412	36%
1916	3,931	2,852	73%	190	5%	1939	671	50	7%	31	5%
1917	4,513	2,737	61%	201	4%	1940	545	33	6%	8	1%
1918	3,805	2,470	65%	222	6%	1941	192	2	1%	2	1%
1919	3,163	1,914	61%	231	7%	1942	236	0	0%	0	0%

進而再以旅行目的比較前往廈門與上海跨境活動的差異，發現排名前二者同樣皆為商業、家族親友活動；然而，前往廈門者，接續是宗教活動、喪葬掃墓，與至上海從事教育研究、觀光旅遊等活動有所不同。參見下表 7。

表 7 前往廈門與上海兩地之旅行目的 (自 1897 至 1941 年)

排名	前往廈門			前往上海		
	旅行目的類別	人次	比例	旅行目的類別	人次	比例
1.	商業活動	63,688	48%	商業活動	14,342	65%
2.	家族親友活動	37,763	28%	家族親友活動	2,920	13%
3.	宗教活動	9,658	7%	教育研究	1,020	5%
4.	喪葬掃墓	5,194	4%	觀光旅遊	923	4%
5.	就職受雇	4,425	3%	就職受雇	903	4%

由上述分析，商業活動一直是跨境最主要的目的，因商業前往廈門和上海的旅券數量各佔近 48% 和超過 65%，並由臺史所檔案館館藏〈長崎泰益號文書〉¹² 收存與臺灣、上海、福建等地 1 千多家商號之往來書信，可見一斑。再如前述提及福建是多數臺灣人的原鄉，應為申請前往廈門參與家族親友和喪葬掃墓的主因；以及移民渡海至臺灣開墾時，亦將原鄉神祇信仰帶入臺灣，即透過「分身」與「分香」兩種形式，在臺灣各地建立分廟。各分廟為與祖廟確立並維繫此一流關係，需每隔一段時間前往祖廟乞火，參加祖廟祭典，即俗稱之「進香」，¹³ 由前往廈門參與宗教活動有 9 千多人之數據，顯示臺灣與廈門兩地宗教信仰的連結關係。

反觀前往上海者，因教育研究與觀光旅遊申請前往的比例較高。如臺灣著名前輩畫家，於 1929 年受聘擔任上海新華藝術大學西洋畫教授，¹⁴ 在其 1930 年旅券申請紀錄中同仁還有妻小 3 女 1 男，陳澄波在上海任教期間，亦創作不少的代表作，如〈清流〉、〈綢坊之午後〉、〈我的家庭〉等油畫藝術作品。

觀光旅遊，反應臺灣經濟與休閒風氣轉變，中國因地緣之便、語言相通，吸引臺人前往，故而，因觀光旅遊前往上海者亦為數不少。如新竹州黃旺成於 1930 年 4 月申請前往上海、南京等地觀光旅券，並在其日記中自 5 月 6 日起逐日記載遊歷福州、上海、南京、青島、北京等地，長達一個多月的旅行見聞；¹⁵ 復將遊記整理，題為〈新中國一瞥的印象〉，自 1930 年 6 月起連載於《臺灣新民報》¹⁶ 共 15 回，以自身觀點向臺人介紹新中國之面貌。

12. 參見《長崎泰益號文書》，「臺灣史檔案資源系統」，<https://tais.ith.sinica.edu.tw>。

13. 據 1987 年統計，全臺民間信仰之神靈共 300 多種，其中 80% 是由中國（主要是福建）分靈而來。（《臺灣大百科全書》，分靈，2021.05.24 閱覽，<https://nrch.culture.tw/>）

14. 參見《陳澄波畫作與文書》，「臺灣史檔案資源系統」，<https://tais.ith.sinica.edu.tw>。

15. 黃旺成著，許雪姬編註，《黃旺成先生日記（17）：一九三〇年》（臺北：中央研究院臺灣史研究所，2017）。亦可連線「臺灣日記知識庫」，<https://taco.ith.sinica.edu.tw/tdk>。

16. 參見《六然居典藏史料》，「臺灣史檔案資源系統」，<https://tais.ith.sinica.edu.tw>。

肆、結論

臺灣總督府旅券檔案是一具有連續性、獨特性之一手史料，對於臺灣史研究及家族史之建構，具有不可忽視的重要性，完整的海外徵集與數位典藏，以及數位資訊建置，不僅可提昇檔案典藏查詢的便利性，以及檔案利用的可親近性，更可進一步運用資料探勘與視覺化等資訊工具，支援現今學界所關注之人流與跨境相關研究主題之深化，亦可結合臺史所檔案館原有數位典藏成果，如〈長崎泰益號文書〉中與各地商號貿易書信，探討臺灣人的海外商業活動和東亞貿易網絡發展史；或連結林獻堂、黃旺成等私人日記中的海外遊記，追溯臺灣人海外足跡與世界觀的形塑。

本文以旅券檔案之數位資訊加值為討論中心，闡述旅券紀錄數位資訊建置和史料著錄原則，所完成之數位資料集，不僅將史料本身從以「券」為中心的國家治理性材料，轉化為以「人」為中心，成為更貼近人文學研究脈絡之素材；經由標準化著錄規範與正規化資訊處理，既維護史料原始脈絡，又消除了文字書寫之訛誤或歧異性，促使數位資訊便於快速檢索及精確擷取，提升史料的研究價值。

並以旅行地名和旅行目的之數位資訊再「加值」，闡釋日治時期臺灣島民海外跨境活動圖像，一方面說明大量的旅券數位內容資訊，有助於拓展宏觀性結構分析的深度與廣度，並為微觀式個案提供整體性之參照數據。整體觀之，自 1897 至 1944 年間，申請旅券 20 萬餘筆資訊，跨境地點最多是前往廈門有 15 萬多人，即平均每 4 人中有 3 人申請至廈門，以及集中於福州、香港、上海、汕頭等沿海城市；其次是下南洋，有 1 萬多人；最少是歐美非，僅 1 千多人，其中臺灣人更是極少數。以跨境目的歸類分析，以商業 8 萬多人最多，家族親友活動 5 萬多人居次，三是宗教活動有 1 萬多人。進而以廈門和上海兩地比較跨境活動之差異性，和探討獨特性個案海外跨境之生命經驗，連結臺史所「臺灣史檔案資源系統」、「臺灣日記知識庫」等數位典藏系統，不僅可提高數位資訊在人文研究之應用性，亦可拓展旅券研究多樣性和深化研究內涵。經由旅券檔案與書信、遊記、報紙、圖像等典藏串連，更使旅券紀錄中的旅程具像如實展開，彰顯旅券檔案看似制式化的申請記錄文字，經過多樣性史料的延伸探索，往往可發現其背後承載著許多豐富而精彩的人生故事，值得細細探索與追尋。

誌謝

1. 臺灣總督府旅券系統建置，係獲得中央研究院數位人文研究計畫經費補助 2019 年~2020 年「臺灣檔案文獻數位典藏與加值應用計畫—以日治時期旅券為核心」（計畫編號 AS-ASDC-108-105、AS-ASDC-109-104）。
2. 本文得以順利完成，承蒙中研院臺史所鍾淑敏研究員費心審閱，和惠賜諸多寶貴修正建議；中研院臺史所檔案館謝明如與林佑芯等同仁，協助臺灣總督府旅券相關資料蒐集和數位內容加值與分析，使本文研究資料更為周延，特此謹致謝忱。

【子題二】

圖書館數位人文研究平台建置與應用

Session II

**Creating and Using Digital Humanities Research
Platforms in Libraries**



刘炜
Wei Liu

刘炜，上海图书馆上海科学技术情报研究所副馆（所）长，研究员，图书情报学硕士，计算机软件与理论博士。复旦大学、华东师范大学和上海大学兼职教授，上海大学博士生导师。兼任上海市图书馆学会副理事长，DC 元数据管理委员会委员，《中国图书馆学报》等专业期刊编委。

Wei Liu is the Deputy Director of Shanghai Library. He is an adjunct professor of Shanghai University, as a doctoral tutor and give lectures on Digital Libraries and Digital Humanities. He also serves as the inaugural director of the China Digital Humanities Alliance (under the Chinese Society of Indexing).



王丽华
Lihua Wang

王丽华，上海大学图书情报档案系副主任，副教授，图书馆学博士，上海大学数字人文研究与发展中心主任，《数字人文研究》编委。

Lihua Wang, Doctor of Library Science, is deputy dean and associate professor at Department of Library, Information and Archives of Shanghai University, and the director of Digital Humanities Research and Development Center. She is also the editorial board member of Digital Humanities Research.

汉学数字人文通用平台的需求与设计

The Functional Requirement and Designing of a Universal Digital Humanities Platform for Sinology

刘炜

Wei Liu

上海图书馆副馆长

Deputy Director, Shanghai Library

王丽华

Lihua Wang

上海大学图书情报档案系

Department of Library, Information and Archives, Shanghai University

摘要

随着人文社会科学向数据驱动型研究范式的转型，图书信息机构提供数字人文服务已成趋势，很多数字典藏系统都在积极应用关联数据、知识图谱、实体识别、机器学习、数据可视化等数据技术，升级为数字人文服务平台。文章讨论了汉学数字人文平台建设的功能需求和设计趋势，从系统、资源、功能、工具与用户五个方面探讨了分布式环境下全球汉学数据典藏机构进行平台共建共享的系统要求、资源规范、功能特征、工具支持与用户界面等整体需求，对数字人文平台为支撑数字人文研究而提供全面的资源获取能力、个人研究环境和数据操控工具等，提出了较为系统完整的设计思考，并以上海图书馆正在开发中的“历史人文大数据平台”为案例阐述这些思考成果的应用。

关键词：汉学研究、数位人文、数据资源、平台设计、需求分析

Abstract

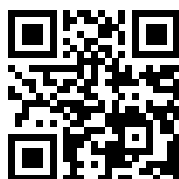
With the transformation of the humanities and social sciences into a data-driven research paradigm, it has become a trend for libraries and other memory institutions to provide digital humanities services. Many practitioners are actively applying data technologies such as Linked Data, Knowledge Graph, Name Entity Recognition, Machine Learning, and Data Visualization

to upgrade digital collection systems to Digital Humanities Service Platforms. This paper discusses the functional requirements and design thinking of a Digital Humanities Platform for Chinese Studies. It also discusses the data and metadata formats, protocols for data exchange, and possible mechanisms for co-construction and sharing of platforms by global sinology repositories in a distributed environment from the aspects of system, resources, function, methods and users. At the same time, the paper proposes a systematic thoughts on how to adapt the digital humanities platform to meet the needs of digital scholarship and provide comprehensive resource access, personal research environment, and data manipulation tools to researchers. The results of these reflections are being applied to the construction of the "History and Humanities Data Platform" currently developed by Shanghai Library.

Keywords: Sinology Study, Digital Humanities, Data Resources, Platform Design, Requirements Analysis

預錄及簡報檔

<https://pse.is/3e37pp>



壹、引言

人文学科是所有科学之肇始，是人文精神之依托，是被称为知识分子的必备和基础素养。无论是古希腊的七艺（文法、修辞、逻辑、算数、几何、天文、音乐），还是春秋的六艺（诗、书、礼、乐、易、春秋），其所创立的知识教育体系在今天多数归属于人文学科的范畴，致力于培养区别于万物的所谓“人性”。而当今社会建立起与工业文明相匹配的极其复杂又高深的现代教育，看似造就了大量知识丰富的“专家”，但却带来了知识分子整体上的灭绝，不仅缺乏对人的价值以及人类未来命运的思考者，连培养基本的责任与担当都成了奢望。在这个机器智能和生命编辑的时代，人文主义越来越遭遇危机，乃至面临灭顶之灾，我们比任何时候都更加需要和呼唤世界意义的守护者¹。

正是这样一个背景，诞生了数字人文。

作为信息技术在人文领域的应用，数字人文目前仍处于非常早期的发展阶段。虽然其历史可以追溯到计算器刚开始用来做文字处理的上世纪中叶，迄今已有七十余年，但“数位人文”一词是 2004 年随着《A Companion to Digital Humanities》一书的出版才得以定名，当前还不具有公认的定义，甚至连边界在哪里也众说纷纭、莫衷一是。即便如此，鉴于数字化社会已势不可挡，印刷品不再是知识生产与传播的主要媒体，Tony Hey 等提出“科学研究第四范式”即“数据驱动型研究”概念²，当所有的研究素材和方法都数字化之后，人文科学也概莫能外，数字人文必然是人文研究的未来。

数字人文是各门具体人文科学采用数字方法的汇聚和总结，是一种“方法论共同体 (Methodological Commons)”³。然而目前这个共同体还不具备库恩所说的共同的“学科范式”特征，还是各种方法、技术的大杂烩，没有形成一定的结构和规律性认识，该领域的研究者甚至对数位人文能不能成为一门“学科”还心存疑虑而争论不休。当然近二十年来数位人文的研究基础设施已得到了很大程度的建设和完善⁴，拥有了颇具影响力的协会、学会和专业期刊，定期召开国际或地区性会议，具有稳定的基金支持，尤其形成了本 - 硕 - 博的专业教育体系，目前的薄弱环节是数据资源相关的平台建设和标准规范尚未成熟，方法论体系也没有发展成型并得到公认。

1. 尤西林，阐释并守护世界意义的人——人文知识份子的起源及其使命 [M]. 上海：华东师范大学出版社，2017：1.

2. Tony Hey 等著，潘教峰等译，第四范式：数据密集型科学发现 [M]. 北京：科学出版社，2012：i.

3. McCarty W. Humanity Computing. [EB/OL].

<http://www.mccarty.org.uk/essays/McCarty,%20Humanities%20computing.pdf>

4. 劉煒等，面向人文研究的國家數據基礎設施建設 [J]. 中國圖書館學報 2016.5.29-39.

以汉学为代表的中文数位人文研究也处在一个刚刚起步的阶段。早期的数位图书馆或数位典藏成果为当下的数位人文研究提供了重要的数据支持，然而从整体上看仍不系统，缺乏规划，各学科发展也很不平衡，研究成果较为零散、微观，多是对数位技术的简单应用、对过去研究的重复验证，或者是对西方研究的一种单纯模仿，还缺乏有影响力的、独创性的成果。究其原因，图书馆等人类记忆机构在数据基础设施建设方面的滞后是一个重要瓶颈⁵。相比西方国家，我们在数据获取方面的困难要大得多：数据系统之间缺乏联通，付费墙壁高耸，造成数据获取的不充分和不完整，或者缺乏必须的数据格式（如中文文献大多以图像方式提供，文本奇缺），影响到项目的成本、成果的水平，以及对数字人文研究方法的归纳总结和教育机构相关人才的培养等，已成为汉学数字人文发展的严重制肘。

本文试图探讨一种开放的数字人文服务平台的设计，不仅满足一般人类记忆机构将数字典藏系统升级为基于数据的服务设施，重点在灵活可迁移的云平台架构设计，以及可互操作、热插入，容器化的应用 app 生态建设，同时考虑所有机构平台之间的互联互通协议和方案，以及如何应用关联数据、知识图谱、实体识别、机器学习等技术，提供人文研究各类文本、图像、社交网络、地理信息和可视化等通用工具的支持，长远支持数字人文项目的全生命周期管理。

貳、数字人文研究与数字人文平台建设

一、传统人文研究与数字人文研究

人文研究一般是人文学者针对特定问题，综合利用各种材料，透过一定方法，经过研究过程而得出结论并发表交流的完整流程。传统人文研究的素材可分为实物、文献（文本或图像）和抽象物（概念、角色等）等，依靠思考和写作完成研究，整个研究过程融合了多种研究方法，是一个从资料收集，到分析归纳，最后得出结论的线型过程。

数字人文研究也基本如此，但稍有不同的是，其“原料”对象可以进一步区分为数字文本、数码图像或由数字对象构成的一个“模型”，有学者称之为“数据态”，这个模型可以很复杂，复杂到作为某个真实系统的仿真（即所谓数字双生 Digital Twins）；其研究方法可以包括传统方法的计算器实现，或任何新颖的计算方法（统计、分析、聚类、可视化）。数字人文研究方法可以具体分类为技术、过程和行为三个方面，其过程也比传统人文研究的线型过程要复杂得多，可以是来回反复的交互过程，其成果发表和交流形式也多利用网络或社交

5. 包弼德等，数字人文与中国研究的网络基础设施建设 [J]. 圖書館雜誌 2018.11.18-25.

媒体，具有迅速、便捷、容易追踪但转瞬即逝的特点，且目前还没有⁶很好的计量与评价方法。

不同的人文学科在迈向数字人文过程中其资源特征和方法特征都有所不同，例如文学或语言学偏重于利用文本处理技术，历史学则关注实体对象的时空呈现及相互关系，哲学需要将文本抽象为特定语义的概念，等等，可以看成是数字人文上述细分要素不同配方的“跨模态”组合。当然这个讨论还只是数字人文研究一般方法的一个思考框架（如图 1 所示），目前无论是具体的人文学科，还是一般性的数字人文，其方法体系都没有定型，还处在发展变化中，也有待进一步挖掘整理。



图 1 数字人文研究的要素组成：数据与方法

二、两者的不同

（一）在研究过程方面

传统人文研究对于素材的收集、加工、处理是研究过程的开始，这是人文研究很重要的有机组成部分，而数字人文可以将数据汇集、处理的通用部分独立出来，作为研究基础设施的一部分，由专门的图书馆、档案馆等相关机构去完成，这就区分了基础设施建设工作和数字人文研究工作。目前数字人文领域大量的工作其实是基础设施建设工作，可以看到中文期刊数字人文的论文发表中大量来自图书馆信息档案学科，就是这个道理。但基础设施建设并不能代替数字人文研究，前者的目的是为了促进后者。

6. 王曉光“新技术”和“文科”不能简单相加 光明日报 2020.12.29.p14. https://epaper.gmw.cn/gmrb/html/2020-12/29/nw.D110000gmrb_20201229_3-14.htm.

(二) 在素材内容方面

传统人文通常透过管理和操控载体化的文献取得内容，限于手工处理的效率，研究的广度、深度都受到限制；而数字人文研究基于数据，平台通常就提供细粒度的知识组织，甚至建立了语义联系，使得材料的操控变得较为容易，能够进行更大范围深入研究，跨学科研究也更为容易。

(三) 在研究方法方面

传统人文研究大都采用定性的思辨方法，透过联想、比较、逻辑推理、思想实验等进行叙事或阐释；而数字人文可以采用建立模型和定量方法，进行文本分析、内容分析、时空分析、社会关系分析、统计聚类、可视化展示等，从某种程度上为人文研究提供了一定的可重复可验证的科学性保证。

(四) 在采用的技术方面

传统人文研究可能会采用田野调查、问卷访谈等；而数字人文可以透过各种技术，如机器学习、神经网络、语意标注、文本分析、量化分析、聚类算法等等。

(五) 在科研协作方面

传统的人文研究大多是学者个人或小规模团队透过多年皓首穷经、苦思冥想，坐多年冷板凳忽然顿悟，取得些许进展；而数字人文更强调大规模协同和社会网络交互，甚至大量采用众包方式，网络平台能否提供相应能力就显得非常重要。

(六) 在成果交流方面

传统人文基本上以出版图书或发表论文为最高标准；而数字人文可以同时推出网站、数据集、工具、软件、课件、博客文章、可视化作品、多媒体电子书等，专著和论文可以只是的副产品。当然数字人文对于基础设施可以更丰富和全面，包含计算设施、云平台、资源库、语料库等等。

三、数位人文平台

数字人文平台是为数字人文研究服务的，也是数字人文研究最重要的基础设施之一，平台建的好不好最终是要透过数字人文研究成果来检验的。因此在建立之初需要了解数字人文研究人员的需求，了解数字人文研究的一般规律，以及方法、过程和行为，否则也无法设计

出好的数字人文平台。当然，数字人文平台首先需要“兼容”传统的人文研究，这种需求在很大程度上数字典藏系统（数字图书馆）就能满足，然后可以进一步升级开发“真正的”数字人文平台，向人文学者全面提供基于数据的研究基础设施。

目前的数字典藏系统是一种初级版本的数字人文平台，由于其大都只是将传统的文献扫描成图像，结合元数据库提供有限途径的查询，功能十分有限，基本上只是传统图书馆的一种载体转换而已，无法满足数字人文研究的进一步需要。虽然有一些平台已开始提供一些工具，例如分词、标点、批注、词云、格式转换、实体提取、人物关系呈现及可视化等，并采用了众包理念，但总体上还较为简单，集成了一些成熟度不一的功能，没有结合人文学者的领域和场景，用户体验不够好。

现有的数字人文平台存在的最大问题还是技术上的，在内容管理上尚未采用知识图谱为代表的语意数据管理技术，还是关系数据库或者全文数据库；在体系结构上虽然已注意借鉴云计算技术，但还没有充分考虑以微服务和容积技术为基础的云原生架构，也没有考虑技术架构和内容架构分离的设计。因此很难满足人物、地点、时代、事件或特定事实主题的数据查询需求、人物或实体之间逻辑或关联关系的延伸查询需求、时空主题范围的统计分析需求以及可视化呈现的需求等等。现在的认知计算技术结合了机器学习和人工智能，已经能够提供语词概念或图像实体的提取与分析、特征比较、相似性聚类等，数字人文平台完全可以应用最新技术，实现最新功能。从平台的角度来看，还有较大可提升空间。

參、平台需求与方案设计

当今时代已不再可能举全国之力穷天下收藏，兴建四库全书那样的项目，开发包罗万象的知识平台，现在甚至连某一学科或主题领域的资源都不可能一网打尽。因此我们在构建数字人文平台或开发人文数据数据库时首先应考虑自身的优势和特点，选取一定的文献类型或学科主题，充分考虑服务对象特点和需求，设定有限目标，并做好长期建设的准备。

目前很多汉学资源收藏机构已经开发了一些颇具特色的数字人文平台，如 CBDB、Docusky、MARKUS 等，应用了许多先进理念和最新技术。本文并非简单地想介绍上海图书馆的想法和平台，而是希望着眼于未来互联互通，对构建一个整体化的汉学数字人文研究的基础设施提出一些设想。包弼德教授曾在 2018 年提出过类似的想法，他建议构建一个“中国研究的网络基础设施”，希望通过各国汉学资料收藏机构的密切合作，开发一个通用平

台⁷，使汉学资源能够互联互通，进一步促成共建共享。

这是一个非常有远见、有现实意义和可行的建议，但这个平台不必是“一个”平台，而可以是整个汉学基础设施共同构成的分布式网络，即可以由相关汉学资源收藏和研究机构各自建设，但遵循共同制订的技术标准和互操作协议，这样就保证了资源获取和服务的互联互通；同时制订一定的合作机制和业务模式，这样又能够促进互惠互利和可持续发展。

因此，本文探讨的平台即是一个在功能上力求完善、能够满足当下研究需求的独立的数字人文平台，又同时在体系架构上兼顾了基于最新语义互联网技术的互联互通，是一个尚未实现但完全具有可操作性的设计方案（如图 2 所示）。



图 2 数字人文平台需求设计

一、平台的应用系统

平台的应用系统需具有一定先进性，这主要从两个维度来考察：系统维度和应用维度，系统维度主要指系统架构的先进性，又可以分为技术架构和内容架构，技术架构提供功能实现，内容架构提供研究视图，共同满足数字人文研究需求，解决数字人文研究的痛点问题；应用维度是指是否能提供应用机构完整全面的解决方案，包括馆藏业务系统、数字保存 / 典藏系统、知识库系统、服务展示 / 应用系统等。

7. 包弼德等，数字人文与中国研究的网络基础设施建设 [J]. 圖書館雜誌 2018.11.18-25.



图 3 数字人文平台系统需求

(一) 系统维度

1. 技术架构

系统纬度首先看技术架构。目前以微服务、容器、容器编排、服务网格、开发运维一体化 (DevOps)、无服务器架构等理念为特征的新一代“云原生”技术正在席卷互联网应用。它拥有传统 IT 无法比拟的优势，可以帮助用户高效享受云技术的弹性和灵活性，使应用进一步微型化、轻型化，支持更加灵活的松散耦合，更加独立于底层基础设施平台，从而能实现热插入、平滑迁移、快速开发、迅速扩展、稳定运维、高容错等，大大降低应用成本，提高运行效率。目前云原生已经成为云时代新的技术标准。

当前还没有数字人文机构采用云原生技术，但图书馆界正在引入基于微服务的新一代图书馆服务平台，即 FOLIO，这是一个开源平台，可望采用云原生技术进行部署实施，其前后台分离的设计和“平台 +App”的架构有助于形成一个开放的软件应用生态，且设计理所当然地支持眼下另一个如日中天的技术概念，即“中台”，独立的数据中台（包括 AI 中台）将支持数字人文的各类应用模块灵活调用，在技术架构上具有无可比拟的先进性。当然该技术毕竟发展还不到十年，其成熟度和标准化程度还不是太高，微服务带来的应用复杂性还难以预料和掌控，这也是新技术必然带来的风险。

从数字人文的应用场景来看，有这样一些需求涉及技术架构：

(1) 知识单元的标识及其管理问题。所有对人文研究具有独立意义的实体或信

息单元，如文献，或人、地、时、事、物、事件、概念，以及各类属性和取值词表等，都需要有独立的标识（即 ID），并统一 ID 编码标准，通常用 HTTP URI，其相互之间的关系如有必要可以通过建立本体知识库来管理。当然建立过程可以采用自动抽取加人工辅助校验方式。

- (2) 支持跨网域的搜索发现或链接功能。例如 OAI-PMH 规范，各类 Restful+JSON 的 API 规范、联邦检索页面分析规范等等。
- (3) 微服务的容器及编排规范。
- (4) 底层关系数据库和图数据库对数据操作的管理。
- (5) 用户及权限管理。

2. 内容架构

内容架构是数字人文应用系统非常独特的架构，也是语义技术逐渐成熟带来的一种能力，具体指数位人文平台中的数字化知识内容是具有一定结构的，这个结构可以以知识本体、关联数据、知识图谱等方式进行描述和表达，例如以各类描述词表对人物、地点、时间、事件和各类对象各类属性和关系进行编码，使计算机可以对表达知识的这些语义数据（可以理解为 RDF 数据）进行操作，从而可以认为这些数据是机器可“理解”的，以至于可以认为整个知识库中的大量内容都是真实世界的一种映像，甚至可以能够让机器进行一定的“事实推理”。传统的数据库只能对字符串或二进制数据（如图像数据）进行操控，如全文检索也就是一种完全基于字符的匹配。数字人文平台对于信息资源的描述和组织可以认为是一种“数据化”过程，这一过程不一定完全依靠人类来做，很多都可以通过目前越来越成熟的机器学习和人工智能来实现。一旦机器能够读“懂”存储的信息所蕴含的知识内容，数字人文平台就能帮人文学者做很多事情，可以成为能力超强的“研究助理”，它不会遗忘任何一个知识细节，并且具有超快的计算能力。

有这样一些需求涉及内容架构：

- (1) 一致性 / 相似性计算。
- (2) workflow 定义对研究流程的支持。
- (3) 各类图像功能（如图像查询、对比、标注等）的支持。
- (4) 文本与图像关联（可提供加工平台，或研究对比）。
- (5) 提供证据链服务（纪录从底层文献到研究结果的整个过程中实体来源及变化，包括引用参考等）。
- (6) 海量数据可视化支持（遥读）。
- (7) 事实的可信度计算及排序（需建立可迭代的可信度模型）。
- (8) 众包数据加工平台的数据管理。
- (9) 数据系统迭代进化的支持（数字化、文本化、数据化）。

（二）应用维度

数字人文平台大多由人类记忆机构，如图书馆、博物馆、美术馆、档案馆等进行建设和维护。作为数字人文基础设施的主要组成机构，他们的主要业务和服务都是围绕人文资源展开的，一个较为完整的平台通常分为四个层次：

1. 馆藏业务管理系统

这主要指对物理藏品或数字藏品的载体，从收集、入藏到转移、剔除或损毁的整个生命周期过程的管理，包括藏品管理系统。它提供了所有馆藏内容最初的来源和版本信息，是循证研究的源头，并通过业务过程的管理保证整个馆藏体系是一个不断发展变化的“活”的有机体。

2. 长期保存 / 典藏系统

即上述业务管理系统中的藏品管理系统的数字化版本，通常是能够保留最真实和完整信息的保存级数位文件，借助显示或其它设备，能够还原物理藏品的内容或形态，高级形式可以看成是每个馆藏的“数字孪生”，可供研究人员进行各种实验、模拟和深度研究。当然任何数字化版本都不可能保留原始对象的所有信息，总是会有所损失，所以依赖技术的不断进步，未来可能需要对馆藏进行再次数字化。这类系统目前主要采用关系型数据库加文件系统的方式实现，更为先进的采用了 NoSQL 数据库的大数据方式，基于云服务架构，而现在应该采用云原生架构加数据中台方式，这样就能够提供底层藏品管理系统与上层知识库系统之间的桥梁，同时提供大量的 API 供知识库系统和服务应用展示前台调用⁸，这些 API 可以以标准方式发布于互联网，从而实现数字人文平台的全网域互操作。鉴于将来的数字人文研究都是基于数据的研究，有了这样的典藏系统，就可以解决绝大多数人文学者研究与教学的需要。

3. 知识库系统

目前似乎还没有一个恰当的术语来描述这样一种系统，最接近的词汇可能就是“语义知识库系统”，指应用了语义万维网技术对领域知识建立相互关联的知识体系，其知识单元是采用 RDF 形式（即主 - 谓 - 宾结构）描述的语义判断，而整个知识大厦是用知识本体语言 OWL 或 OWL2 组织起来，其背后的数学基础是一元谓词逻辑。数字人文平台的内容架构主要是由知识库系统提供的。其简化版就

8. 鲁丹、李欣、陈金传，基于 API 技术的数字人文基础设施的构建 [J]. 图书馆学研究，2019(13):42-46,57.

是采用关联数据的系统，更简化的一个版本是目前十分热门的利用“知识图谱”技术所支持的系统。这类系统在人工智能领域属于“符号学派”，与过去的专家系统同属一类，是将人的知识代码化形成规模之后，就具备了某种智能，现在与连结学派和概率学派有融合的趋势，作为人工标注或结构化的数据提供机器学习，从而具有自动的知识获取能力。数字人文平台需要大量的底层“知识库”来支撑各类数据的语义解释和关联关系，例如人名、地名、机构名、朝代、官职、谱系、辞典、词表等等，几乎所有的工具书都可以提供知识关联，所有的知识生产都是建立在过去知识的基础上，与这些底层知识库都可以建立起逻辑联系，最强大的是这些知识库都是以某种方式在整个互联网上提供共享，所有基于知识库和标准描述方式的术语词表都可以达成全网域的语义互操作。

4. 服务应用展示系统

这是数字人文平台中绝大多数功能得以实现和展现的前台，也是各类工具与后台数据进行链接的中介，通常以桌面或移动应用，以及浏览器方式提供。所有的搜索、浏览、展示（包括可视化）、众包和用户空间功能都在这里以 App 方式提供，这样有助于达成大量的第三方应用 app 的开发和发布，行成一个开放强大的数字人文应用和工具的生态环境，从而很容易实现包弼德教授提出的、为第三方数据、第三方工具、第三方图书馆免费公开的元数据访问和数据共享的规范和方案⁹。

二、平台的资源、功能与界面需求

资源完整、功能完备、界面友好，是任何一个信息系统的基本要求。当然不同的系统对这三个方面的具体需求是不同的。一个好的数字人文平台至少要在这三个方面达到最低要求，同时要注意三者之间的平衡。

（一）资源完整性

人文研究者在选定了研究问题之后，第一步就是要查询数据。很多机构在建设数据库或提供查询时只从自己已有的或订购的资源入手，这是不够的，还必须考虑到是否有办法提供外部资源的发现，甚至直接获取。要实现这一点，就要应用元数据收割方案，例如 OAI-PMH，或开发标准或个性化的 API，其中涉及很多考虑因素和资源互操作的具体技术，包括利用知识库系统实现不同系统间的语义互操作。如下图所示。

9. 包弼德等，数字人文与中国研究的网络基础设施建设 [J]. 圖書館雜誌 2018.11.18-25.



图 4 数字人文平台资源需求

(二) 功能完备性

数字人文平台需要考虑很多与过去数据库检索系统不同的功能，过去的系统主要是以文献为主要内容，根据数据库字段（即高级检索）或全文检索能够定位到具体的文献，再通过链接解析或其他方式获得原文。而数字人文系统由于提供了以“数据”为基础的存储、关联和查询能力，因此多了与“知识库”相关的很多语义功能，而且在搜索、浏览、管理等方面都能够全面支持基于知识的操作（例如 SPARQL 查询、分面组配等），有时甚至有包含逻辑推理的功能实现（如启发式搜索）。

数字人文平台还有一个特质是要利用众包让用户参与到系统的建设中来，这是当前几乎所有数字人文应用都采取的方式，因为仅仅通过图书馆或相关机构工作人员的工作是不可能实现海量高质量数据加工的。



图5 数字人文平台功能需求

(三) 用户友好性

当前的信息系统对用户友好性的要求越来越高，这也是对系统界面提出的要求，除了一般的方便友好、美观简洁之外，能否提供良好的个性化服务成为系统是否具有粘性、能否留住用户的重要特性，而且个性化服务大量采用了人工智能技术，详见下图。当然，由于个性化的前提是需要有用户注册登录等用户管理功能，且对用户的行为也会进行一定的收集，这涉及到用户隐私问题，平台在设计开发时必须考虑到隐私保护与个性化之间的平衡，很多研究工具的提供应该能同时支持本地脱机版和上传网络版两种不同的运行方式，当然两者在功能细节上可以有所不同。



图 6 数字人文平台界面需求

三、平台的工具开发

利用大量的数字人文工具进行研究是数字人文区别于传统人文最重要的特点之一。工具方法是方法的重要组成部分，成熟的方法往往通过工具的开发而得以固化，并且负载了大量前人的经验总结。传统人文研究能够独立的工具不多，且数据的收集、阅读和加工处理往往是一体化、个人化的，工具很难独立于资料，有的甚至很难独立于研究团队。这也是为什么有许多人文社会科学学派往往是得益于独特的方法。

工具要求越丰富越好，但这里讨论的只是人文研究可能用到的具有一定通用性的工具，以及这些工具的常见功能，数字人文学者可以通过这些工具的组合，结合资源和研究过程，发展出自己独特的方法。这些工具可以有一定的独立性，但依附于平台能够更好地发挥作用，因此平台将致力于深入研究人文学者的需求，推出大量的标准规范，从而让大量第三方都能够开发自己的独特工具，甚至工具与资源或知识库的结合体，从而有助于形成一个应用生态，以及工具 App 市场。

这里将工具划分为平台性工具（包括数据工具、IIIF、GIS、文献计量工具、阅读工具、社会关系工具）、文本工具、图像工具、知识图谱工具、机器学习工具和可视化工具六大类。上述分类的合理性需要进一步探讨，其中涉及的内容也远不是对各类工具的穷尽例举，仅仅作为一个讨论的基础，供具体进行工具开发和平台建设参考。

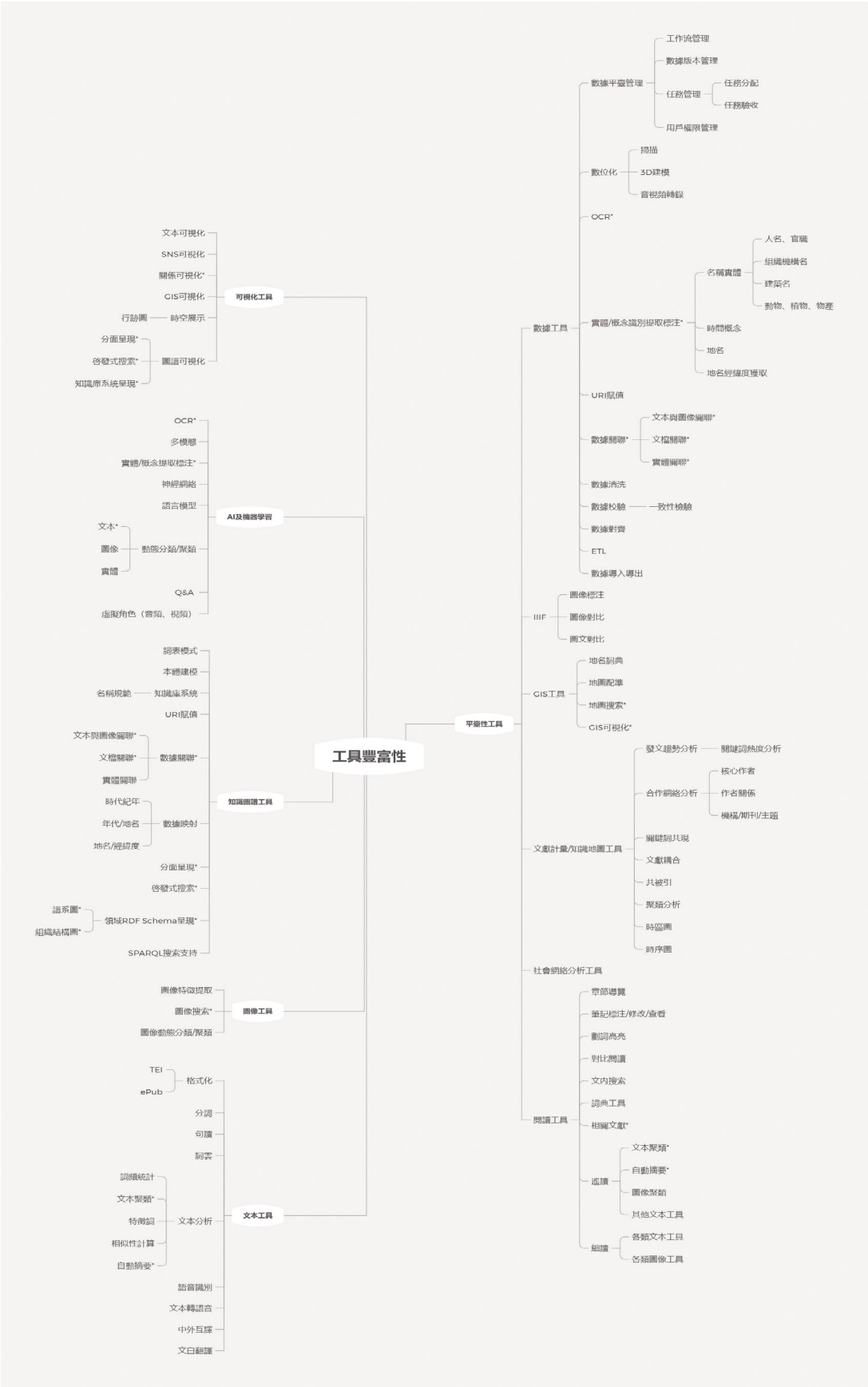


图 7 数字人文平台工具需求

(一) 平台性工具

这里的平台是指网络上可以实现一定的功能、有特定输入输出的环境，平台性工具就是依附于平台或自身就可以独立成为一个平台的工具，它通常不是一个简单工具，而需要结合一定的数据，需要一些组件的配合，并经过一定的流程才能达到目的。例如 IIIF（国际图像互操作框架）就是一个功能强大的综合性图片平台，由多个服务器灵活组合而成，它本身就可以成为数字人文的服务平台，这里之所以作为一种工具，因为它提供了大量的关于图像的操作功能，如搜索、缩放、旋转、标注、比较等等，可以应用于人文研究，非常强大。类似的还有数据处理平台工具、GIS 平台工具、文献计量平台工具、社会网络分析工具以及阅读平台工具等等。

(二) 文本工具

文本是数字人文利用最多的资源类型，文本工具也是数字人文工具中种类最多、使用最频繁的工具，也是目前开发最成熟的工具类型。上图列出的是常用工具，一些综合性的文本工具，如“远读”、“细读”列在平台性工具类目下。

(三) 图像工具

通常所有的图像扫描、处理软件都可以作为数字人文的图像工具，这里仅列出数字人文项目非常常用的工具类型，如图像特征提取工具、图像分类 / 聚类工具和基于图像的搜索工具等，图像平台 IIIF 已作为平台类工具栏出。

(四) 知识图谱工具

知识图谱是数字典藏向数字人文进化的关键技术之一，这里将关联数据、语义网技术都归入知识图谱。这类工具包括了实体提取、URI 赋值、词表模式、本体构建等语义化工具，本体 / 词表管理、语义映像、RDF 语义数据存储等语义管理工具以及 SPARQL、启发式搜索、分面呈现等语义搜索、展示和利用工具等。

(五) 机器学习工具

当前数字人文的大量应用都用到了人工智能领域的机器学习技术。从 OCR 到实体提取，从神经网络到深度学习，无一不能应用于数字人文研究的各个过程。机器学习最大的特点是离不开数据，尤其是海量的数据，因此数字人文平台中的数据是其产生作用的前提条件，而由数据训练出来的机器学习模型又可以应用于更广泛的数据中，这是它的运作方式，也是它的价值所在。

(六) 可视化工具

可视化是数字人文进行数据操控、展示和结果呈现必不可少的工具，也是数字人文区别于传统人文的重要特质。可视化虽然有很多工具，但现在基于互联网的工具已成为主流，正在成熟起来。它后台连接的数据可以是平台上已有的数据，或者挖掘出来的数据，或者是用户上传的数据，是否支持多种应用方式取决于平台架构设计的灵活性。

肆、案例：历史人文大数据平台

上海图书馆正在建设的历史人文大数据平台，就是应用上述理念和技术，依托自身资源，向全社会提供一个先进、开放、全面的数字人文服务平台。打造这个平台主要有三个目的：一是升级原有的数字典藏系统；二是提供基于“知识”的数字人文服务；三是试验一些互联互通共建共享的新协议新模式。其实就是作为对前述数字人文发展趋势进行应对的一种尝试。

实现这三个目的有两条现实可行的路径：其一，从现有的数字典藏系统出发，也就是从目前上海图书馆馆藏特色资源出发，升级技术架构和内容架构：技术架构全面微服务化、容器化和平台化，支持外部资源与服务通过各种标准或非标准方式（推荐 Restful API）接入；内容架构进行“数据化”改造，支持“基于知识的服务”。其二，从数字人文研究者的角度出发，规划所有人文资源的整合方案，从提供资源、到提供平台环境（包括工具），努力实现主要数字人文应用场景的“一站式”的服务。

一、历史人文大数据平台的建设规划

上海图书馆走上数字化道路已经有四分之一世纪。从 1996 年位于上海淮海中路的“新馆”开馆，就开始古籍数字化项目，并且参与了中国最早的由国家图书馆牵头的“试验性数字图书馆计划”，成立专门部门，每年耗费巨资进行特色资源的数字化工作，从无间断。

仅仅数字化是不够的，提供知识服务是图书馆的根本宗旨。早期重视数字化，但对于数字典藏系统的建设并没有充分重视，因此数字资源的整合服务一直没有充分开展。到 2016 年，上海图书馆尝试以最具特色的馆藏家谱资源为案例，开始了以服务为导向的系统开发尝试，取得了不错的效果，迄今家谱系统一直是数字典藏中利用效果最好的资源之一。

为了建设具有知识关联的数字人文服务系统，底层知识库平台建设是必不可少的，这也

是数字人文基础设施最困难的内容。近几年我们陆续构建了人名规范、地名规范、地理名称规范、机构规范等等规范知识库，可以支持目前列入计划的特色资源库的底层知识关联，并开始开发一些工具，提供众包、标注、分析、可视化等。

正是由于有了底层知识库的支持，上海图书馆的特色资源库才有可能做一个全面规划，将来各类数字人文系统可以在一个统一的平台上，我们称之为历史人文大数据平台。虽然这一平台尚未建成，但已经经过了初步尝试，证明了技术和工程上的可行性和可能性，且数据也有一定规模。目前我们除家谱库外，正在开发的还有古籍库（包括精品善本库）、碑帖库、地方志库、手稿尺牍库、名人档案库（如盛宣怀档案、张佩纶档案等）、民国资源库（包括书刊报）等，这些文献如按照数字人文研究的要求，可以建立无数个基于各类学科或主题的知识库，可以汇总在一个平台上提供满足各类需求的统一服务，通过一定的开放链接协议，可以将全网域的各类资源连为一体，组成一个虚拟汉学数字人文平台。

二、平台应用场景

对于一个资源众多、用户复杂、目标并不单一的服务平台来说，“主页”概念是不适用的。历史人文大数据平台虽然设计了一个主入口（主页，如图），但它的作用只相当于“游客中心”甚至是“疏散中心”，主要起到宣传、导航、资源发现和用户培训的作用。任何一个简单的搜索，都可以返回所有资源库中（甚至外部联邦检索或搜索引擎）的命中内容，这样能够让“随便逛逛”的读者也有所收获，同时用户对自己感兴趣的主题可以通过哪些资源库获得有一个非常直观的认识，从而带有目的的读者能够迅速找到属于自己的入口。



图 8 历史人文大数据平台主入口

平台对所有的专题库（包括文献库、知识库和工具库三类）都有一个入口，有一个页面提供了一个完整列表（如下图）。其中大多数文献库都是元数据库加扫描图片方式提供，个别有全文，知识库和工具库都支持响应式 H5 接口（可嵌入各类 App）。

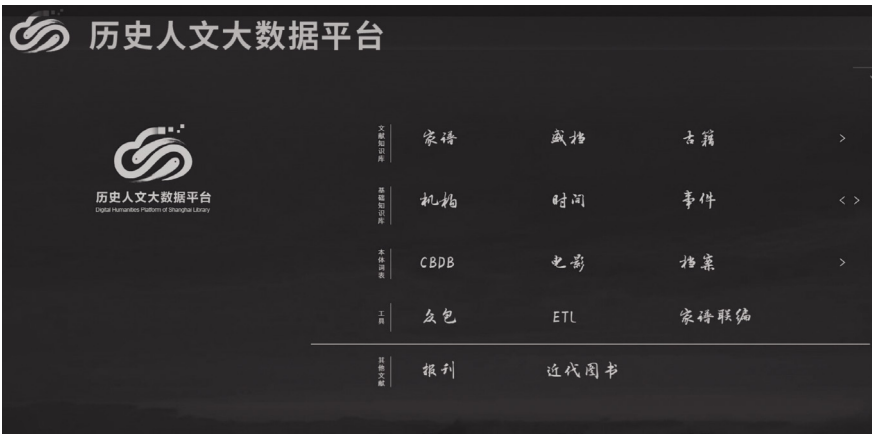


图 9 平台各类资源库和功能入口

我们把平台用户分为四类：普通用户、专业用户、系统用户和机器用户。普通用户是无需用户认证即可来“随便逛逛”的用户，平台会有很多在线展览、人文讲座、推广活动、技能培训等内容发布。专业用户是平台服务的主体，通常是经过注册的研究人员或大学师生，也可能是相关机构中的个人用户（登录为单位用户或以 ip 控制方式提供权限管理），这类用户除非使用主页中的搜索框进行资源发现（搜索框在各相关页面也都会出现），一般无须从主入口进入，只要浏览器保留了登录 Cookie，域名会直接将其定位到他自己的个性化页面，该页面已经将其经常使用或可能用到的专业资源入口与各类服务功能集成在一起了，每个用户的专业入口都是个性化的，与“我的空间”捆绑，用户如果不满意，也可以在“我的空间”中修改参数设定。系统用户是那种参与数据加工或项目研发的用户，这是“平台性”的具体体现，作为平台，不是一个私有的封闭系统，而应该有一定的开放性，属于整个小区，允许大家参与共建、分享成果，因此必然有一类用户通过贡献内容、参与数据加工或功能开发而具有更多的权限。最后一类“机器用户”就是指通过 API 或其他接口直接消费数据的计算机程序，这样能将平台与互联网上其他应用连为一体，使“一站式”服务成为可能。

平台提供的所有服务可以分为“场景”、“故事”和“功能”三个层次，区别三类用户，提供不同的功能组合，详细如下表所示。“场景”可粗略地对应于数字人文研究的“行为”，例如搜索、浏览、下载、阅读等等；“故事”是组成场景的若干种应用；而“功能”是平台提供的最小单位模块，通常对应于目前云原生应用架构中的“微服务”。这里的服务基本都是用户直接可用的部分内容，后台其实还有大量的微服务，由于与平台用户并无直接关系，这里就不详述了。

表 1 历史人文大数据平台主要功能一览

用户类别 场景类别		普通用户（非注册）	专业用户（注册）	系统用户	
Epic	Story 故事	Feature 功能			
搜索	简单搜索 / 全文搜索	精确或模糊匹配可选			
		支持提问自动分词			
		支持前后缀搜索			
	高级搜索	分字段限定	支持专业检索，即逻辑表达式在一个搜索框内实现高级搜索		
			支持字段按数据类型（如年代）限定，支持范围限定		
			位置限定搜索		支持正则表达式搜索
				统计式搜索	
		智慧搜索（框式搜索）：I feel lucky（用各种技术猜测用户需求的检索）			
		二次搜索：在任意结果集中再进行限定搜索（支持逻辑式限定）			
		知识搜索		基于实体名称 / 概念的搜索，如人名、地名、机构名、事件、物体	
			启发式搜索		
				SPARQL/CQL 搜索（部分）	
	图像搜索		基于输入图像的匹配（只支持部分图像库）		
浏览	排序	检索结果支持多字段排序（分主次），支持拼音 / 笔画 / 时间等排序方式，支持正序 / 逆序			
	分面组配浏览导航	按结果集分面情况选择			
	知识导航		按知识本体呈现结果（相关图式参见可视化部分）		
	地图浏览	检索结果具有地理或时代属性的，可以在地图上进行时空呈现			
	对比		不同检索结果集比较呈现，可多窗口（参数可选？）对象比较参加阅读场景		
	结果可视化	结果集基本属性的可视化	下载至个人空间，再支持本地下载	可视化参数可调	
下载	元数据下载到个人空间		需经授权才能下载	可订制 Schema 数据格式，与高清对象数据一起打包下载	
	对象数据下载		可选择一定的参考文献各式，批量下载		
	参考文献格式下载		可选择一定的参考文献各式，批量下载		
阅读	结构导览	目录章节导览、各类跳转、超链接			
	文内搜索	搜索词高亮并可导航（下一个）			
	辞典工具	可外挂多语种 / 专业辞典			
	文本朗读	自动机器语音朗读 / 生僻字词朗读（随辞典工具）			
	书签 / 高亮	非专业注册用户存于本地，专业用户上传个人空间，并提供社会化（分享）选项			
	批注 / 修改		以 W3C annotation 标准方式（RDF）实现，可分享		
	对比		选择不同对象多窗口打开比较和分别批注（元数据结果集比较参见浏览场景）		
	文献推荐		阅读及写作过程中不断推荐相关文献，推荐模型相关参数 / 阈值可选		

用户类别 场景类别		普通用户（非注册）	专业用户（注册）	系统用户
Epic	Story 故事	Feature 功能		
高级 阅读 / 阅读 工具	细读		文本标注：根据词表（内建 / 挂接词表或用户导入词表）标注	
			文本标注：句读、分词、词性标注等	
			实体识别并标注	
			文白翻译	
			文本互译（外挂翻译工具）	
			文本格式化，格式标注 / 转换 （如 TEI、ePub 等，参见数据加工工具）	
	通读		特征词 / 词云生成	
			自动摘要（模型参数可调）	
			文本 / 图像相似性计算、聚类分析	
			风格分析	
			情感分析	
	共读		支持用户书签 / 标注 / 批注等信息的共享和挖掘	
众包 平台	任务管理			任务发布、分配
	用户管理			注册及权限管理、资金监管
	工作认领			工作量计算、统计反馈
	质量管理			质量校验、纠错、版本控制
可视 化	文本可视化	各种词云	各种高阶词云、关系图、热力图	参数可调
	关系可视化		各类 d3/eCharts 图示	参数可调
	GIS 时空	行迹图	时空叙事 / 现地研究	
	图谱可视化		各类知识本体可视化，包括人物生平 / 大事年表 / 互动展览 / 组织机构图标等	

伍、结论

数字人文平台建设的愿景是让人文研究不再困难。从雅典学园到文艺复兴，从鲁国杏坛到康梁变法，两千年来人文学者的创造性思考从来都是依靠个体的博览群书与博闻强记，依靠师徒私授或学派论战，思想的诞生、学说的完善，以及对社会实践的影响主要依靠的是个人的能力，人文知识的产生、发展和传播的整个过程是偶然、不清晰和不确定的，每位学者都要从最原始的篇章学起，遍历所有典籍并考察整个源流，穷极一生只能成为专家而无法成就大家，而数字人文正在第一次给人文研究带来革命。

针对人文研究的完整过程，数字人文已能够分而治之：首先使数据查询和获取不再困难，其次使知识存储、传播和利用不再困难，然后让分析、比较，形成观点不再困难，最后

使结果展示、交流和争鸣不再困难，人文学者不再是单打独斗而是集团作战，无须管中窥豹而是直接综揽全局尽情把握，人文研究的规律与方法将得到更好的揭示，人文成果的发表形式将不再是报刊杂志，人文学说的比较与评价将更方便地在实践中得到检验和反馈，为人文研究提供的服务能力将更快地得到迭代和提高。

照此发展下去，那么问题来了：如果数字人文充分采用了人工智能技术，推向极致，可能机器也能自动进行人文研究。此时的人文，还是人文吗？

其实数字人文的终极意义还是在于以科技强化人文，而不是将人文变成被动机械的对象，进行去价值化和无意义化。最终的意义呈现，其主体是人类自身。当所有的人文都是数字人文时，“数字”与“人文”才能够真正合为一体，那时数字的工具性特征便不再重要，人文研究此时便能回归本源，真正彰显人类的价值和生命的意义。

这也是我们要用尽所有先进技术，推进数字人文平台的开发与建设的根本原因所在。

参考文献

- ¹ 尤西林. 阐释并守护世界意义的人——人文知识分子的起源及其使命 [M]. 上海：华东师范大学出版社，2017：1.
- ² Tony Hey 等着，潘教峰等译. 第四范式：数据密集型科学发现 [M]. 北京：科学出版社，2012：i.
- ³ McCarty W. Humanity Computing. [EB/OL].
<http://www.mccarty.org.uk/essays/McCarty,%20Humanities%20computing.pdf>
- ⁴ 刘炜等. 面向人文研究的国家数据基础设施建设 [J]. 中国图书馆学报 2016.5.29-39.
- ⁵ 包弼德等. 数字人文与中国研究的网络基础设施建设 [J]. 图书馆杂志 2018.11.18-25.
- ⁶ 王晓光“新技术”和“文科”不能简单相加 光明日报 2020.12.29.p14. https://epaper.gmw.cn/gmrb/html/2020-12/29/nw.D110000gmrb_20201229_3-14.htm.
- ⁷ 包弼德等. 数字人文与中国研究的网络基础设施建设 [J]. 图书馆杂志 2018.11.18-25.
- ⁸ 鲁丹、李欣、陈金传. 基于 API 技术的数字人文基础设施的构建 [J]. 图书馆学研究, 2019(13):42-46,57.
- ⁹ 包弼德等. 数字人文与中国研究的网络基础设施建设 [J]. 图书馆杂志 2018.11.18-25.



項潔
Jieh Hsiang

項潔，目前為國立臺灣大學資訊系特聘教授及臺大數位人文研究中心主任，曾擔任臺大圖書館館長及臺大出版中心主任，也曾任臺灣人工智慧學會理事長，並為臺灣數位人文學會創會理事長。於 1993 年返臺前，為紐約州立大學（石溪分校）計算機科學系專職教授，亦曾擔任法國巴黎大學、法國南錫第一大學、義大利米蘭大學的客座教授。1998 年至 2001 年，項教授擔任國立暨南國際大學科技學院的創始院長，並曾於 2002 年至 2008 年擔任國立臺灣大學圖書館館長，於 2008 年至 2016 年擔任國立臺灣大學出版中心主任。項教授自 1996 年以來即持續投入於數位典藏及數位人文的相關研究工作，並成立了臺灣數位人文學會，並擔任第一任理事長。項教授已撰寫發表了 200 多篇學術論文，以及 6 本關於數位人文領域的專書著作。

Jieh Hsiang is a Distinguished Professor in Computer Science of the National Taiwan University and the founding Director of its Research Center for Digital Humanities. Before returning to Taiwan in 1993, he was a full professor at Stony Brook University. He had also been a visiting professor at the University of Paris, University of Nancy and University of Milan. He was the founding Dean of the College of Science and Technology of the National Chi-Nan University from 1998 to 2001. He also served as the University Librarian of NTU Library from 2002 to 2008, and the Director of University Press of NTU from 2008 to 2016. He has been working on Digital Archives and Digital Humanities since 1996, and founded the Taiwanese Association for Digital Humanities and served as its first President. Professor Hsiang has written over 200 scholarly papers and chapters, as well as 6 books on Digital Humanities.



張毓哲
Yu-Che Chang

張毓哲，國立臺灣大學數位人文研究中心研究人員。

Yu-Che Chang is a research associate at Research Center for Digital Humanities, National Taiwan University.



傅莉雯
Li-Wen Fu

傅莉雯，國立臺灣大學數位人文研究中心計畫研究專員。

Li-Wen Fu is a research associate at Research Center for Digital Humanities, National Taiwan University.

歷代寶案脈絡分析系統之建置¹

A Context Discovery System of LidaiBao'an

張毓哲²

Yu-Che Chang

國立臺灣大學數位人文研究中心研究人員

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

傅莉雯³

Li-Wen Fu

國立臺灣大學數位人文研究中心計畫研究專員

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

項潔⁴

Jieh Hsiang

國立臺灣大學資訊工程學系暨研究所特聘教授

Distinguished Professor in Computer Science, National Taiwan University

摘要

《歷代寶案》是琉球王國與周邊國家往來的外交文書史料，自 1424 年起，迄於 1867 年，時間涵蓋 444 年，空間範圍從東北亞延伸至東南亞，包含了與中國、臺灣、東南亞各地相關的往來紀錄，體現了朝貢秩序的運作。

現存於臺大圖書館的《歷代寶案》抄本，由臺北帝國大學小葉田淳教授，委託沖繩久場政盛進行抄寫，為全世界最完整且接近原貌的版本，也是遺存的諸多抄本中，完整度最高的一部。

臺大圖書館與數位人文中心合作，透過全文化、重要詞彙擷取、正規化與標記，建置一個以研究者為導向，具有後分類、脈絡關係、可分析統計、可視覺化觀察的數位人文系統，目前系統內收錄有第一集（42 卷）、第二集（109 卷）、第三集（13 卷）及咨集（1 篇）、別集

1. 本文的完成特別感謝國立臺灣大學數位人文研究中心董家兒小姐和胡乃文小姐的協助。

2. 國立臺灣大學數位人文研究中心研究人員，負責歷代寶案內容解析與詮釋。

3. 國立臺灣大學數位人文研究中心計畫研究專員，負責系統研發與建置。

4. 國立臺灣大學資訊工程學系暨研究所特聘教授。

(2 篇)，共 4,294 件、全文約 330 萬字；重要詞彙擷取物名 5,669 個、官名 3,124 個、人名 8,613 名、地名 2,073 並進行 GIS 地理資訊座標的定位，透過系統不僅能觀察《歷代寶案》文獻的整體資訊，並可進行檢索、分析與觀察，使這批珍貴研究史料能更多元的應用。

關鍵詞：《歷代寶案》、數位人文、臺大圖書館、脈絡發掘、朝貢貿易

Abstract

Lidaibao'an is an archive of diplomatic documents between the kingdom of Liuqiu (present day Ryukyu Islands) and its surrounding countries. Starting from 1424 until 1867, when it was annexed by Japan, it spans across 444 years and contains documents concerning South Asian countries, Japan, Taiwan, and in particular tributary records with China.

The copy of Lidaibao'an, kept at the National Taiwan University Library, was transcribed by Atsushi Kobata in the 1940s, who then brought it back to the Taipei Imperial University where he was teaching at the time. The NTU library's copy is considered the most complete among all existing versions.

This talk describes a digital system of Lidaibao'an developed jointly by the NTU Library and the Research Center of Digital Humanities of NTU. Through the transcription of the full text, extraction of named entities, normalization of dates, and taggings, we have built a research-oriented full-text retrieval system that features post-query classifications, context discovery (through both metadata and tagging), statistical contexts, and visualization. The system contains the NTU version of Lidaibao'an in its entirety, with 4,294 documents and a full text of more than 3.3 million characters. Important named entities extracted include 5,669 product names, 3,124 office titles, 8,613 person names, and 2,073 location names with geographic coordinates. We hope such a context discovery system can encourage research using Lidaibao'an.

Keywords: Lidaibao'an, Digital Humanities, NTU Library, Context Discovery, Tributary Trade

壹、前言

《歷代寶案》是琉球王國與其周遭國家往來互動過程產生的外交文書，收錄時間起於 1424 年（明永樂 22 年、尚巴志 3 年），迄至 1867 年（清同治 6 年、尚泰 20 年），是探究琉球國與中國之間朝貢貿易運作的外交史料，也是了解東亞進入大航海時代以前琉球國與東亞各國的貿易史料，為呈現琉球王國自身歷史與東亞海洋史多元性的重要文書。

臺灣保存了《歷代寶案》最完整的版本，係源自於 1935 年臺北帝國大學（今臺灣大學）助教授小葉田淳委託久場政盛抄寫「久米村保管本」，1972 年經由國立臺灣大學圖書館影印出版。⁵ 沖繩方面，因著史料的語言限制，沖繩縣立圖書館史料編集室、沖繩縣文化振興會公文書管理部史料編集室、沖繩縣教育廳文化財課史料編集班等公部門和財團法人為使內容平實易懂，長時間投入編修《歷代寶案：校訂本》、《歷代寶案：譯注本》，致力於推廣與普及。

隨著各國檔案的開放和出版，近年來越來越多琉球相關的史料出土。為了擴大《歷代寶案》的使用性，國立臺灣大學數位人文研究中心和國立臺灣大學圖書館進行合作，在既有的整理基礎上，進一步針對此文書的資料特性與便利學術研究使用等目標進行「歷代寶案脈絡分析系統」規劃，以使用者需求為導向，進行系統相關功能的開發與建置，方便使用者瀏覽、檢索與分析資料，讓使用者得以在龐大的資料中更快速且精確的得到檢索結果，從中獲得檢索結果的重要特徵，甚或是發現可供探討的新議題，達成本系統的建置目標。

本文首先就歷代寶案數位化內容來說明系統包含的檔案特性與限制，進一步說明如何針對《歷代寶案》的內容特性、學術研究使用目標的需求，進行系統規劃建置與研發相關功能的特色，並透過實作範例說明使用系統功能來輔助檔案應用與研究分析，期許藉此吸引更多使用者運用歷代寶案脈絡分析系統（以下簡稱為歷代寶案系統）進行各領域的研究，持續發揮《歷代寶案》的史料價值，提供更多面向的東亞史研究議題。

貳、歷代寶案數位化內容與特性

《歷代寶案》的原本王府保管本與久米村保管本分別毀於 1923 年關東大地震和 1945 年沖繩島戰役。⁶ 今日所見的各版本大多是根據久米村保管本產生的寫本和影印本，包含國立國會圖書館寫本（約 1802 年）、東恩納寬淳寫本（1932 年）、東恩納寬淳影印本（1934-

5. 小葉田淳，〈歷代寶案について〉，《史林》第 46 卷第 4 號（1963），頁 528-530。

6. 沖繩縣教育委員會，《歷代寶案の栞》（沖繩：沖繩縣教育委員會，2018），頁 3。

1935 年)、鎌倉芳太郎影印本(1933 年)、舊沖繩縣立圖書館寫本、臺灣大學藏寫本(1935 年)、東京大學史料編纂所寫本(1941 年)，另有鄭良弼寫本(約 1814 年)。⁷ 各版本缺失部分略有不同，以臺灣大學藏寫本最為完整。

國立臺灣大學圖書館在 1972 年將臺灣大學藏寫本出版為《歷代寶案》15 冊，本系統的數位資料即以此為基礎。數位化過程首先將文本按照頁數依序掃描，共掃出圖片 9,398 張，再依照內容分件、貼標，完成 4,381 件。

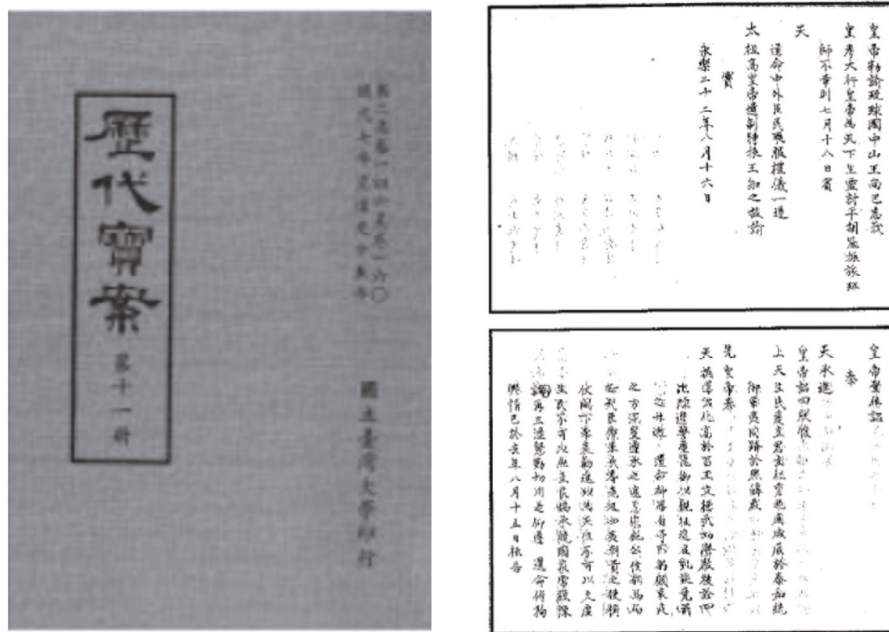


圖 1 臺灣大學圖書館 1972 年出版之《歷代寶案》封面與內頁

本系統提供全文內容打字而成的全文檔。全文檔經過文本比對、一校、二校與專業審查等流程，除了修訂繕打過程的人為錯誤，更針對當年原件的抄寫錯誤、內容錯置與正異體字等問題，逐一釐清修改，透過將上下文錯置的內容拆件、將分散在不同頁的同一內文併件，或將不合同件內容的文字挑出做為增件，最終總共完成 4,293 件，約 330 萬字（包含標點與分段）之全文檔。

7. 沖繩縣教育委員會，《歷代寶案の榮》，頁 3。

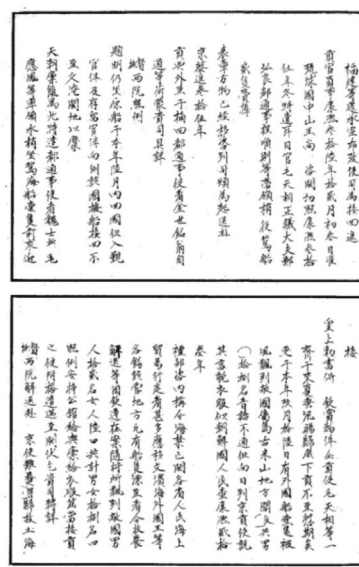


圖 2 以共同協作之後臺進行全文標點校對

說明：由於全文標點校對為耗費人力的工作，故訂有校對原則供標校者作為統一的全文標校規範。此外，因標點符號涉及編碼問題，後臺建有標點符號小鍵盤供標校者使用。

全文檔完成以後，下一步是建置詮釋資料。首先利用程式自動擷取出每一件的成文日期、標題、人名、官名、地名、物品等欄位，再使用共同協作之後臺，人工訂正詮釋資料，並對內文再次進行檢視。如此全文、詮釋資料、圖片三者對照，同步修改與更新。

以下針對詮釋資料各欄位的建置方式略作說明：

一、成文日期

琉球國受到東亞文化圈的影響，《歷代寶案》收錄之文書皆以中國朝代為紀年。本系統除了將程式自動擷取的明、清朝代與年份正規化外，也轉換成西元年的表示方式，並可以成文日期對文件進行排序，跳脫原件依照集／卷的編排方式。

二、標題

抓取每件文書首行首句的段落內文作為標題，再由人工逐件修改無意義或有錯字者。

三、人名、官名、地名

首先，利用本中心過往製作其他中文資料庫的關鍵字做為種子詞彙，將這批種子詞彙在歷代寶案系統的全文中找出頻繁出現的前後綴詞做為詞夾子，再利用詞夾子自全文夾出另一

批詞彙，再以人工進行判讀與修正。⁸ 大約由人工判讀 1,000 件後，再以這 1,000 件抓出來的詞彙倒回去重新抓取，如此往復，程式與人工互相輔助，以提高詞彙擷取的準確度。

四、物品名

同樣使用如上所述之程式擷取與人工校正的流程，但本系統不只是單純記錄物品品項，而是以「物品名稱：數量（阿拉伯數字）：單位」之特殊形式，詳實記錄文本提及之物品狀況。

件號 ntu-1088321-0011600117.txt 檔案ID pdsg-ntu-1088321-00059

來源：歷代寶案第一集 卷三

官名分隔符號「,」

皇帝,琉球國世子,兵部副理官,行人司行人,琉球國中山王,妃,國王,王舅

國入參考：琉球國中山王,琉球國世子,琉球國國王,王舅,皇帝,皇

人名分隔符號「,」

尚質,馬宗毅,張學禮,王核

國入參考：尚質,馬宗毅,張學禮,王核,世子,尚,張,王,皇

地名分隔符號「,」

琉球國

國入參考：琉球國,中山

物品名分隔符號「,」

文幣,錦綉:2:匹,青綵綵:3:匹,藍綵綵:3:匹,藍素綵:3:匹,衣素:2:匹,綢綵:2:匹,錦:3:匹,綢:4:匹,羅:4:匹,紗:4:匹,青綵綵:2:匹,藍綵綵:2:匹,妝綵:1:匹,綢綵:1:匹,藍素綵:2:匹,衣素:2:匹,錦:2:匹,羅:4:匹,紗:4:匹

國入參考：綵綵:3:匹,羅綵:2:匹,藍素綵,藍素綵:2:匹,羅綵:2:匹,門綵:2:匹,門綵:1:匹,馬,紗:4:匹,綢:4:匹,羅:4:匹,綢綵:2:匹,衣

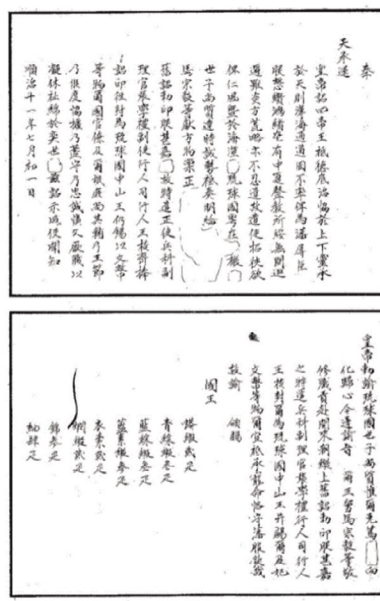


圖 3 以共同協作之後臺進行詮釋資料建置

說明：詮釋資料欄位下方列有程式擷取的參考詞彙，提供詞彙擷取人員判讀與修正。

再者，就文書編纂之冊別而言，《歷代寶案》依據編纂時間，分為第一集、第二集、第三集，以及咨集、別集各一。1972 年國立臺灣大學圖書館將《歷代寶案》分成 15 冊出版，文書內容的數位化與分件時也以 15 冊之頁碼作為參考依據。為了不使原有的文書脈絡消失，在建立詮釋資料時，每件全文的集／卷資訊亦被單獨提取出來，在本系統與書籍資訊一同呈現，提供清晰的文件出處來源。也就是說，使用者可以透過以下二種方式回溯歷代寶案的出處與來源：

8. 文本截詞工具，DocuSky 數位人文學術研究平臺，項潔教授主持，杜協昌博士設計開發，<https://docusky.org.tw/DocuSky/docuTools/TermClipper/TermClipper2020.html>。

(一) 《歷代寶案》卷目與書訊概況

集	卷	涵蓋時間 (中曆)	編纂方式	編纂時間
第一集	卷一至卷四三	永樂 22 年 - 康熙 35 年	文書性質	康熙 36 年
第二集	卷一至卷二〇〇	康熙 36 年 - 咸豐 8 年	年代順序	雍正 4-7 年
第三集	卷一至卷十三	咸豐 9 年 - 同治 6 年	年代順序	缺
別集	咈英情狀、冠船之時唐人持來品貨物錄	道光 24-27 年、康熙 58 年	專題	缺
咨集	文組方	乾隆 38-49 年	專題	缺

(二) 國立臺灣大學圖書館 1972 年出版之書籍資訊

包含冊數、書冊登錄號、書冊頁碼 (圖片頁碼)、出版者等。

編號	pdsg-ntu-1088321-00006
出處	歷代寶案脈絡分析系統第一集卷一
標題	皇帝敕諭琉球國中山王世子尚〔巴〕志
日期	明洪熙1年2月1日(1425-02-18)
官名	內宮,皇帝,太宗文皇帝,世子,琉球國中山王,琉球中山王
人名	尚巴志,思紹,柴山
地名	琉球國,琉球
物品	
附註	
書冊資訊	
書冊編號	ntu-1088321
文集	歷代寶案
冊	第1冊
出版者	臺大圖書館
頁碼	8
件號	ntu-1088321-0000800008-0000001.txt

洪熙元年二月初一日

皇帝敕諭琉球國中山王世子尚〔巴〕志：昔我皇考太宗文皇帝，恭膺天命，統御萬方，恩施一視，遠邇歸仁。〔爾〕父琉球〔國〕中山王思紹，聰明賢達，茂篤忠誠，敬天事上，益久弗懈，朝貢有常，罔愆于職。我皇考嘉乃勤款，良用褒錫。肆朕繼承大統，弘敷治化，尤重繼承，念爾父告終，已逾再歲，非有嗣嫡之賢，曷膺傳襲之重。茲特遣內官柴山齋敕：命爾世子尚巴志為琉球國中山王，以繼其世。於戲，尚立忠立孝，恪守藩服，修德務善，以福一國之人，則爵祿之榮，延于無窮，尚其祇承朕命，無怠無忽。故茲敕諭，宜體至懷。

圖 4 詮釋資料 (左) 與數位化全文 (右) 在系統上的呈現

說明：左上為核心欄位，左下為書冊資訊，右為全文內容 (可選擇橫向文字或直行文字)。

如此既保留《歷代寶案》原始的文書脈絡，同時亦呈現數位化所依循的編輯概念，使用者能根據其需求交互參照。

綜而言之，《歷代寶案》雖然以中文書寫，但內容涉及琉球國的歷史文化與外交事件，

充滿該國獨有的詞彙。不論是人名或官名，都與明、清兩朝的中國文書有很大差別，即使利用常見的中文詞庫去擷取，也會漏掉許多特殊詞彙。因此，透過人工方式標記新詞彙，再倒回去擷取，才能逐漸累積該文書的詞彙量。地名除了明、清兩朝中國與琉球國周遭常見的地名外，亦包含東亞國家與其沿海地區地名，具有跨區域之特殊性。物品名則有別於本中心過去處理的文本資料性質，故以特殊形式記錄物品，期盼能對研究提供協助。透過往復擷取和校正的過程，塑造出一份屬於琉球國與東亞海域的特殊詞彙清單。

參、系統建置規劃與功能說明

為了發揮《歷代寶案》價值和數位化特色，由國立臺灣大學數位人文中心、國立臺灣大學圖書館、研究中國與琉球關係的學者專家，針對資料特性、使用者需求等面向進行文獻分析與系統規劃，建置「歷代寶案脈絡分析系統」，目標在於建立一個橫跨 400 餘年的全文檢索系統，重新建構《歷代寶案》在原始書冊之外的時間、空間、人物、官職、物品等多重脈絡，提供使用者鳥瞰《歷代寶案》全貌的環境以及重新解構與呈現的系統，系統規劃建置需求分析與系統特點如下：

一、文獻特性與使用者需求分析

數位人文系統的建置目的在於建置一個好的「使用環境」，而一個好的使用環境不僅在於符合研究者與公眾的需求，更重要的是分析「文獻特性」，並針對文件的特性進行系統功能建置。因此，在系統規劃上就《歷代寶案》的特性分析如下：

(一) 時間範圍：橫跨 444 年，以中國年號紀年

年代跨距範圍長，除琉球國歷任國王外，涉及中國（明朝、南明、清朝）、朝鮮、暹羅（今泰國）等東亞各國朝代歷史。

(二) 資料特性：外交文書，包含外交官職、物品名稱與數量單位

由閩人三十六姓所掌管的琉球國外交文書，內容以琉球國向中國的朝貢事務為主，因此外交官員和朝貢物品屬於本文書的重要資訊。

(三) 區域特性：內容包含大東亞地區，不限於單一國家內的小地名，但與中國沿海城市息息相關

琉球王國仰賴與周邊國家進行朝貢貿易以獲取資源，除了中國之外，與日本、朝鮮，甚至東亞各國如暹羅、佛大尼（今泰國北大年府）均有所往來，透過物品的交換獲取利益。因此本文書所涉及之地名，琉球本國甚少，以外國佔大多數，且《歷代寶案》的重點在於與中國的文書往來，因而涉及中國沿海地區的重要城市。

（四）數位檔案特性：全文的文獻與特殊詞彙詮釋資料

本系統所應用的數位化文獻為全文打字標點的全文內容，除搭配書冊影像之外，透過特殊詞彙的擷取而有詮釋資料內容，因此系統建置規劃過程中全文與詮釋資料均為系統研發與應用的重要範疇。

二、多面向的鳥瞰文獻全貌

透過上述的文獻特性與使用者需求分析，歷代寶案系統首先提供使用者一個觀察文獻全貌鳥瞰的環境，透過「瀏覽」功能，提供使用者慣用的原始編輯與出版型態的脈絡瀏覽，並提供一個以年代為脈絡的瀏覽方式。

（一）原始編輯「集／卷」瀏覽模式

依照《歷代寶案》的原始編輯，提供「集／卷」瀏覽模式，將《歷代寶案》第一、二、三集、別集、咨集提供「集」及「卷」兩階層的瀏覽方式。

5 集，248 卷，4293 件。	
第一集 +	(1032) 1111
第二集 +	(3079) 1111
第三集 -	(163) 1111
卷一	(14) 1111
卷二	(4) 1111
卷三	(21) 1111
卷四	(5) 1111
卷五	(9) 1111
卷六	(10) 1111
卷七	(9) 1111
卷八	(42) 1111
卷九	(4) 1111
卷一〇	(21) 1111
卷一一	(15) 1111
卷一二	(2) 1111
卷一三	(7) 1111
別集 -	(14) 1111
倭英情狀	(13) 1111
冠船之時唐人持來品或物錄	(1) 1111
咨集 -	(5) 1111
文組方	(5) 1111

圖 5 原始編輯「集／卷」瀏覽模式

(二) 出版型態「收錄書冊」瀏覽模式

國立臺灣大學圖書館出版的《歷代寶案》「書冊」模式，為近數十年多數學者參考引用的版本，同時缺失的部分以沖繩縣教育委員會出版的《歷代寶案：校訂本》補足，內容在「收錄書冊」瀏覽模式中呈現。

臺大圖書館出版，《歷代寶案》		
15 冊，4262 件。		
ntu-1088321	第1冊	(425) 冊
ntu-1088322	第2冊	(583) 冊
ntu-1088353	第3冊	(275) 冊
ntu-1088309	第4冊	(307) 冊
ntu-1088325	第5冊	(332) 冊
ntu-1088311	第6冊	(327) 冊
ntu-1088327	第7冊	(253) 冊
ntu-1088313	第8冊	(264) 冊
ntu-1088329	第9冊	(174) 冊
ntu-1088315	第10冊	(210) 冊
ntu-1088316	第11冊	(248) 冊
ntu-1088317	第12冊	(233) 冊
ntu-1088333	第13冊	(249) 冊
ntu-1088334	第14冊	(200) 冊
ntu-1088320	第15冊	(182) 冊
沖繩縣教育委員會出版，《歷代寶案校訂本》		
3 冊，31 件。		
ntu-1115751	第2冊	(24) 冊
ntu-3642149	第10冊	(5) 冊
ntu-1769586	第13冊	(2) 冊

圖 6 出版型態「收錄書冊」瀏覽模式

(三) 時間的鳥瞰模式

包含了「中國朝代年份」及「西元」兩種模式。透過年代正規化將中曆與西曆進行標準化，提供使用者中國朝代年號與西元年月日兩種瀏覽模式。

瀏覽			
年份	集/卷	西元	收錄書冊編號
4293 件。			
明 +	(721) 冊		
南明 -	(25) 冊		
弘光 -	(7) 冊		
1年 (西元1646年)	(7) 冊		
隆武 +	(18) 冊		
清 -	(3491) 冊		
順治 +	(24) 冊		
康熙 +	(456) 冊		
雍正 +	(181) 冊		
乾隆 +	(970) 冊		
嘉慶 +	(599) 冊		
道光 +	(906) 冊		
咸豐 +	(252) 冊		
同治 +	(103) 冊		

瀏覽			
年份	集/卷	西元	收錄書冊編號
1424~1867, 4293 件。			
1643 (明崇禎16年) +			(2) 冊
1644 (明崇禎17年) +			(15) 冊
1645 (南明弘光1年,南明隆武1年) +			(11) 冊
1646 (南明隆武2年) +			(13) 冊
1647 (清順治4年) +			(2) 冊
1648 (清順治5年) +			(1) 冊
1649 (南明隆武5年,清順治6年) -			(7) 冊
1649-03 +			(1) 冊
1649-12 +			(6) 冊
1651 (清順治8年) +			(1) 冊
1653 (清順治10年) +			(7) 冊
1654 (清順治11年) +			(5) 冊
1656 (清順治13年) +			(2) 冊

圖 7 中國朝代年號瀏覽模式 (左) 和西元年月日瀏覽模式 (右)

另外，本系統同時將所有數位化文獻以時間軸的方式呈現，將各年累加的資料以柱狀圖呈現，並標示中國朝代更迭的分界。

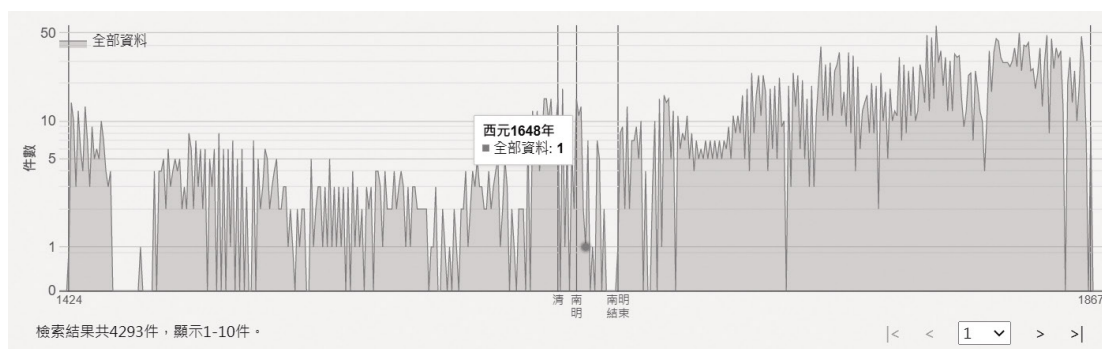


圖 8 時間軸瀏覽模式

(四) 空間的鳥瞰模式

透過地名的探勘，將數位化文獻內文提及的地點以 GIS 的空間概念進行瀏覽，以視覺化的形式對文獻內文具有初步概略的理解。

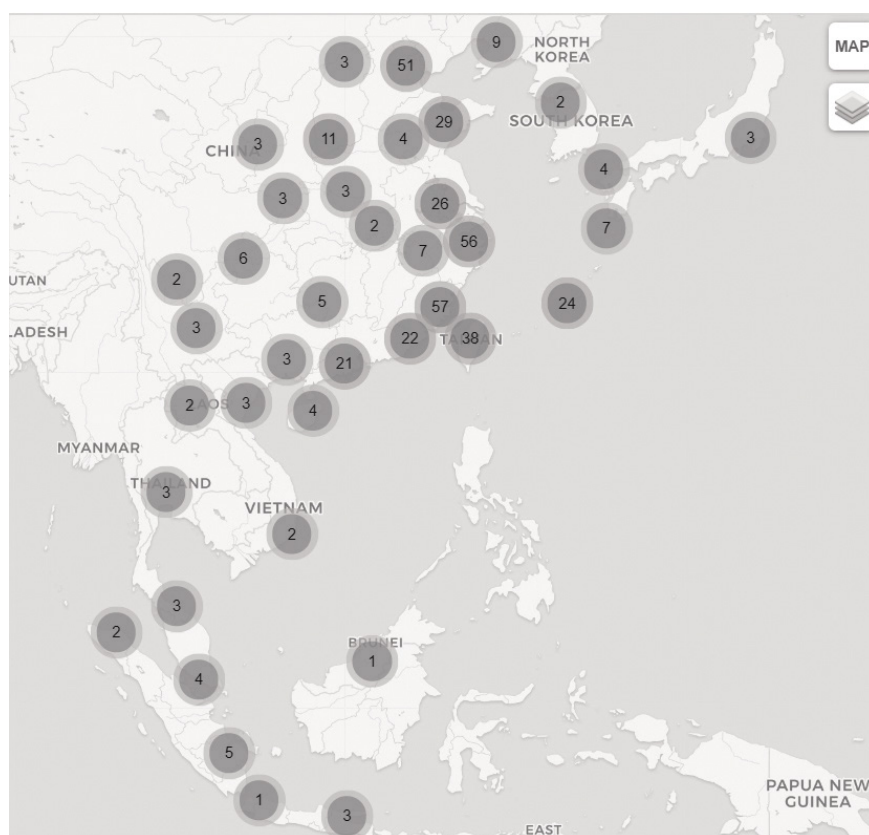


圖 9 地點瀏覽模式

三、多重脈絡的探勘與分析

數位化文獻透過脈絡的探勘與分析後，探勘「時間」、「官名」、「人名」、「地名」、「物品」等特殊詞彙，進行多重的分析與呈現，主要採用「後分類」與「文件頻率」提供使用者針對檢索結果進行觀察與分析的操作環境。

(一) 後分類的多重脈絡分析

檢索結果的後分類意指針對搜尋的結果進行再分類，甚至是後分類的後分類，即是對於檢索的結果透過後分類操作進行觀察與分析。本系統以年份（中國朝代）、西元年份、月份、官名、人名、地名、物品等進行檢索結果的再分類。

以關鍵詞「胡椒」為例，針對年份中明代的 168 件，可再進行西元年、月份、官名、人名、地名及物品的檢索結果分類統計，透過分類檢視的操作與分析，可直觀式的發現胡椒的檢索結果中，物品數量第二高者為蘇木。

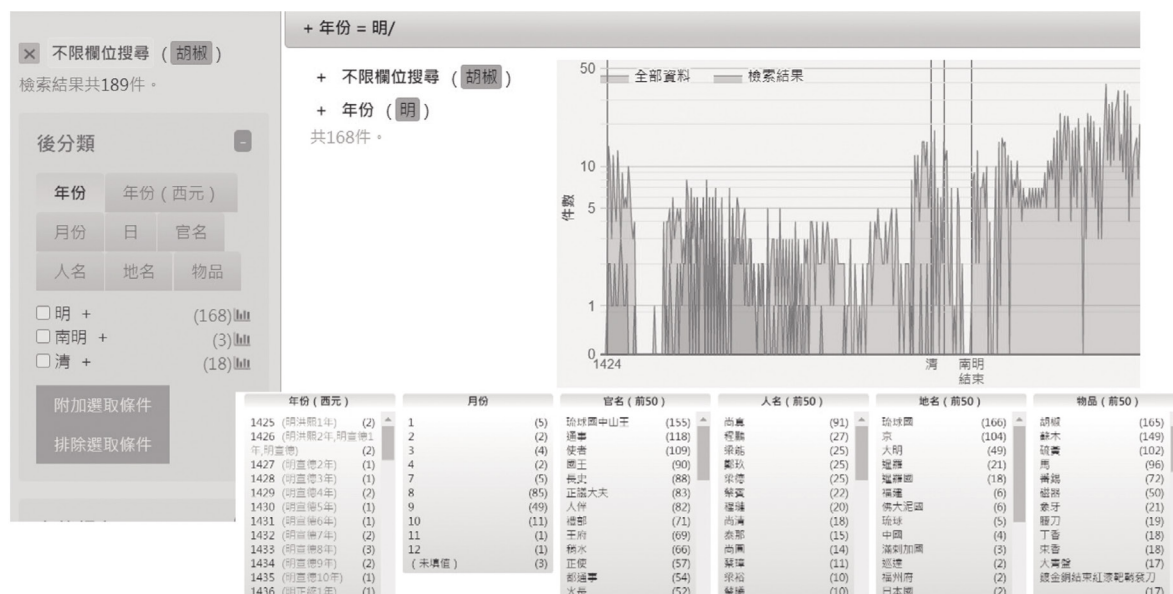


圖 10 後分類的多重脈絡分析

(二) 數位化文獻的關係再構—文件頻率

文件頻率則提供另一種觀察檢索結果的方式，除了檢索結果的數量統計之外，同時透過文本與文本間的運算分析，計算「相關度」。本系統相關度的計算方式，以束香為例，被歸為後分類物品的束香有 32 件；而在關鍵詞「胡椒」及年份「明代」的條件下，束香有 18 件， $18 \div 32 \approx 0.56$ ，0.56 即為束香與胡椒的相關度。此外，以關鍵詞

「胡椒」的檢索結果為例，「束香」、「檀香」的文件數同為 18 件，但相關度分別為 0.56 與 0.82，可見在這批文本中，於關鍵詞「胡椒」及年份「明代」的條件下，檀香與胡椒的相關程度高於束香。

文件頻率			
官名	人名	地名	物品
		文件數	相關度
詞彙 t		20	0.80
束香		18	0.56
檀香		18	0.82
鍍金銅結束紅漆靶鞘袋刀		17	0.4
大青盤		17	0.49
硫磺		15	0.07
小青盤		15	0.39
小青碗		14	0.67
金靶鞘腰刀		13	0.15
紅銅		13	0.02
摺紙扇		12	0.92
銀粉匣		12	0.2
銀靶鞘腰刀		11	0.13
金粉匣		11	0.19
金彩畫屏風		10	0.91
紅苳		10	0.4

圖 11 文件頻率

四、檢索結果的兩層式呈現

本系統針對檢索結果的呈現，提供「一般檢視」、「地點檢視」與「統計檢視」三種模式，每種模式均提供數位化文獻第一層的簡易檢視模式以及第二層的詳細結果呈現。

(一) 一般檢視：針對檢索結果提供簡要的「題名」、「集／卷」、「日期」，並就檢索的詞彙於全文中進行「標亮」和顯示所在的特殊欄位

檢索結果共189件，顯示1-10件。			<<1>>	
			每頁10條，排序依集/卷	
1.	皇帝敕諭琉球國中山王尚圖			第一集/卷一
	日期 明成化11年4月20日(1475-05-24)			
	物品 胡椒, 糖, 雞			
	王正貴外，許照例附搭胡椒等物，其餘正副使人等，不許夾帶私貨前來買賣，及在途生事，擾害平民，打擾官府，有累國王忠順之意。王其省之，省之。故諭。 廣德 成化十一年四月二十日 之寶			
2.	〔禮部〕為進〔貢事〕			第一集/卷六
	日期 清順治11年6月15日(1654-07-28)			
	物品 土夏布, 檀香, 硫磺, 胡椒, 蘇木, 犀角, 象牙, 綢布, 綢生綢, 馬刀, 金銀酒海, 海巴, 擺子扇, 錫, 生熟夏布, 牛皮, 降香, 本香, 黃熟香, 烏木, 磨刀石, 王詔, 生綢			
	熟香、蘇木、烏木、胡椒、硫磺、磨刀石在案。該臣等謹得：琉球進貢方〔物數〕目及貳年貢貢，俱應照會典例，移〔咨〕琉球國中山王，永為定例，欽遵施行等因。〔順治〕拾壹年參月貳拾捌日題，肆月初壹〔日〕奉〔聖旨〕：是。依。欽遵。抄部送司。案呈到部。〔擬合就〕行。			
3.	禮部為進貢事			第一集/卷六
	日期 清康熙5年8月17日(1666-09-15)			
	物品 瑪瑙, 烏木, 降香, 刀, 金銀酒海, 金銀粉匣, 象牙, 海巴, 擺子扇, 泥金扇, 生紅銅, 錫, 生熟夏布, 牛皮, 木香, 達香, 丁香, 檀香, 黃熟香, 蘇木, 胡椒, 硫磺, 磨刀石, 馬鞍, 腰刀, 紅銅, 馬鞍, 馬匹			
	黃熟香、蘇木、烏木、胡椒、硫磺、磨刀石。該部謹得：琉球進貢方物數目及貳年貢貢，俱應照會典例，移咨該國中山王永為定例等因。奉聖旨：是。依議行。欽遵。移咨該國王在案。今該〔王〕稱係彈丸小國。生物至〔少〕，其瑪瑙、烏木等拾〔件〕，原產于交〔趾〕等處，難以具貢。其貳年			

圖 12 一般檢視

(二) 地圖檢視：針對檢索結果在地圖上進行標示，並提供座標分布與選單，可進行排除與附加選取條件

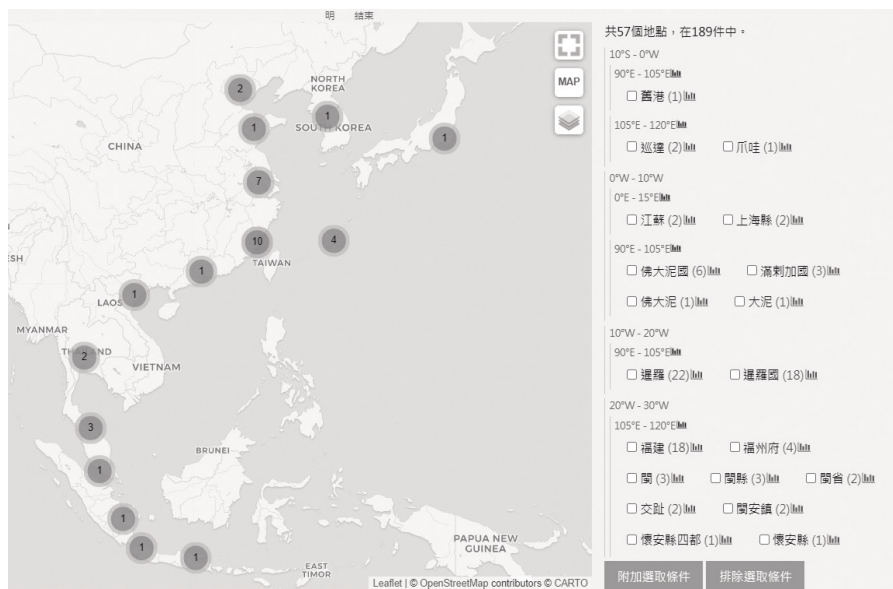


圖 13 地圖檢視

(三) 統計檢視：針對檢索結果提供「橫向欄位」和「縱向欄位」進行表格製作，並提供「下載表格」功能，可下載 CSV 檔的表單（可用 EXCEL 開啟）

橫向欄位： 縱向欄位： 製作表格

*選擇兩個欄位且資料量過大時網頁會當掉，請先檢索縮小範圍後再進行統計。

檢索條件：不限欄位搜尋 (胡椒)
檢索結果共189件。
橫向欄位：地名
縱向欄位：物品

下載表格

	琉球國	京	大明	暹羅	福建	暹羅國	琉球	佛大泥國	那霸港	中國	福州府	滿刺加國	王城	閩縣	閩	江蘇省城	浙	閩安鎮	山東
胡椒	185	117	49	22	17	17	8	6	4	4	4	3	3	3	3	2	2	2	2
蘇木	159	96	47	22	10	17	7	6	0	3	1	3	3	1	1	0	0	0	0
硫黃	103	90	9	0	3	6	2	0	0	0	1	0	0	1	1	0	0	0	0
馬	99	93	4	1	5	0	3	0	0	0	1	0	0	1	1	0	0	0	0
番錫	71	68	4	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
磁器	50	1	42	21	1	13	3	6	0	3	1	3	0	1	0	0	0	0	0
象牙	23	20	3	1	3	0	4	0	0	0	0	0	0	0	1	0	0	0	0
腰刀	21	4	8	1	2	13	2	0	0	2	0	0	0	0	1	0	0	0	0
丁香	18	18	2	1	4	1	3	0	0	0	0	0	0	0	1	0	0	0	0
束香	17	17	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0
檀香	17	14	1	1	3	0	4	0	0	0	0	0	0	0	1	0	0	0	0

圖 14 統計檢視

(四) 詳目頁檢視：不論在一般檢視、地圖檢視或統計檢視，點選檢索結果均會以另開分頁的形式進入第二層的詳目內容

詳目頁各項功能說明如下：

可調式的雙面板呈現：共有「詮釋資料」、「全文」、「圖片」以及「地圖」四種內容，以勾選的方式，提供使用者檢視操作。



圖 15 可調式雙面板詳目呈現

全文的標籤顯示及直書橫書模式：全文內容除了可以放大、縮小顯示之外，並可調整直書、橫書模式與圖片對照查看內容。

另外，亦可將特殊詞彙，包含人名、地名、官名、物品、時間等進行不同顏色的標籤顯示與取消顯示，提供使用者操作檢視。

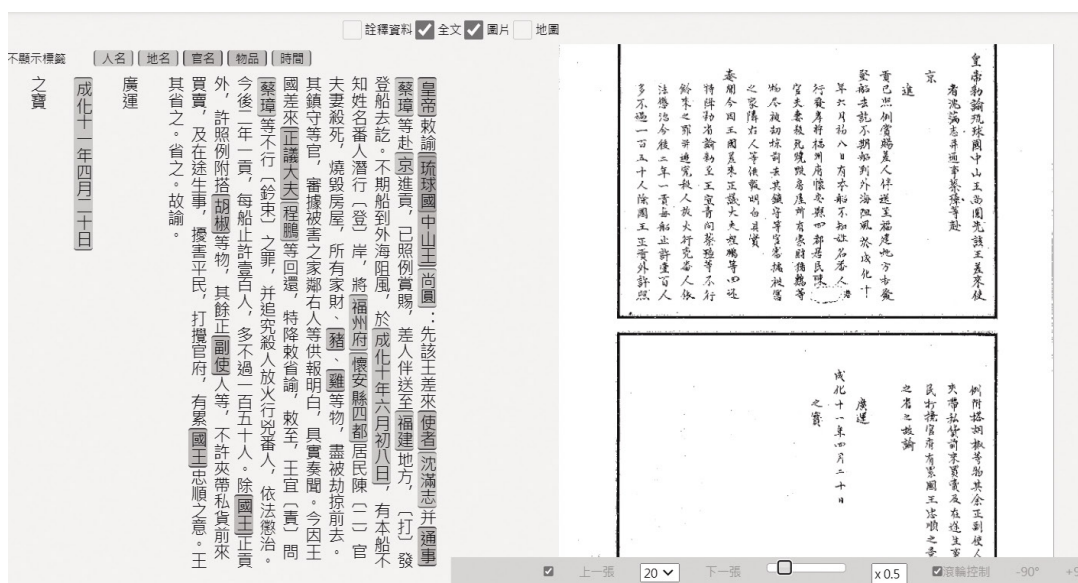


圖 16 全文標籤顯示與直書、橫書模式

目前歷代寶案脈絡系統 (<http://lidaibaoan.digital.ntu.edu.tw/>) 已建置完成，期待各界研究者參與使用。

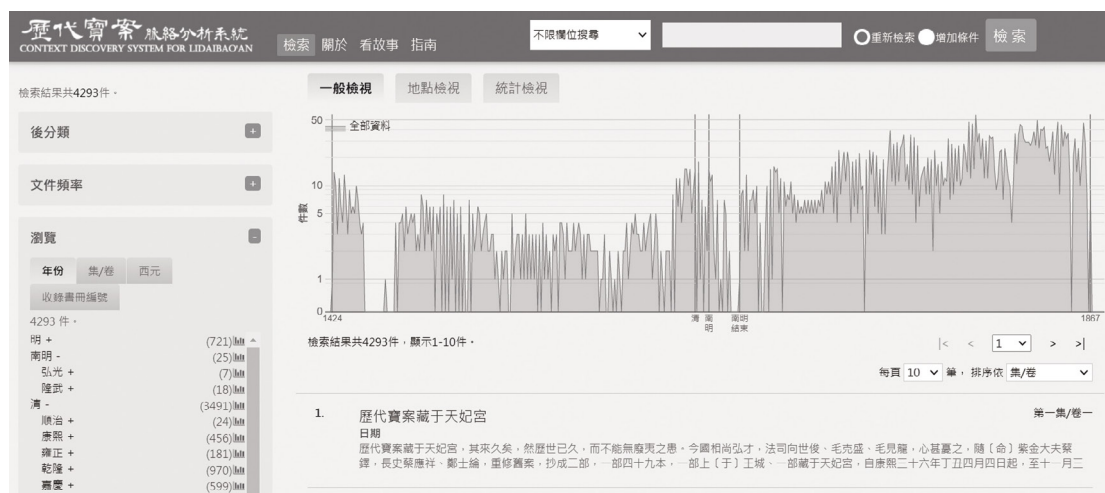


圖 17 歷代寶案脈絡系統首頁

除上述所提供的檢索結果脈絡分析功能之外，本系統亦建置「關於系統」、「看故事」以及「指南」等網頁，提供系統收錄資料的說明、朝代對照參考及對於歷代寶案故事書寫的應用範例，提供除了文獻研究外，大眾書寫的應用實例以供各界參考。

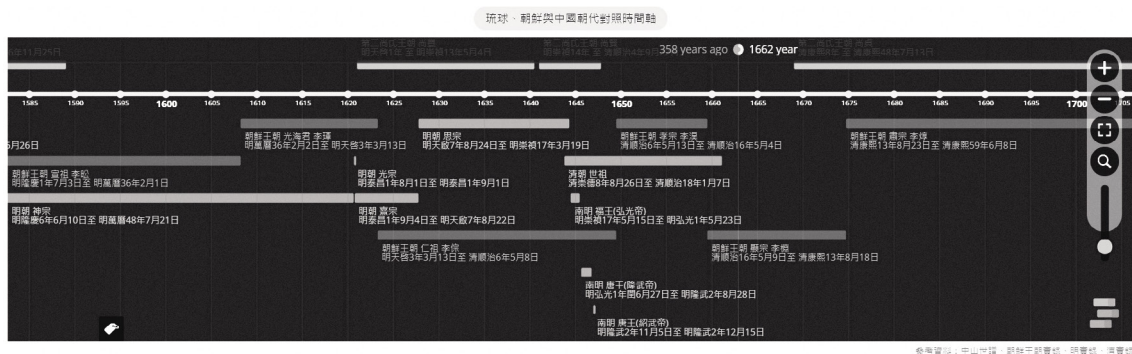


圖 18 系統頁面：琉球、朝鮮與中國朝代對照時間軸



圖 19 看故事：歷代寶案故事集

肆、系統應用實作範例

《歷代寶案》作為外交性質濃厚的文書，其空間範圍涉及今日中國、韓國、日本、越南、泰國、馬來西亞、印尼等國家，資料內容提及的大小地名更是多達 2,000 個。

基於臺灣和琉球同樣為西太平洋東亞島弧一環且具有海洋文化的特性，故筆者以臺灣作為範例，透過歷代寶案系統檢索查詢，探討臺灣是以何種面貌出現於《歷代寶案》之中，以怎樣的外交型態和琉球國產生互動，以及當時的臺灣在東亞海域扮演什麼樣的角色。

以下藉由系統中「關鍵字查詢」與「後分類瀏覽」兩項檢索查詢的運用，配合系統相關功能的操作，以關鍵詞「臺灣」為例，說明如何利用歷代寶案系統展開海洋史相關研究的輔助與應用。

一、關鍵詞「臺灣」的操作

以「臺灣」做為關鍵詞進行資料檢索，合計有 165 筆資料，⁹ 大部分的內容和琉球國人風飄到臺灣附近海域的海難事件相關。¹⁰ 透過後分類瀏覽的功能，筆者觀察到歷代寶案系統中與「臺灣」有關的資料有幾點特色，分別說明如下：

(一) 年份

主要分布於清代，最早出現的時間為 1721 年（康熙 60 年），最晚為 1863 年（同治 2 年）。以西曆來看，1804 年 10 筆最多，1809 年 9 筆次之；以中曆來看，嘉慶 9 年 9 筆最多，嘉慶 13 年 7 筆次之，其筆數落差與中西曆換算有關。此兩年筆數較多的原因並非海難事件特別多，而是海難對象的層級不同，為琉球國的貢船與朝貢人員，其牽涉議題大致分為人員和物品兩個部分。人員方面，清朝如何提供眾多的受難者急難救濟措施、是否有大小適宜的船隻可供搭載返國、清朝是否追究責任與琉球國是否懲處琉球貢使；物品方面，則涉及朝貢物品是否遺失、琉球國是否需要補貢，自然引起清朝和琉球國雙方的高度重視。

9. 本系統已整合正異體字，使用「臺灣」或「台灣」查詢，皆可得到相同筆數的檢索結果。以不限欄位搜尋，可得 167 筆檢索結果，扣除 2 筆指涉「南臺灣泊」，計有 165 筆。

10. 海難議題研究成果豐碩，琉球人風飄臺灣之相關研究成果如楊彥杰，〈清代臺灣撫卹琉球遭風難民的案例分析〉，《第七屆中琉歷史關係國際學術會議中琉歷史關係論文集》（臺北：中琉文化經濟協會，1999），頁 655-658；徐恭生，〈清代海上漂風難民拯濟制度的確立及演變〉，《第八回琉中關係國際學術會議論文集》（沖繩：琉球中國關係學術會議，2001），頁 60-72。

明代查無結果乃因當時的臺灣被稱為「雞籠」，¹¹ 本系統中僅有 3 筆。內容主要和 1609 年（明萬曆 37 年）「慶長琉球之役」有關，琉球國向明朝報告薩摩藩入侵之事件經過，以及因戰爭致使貢期延遲。而「雞籠」被提及是因為日本有進攻雞籠的計畫，其戰略地位亦牽涉到中國閩南沿海一帶的安危，琉球國提醒明朝須於海防上加強警戒。¹² 此外，14 至 16 世紀這段時期的琉球國因對外貿易繁盛而被稱為「大交易時代」，由明代臺灣與琉球國幾乎毫無關聯來看，推論臺灣並不屬於琉球國海洋貿易網絡的一環。

（二）月份

各月份皆有，以農曆八月份（西曆九月份）36 筆最多，農曆五月份（西曆六月份）26 筆次之，似與夏、秋二季為颱風盛行季節，導致船隻易發生事故相互呼應。然而本系統的月份是擷取公文的成文日期，與海難發生的時間或許有些許差距，詳情請見後續探討。

機械動力廣泛應用於船隻以前，海上交通工具以帆船為主，其動力依靠風力和洋流，並且搭配準確的針路（羅盤），以防迷失方向。因此，福建往琉球以夏季之西南風和黑潮為佳，風過盛船隻易飄至沖繩群島以北各島，琉球往福建以冬季之東北風為佳，風過盛船隻易飄到廣東或南洋各地。¹³

（三）官名

在後分類瀏覽近 600 個官名當中，以文件頻率之文件數前 20 高的官名來看，文件數最多的福建等處承宣布政使司、琉球國中山王為公文行文主要的收、發文對象。

清朝方面，以直接涉及處理難民相關事務的地方機關或首長及其上級機關或首長為多，與清代奏報制度、文武官職負責職務有關。文官部分，以直屬機關或首長由下到上的層級為例，臺灣府知府—福建等處承宣布政使司—福建巡撫、閩浙總督皆為常見官名。另有福防廳／福防同知／福州府海防同知協助福州府管理海防事務。武官部分，由

11. 明代以降，「臺灣」漸以雞籠山、北港、東番、臺灣而定名。參見臺灣銀行經濟研究室編，《流求與雞籠山》（臺北：臺灣銀行經濟研究室，1964）。

12. 〈琉球國中山王府攝王妃馬氏王弟尚宏暫看國事法司馬良弼為飛報事〉，明萬曆 38 年 1 月 20 日（1610-02-13），編號 pdsg-ntu-1088321-00349，歷代寶案脈絡分析系統。〈琉球國中山王府攝王妃王弟尚宏暫看掌國事法司馬良弼為懇乞天恩恤憐遭亂贖修貢職事〉，明萬曆 38 年 1 月 30 日（1610-02-23），編號 pdsg-ntu-1088321-00350，歷代寶案脈絡分析系統。〈琉球國中山王尚寧為急報倭亂致緩貢期事〉，明萬曆 37 年 5 月（1609-02-04~1610-01-24），編號 pdsg-ntu-1088321-00348，歷代寶案脈絡分析系統。

13. 梁嘉彬，〈明清兩代福州那霸間風帆去程之研究〉，《第一屆中琉歷史關係國際學術會議論文集》（臺北：聯合報文化基金會國學文獻館，1987），頁 335-336。

閩安協相關人員通報該管轄範圍的船難事件或撥兵護送琉球國船隻出洋。琉球國方面，以處理難民接回返國事宜的朝貢團成員為多，有航海人員如水梢、水手，外交人員如通事、都通事、貢使。

藉由文件頻率的相關度低可得知，這些具有外交性質、航海技術相關的機關、官職或職務，以及福建一帶居中負責藩屬國相關事務的機關或官職，同樣也頻繁出現在歷代寶案系統的其他議題中。惟臺灣各級地方官職，如臺灣道、淡水同知、臺灣府知府、噶瑪蘭通判、鳳山縣知縣等，在後分類中筆數少，相關度卻極高，可見《歷代寶案》中的臺灣各地行政長官或官府大多是琉球國人在臺灣附近海域遭難後第一個官方介入的救助者。

文件頻率			
官名	人名	地名	物品
詞彙 t		文件數	相關度
福建等處承宣布政使司		146	0.07
琉球國中山王		118	0.05
水梢		106	0.11
福防同知		103	0.25
巡撫部院		96	0.22
國王		96	0.09
閩安協		92	0.24
總督部堂		87	0.3
禮部		78	0.04
官伴		77	0.12
皇上		76	0.07
都通事		72	0.04
督部堂		71	0.28
福防廳		69	0.23
閩浙總督		64	0.22

圖 20 後分類官名之文件頻率

(四) 人名

在後分類瀏覽 1,500 餘個人名當中，由名字的組合方式可知道以琉球國平民佔大多數，如宮平、比嘉、金城，此為只有名字、沒有姓氏的童名，童名對琉球國百姓而言即是正式的名字。部分人名橫跨的時間向度過長，為不同時期的同名者。此外，同一時間向度亦有可能出現同名者，需由文本中是否記載出身地作為判斷依據。

相較於此，琉球國士族亦有童名，使用場合為私領域，公領域則使用家名、名乘、唐名、家號（和名），其中唐名用於與中國的外交場合，和名用於與日本的外交場合。¹⁴ 本文書的外交對象主要為明、清朝，出現的外交人員如孫光裕、梁文翼即為唐名。

14. 呂青華，〈琉球久米村人的民族學研究〉（臺北：國立政治大學民族研究所博士論文，2007），頁 93-94。

文件頻率			
官名	人名	地名	物品
詞彙 t		文件數	相關度
宮平		25	0.49
仲村渠		23	0.4
島袋		22	0.35
金城		21	0.23
宮城		21	0.29
比嘉		21	0.25
內間		20	0.5
國吉		19	0.54
安里		18	0.4
宮里		16	0.44
大城		16	0.25
玉城		16	0.24
知念		15	0.28
尚瀨		15	0.11
言尚焜		14	0.38

圖 21 後分類人名之文件頻率

(五) 地名

在後分類瀏覽 500 餘個地名當中，以文件頻率之文件數前 20 高的地名來看，除了本文書的產生國家琉球國／琉球最為常見，琉球國各地以那霸府出現次數最多，因其為沖繩本島主要港口所在地，且那霸官掌管錢穀，差該地人前往各島催納賦稅。¹⁵

其餘以中國地名居多。福建／閩／閩省是中國與琉球船行來往的重要省會，其中閩江下游近海口處的閩安鎮是出入福州城的海上交通要道，¹⁶ 而明清時期接待琉球人的柔遠驛位於福州府，京（北京）是琉球朝貢團入貢的目的地。此為琉球人最容易與中國產生連結的地名。另外在 1858 年（咸豐 8 年）臺灣開港以前，受限於對渡政策，廈門、臺灣府則是清代臺灣與中國通航的對渡口岸之一。¹⁷

此外，臺灣相關的各層級行政區或地名，如臺灣府、鳳山縣、淡水、噶瑪蘭等，文件數不一定多，但文件頻率的相關度數值高，可得知絕大部分僅出現在本文所檢索的關鍵詞臺灣之中。

15. 徐葆光，《中山傳信錄》卷五，中國哲學書電子化計劃，<https://ctext.org/wiki.pl?if=gb&chapter=825749>，2021 年 2 月 27 日瀏覽。

16. 何孟興，〈明代福州海防要地「竿塘山」之研究（1368-1456 年）〉，《止善》7（2009.07），頁 8-9。

17. 「對渡口岸」，臺灣大百科全書，<https://nrch.culture.tw/twpedia.aspx?id=3634>，2021 年 2 月 28 日瀏覽。



官名	人名	地名	物品		文件數	相關度
詞彙 t						
		琉球國			157	0.04
		福建			146	0.06
		臺灣			139	1
		閩			136	0.11
		琉球			78	0.08
		臺灣府			68	1
		閩省			66	0.14
		京			65	0.03
		那霸府			63	0.32
		浙江			50	0.17
		福州府			42	0.2
		淡水			40	1
		浙			36	0.2
		閩安鎮			35	0.22
		鳳山縣			35	1

圖 22 後分類地名之文件頻率

(六) 物品

在後分類瀏覽 500 餘個物品當中，以文件頻率之文件數前 20 高的物品來看，以急難救助和例禁收買兩類物品為主。清朝官方施予琉球國難民的急難救助物資和金錢所佔文件數最多，如行糧、米／口糧、鹽菜銀、衣服／布棉、茶葉等，其根據為《大清會典》。例禁收買則是所佔詞彙數量最多，1644 年（順治元年）將史書、黑黃紫皂大花西番蓮緞匹、焰硝、牛角、兵器、黃紅銅器列為禁止外國貿易收買之物。¹⁸ 1731 年（雍正 9 年）根據廣東布政使奏稱外國船隻大量出口粵東出產的鐵鍋，應照廢鐵之例嚴禁漢船、夷船買賣出洋，故增加鐵鍋一項。¹⁹ 另有桐油、緞匹、綾絹、紗羅屬於例禁不許收買的物品。

由於這些物品亦會出現在其他國人海漂中國的急難救助、柔遠驛開館貿易等議題中，故文件頻率之相關度並不高。

18. DQHDZL-094-050_030_01b0_030_00-004，禮部 / 主客清吏司 / 朝貢下 / 禁令 /，乾隆朝大清會典與則例對照檢索系統，臺灣大學數位人文研究中心。

19. DQHDZL-024-030_020_3500_010_00-020，吏部 / 考功清吏司 / 海防 / 海防 /，乾隆朝大清會典與則例對照檢索系統，臺灣大學數位人文研究中心。

文件頻率			
官名	人名	地名	物品
詞彙 t		文件數	相關度
行糧		92	0.32
米		84	0.27
鹽菜銀		82	0.31
口糧		56	0.25
銀兩		53	0.19
布匹		51	0.3
鐵鍋		44	0.3
衣服		38	0.21
藥材		37	0.27
茶葉		36	0.22
紅銅		33	0.05
桐油		33	0.29
牛角		33	0.29
焰硝		33	0.29
兵器		33	0.29

圖 23 後分類物品之文件頻率

二、海難事件案例分析：以琉球人風飄臺灣為例

觀察完歷代寶案系統所呈現的大數據脈絡分析後，回歸文本內容分析，先自 165 筆資料中，去除 30 筆無關琉球人風飄臺灣的議題，再從 135 筆資料，整理出 46 件案例（詳見附表 1）²⁰。文本中大多會記載船隻出航時間和事由、航行起訖地、船型、乘船人數和遭難死亡人數、遭風和飄臺時間、飄臺地區、遭風與飄臺的海難過程、清代官民的救濟措施、臺灣與中國之間對渡的出船時間、到省和安置館驛時間、返國之離驛登舟時間。以下筆者將海難事件依照時序分成三個部分探討。

（一）海難前

整體來看，這些琉球船隻大多是在琉球國所屬的三十六島範圍內活動而遭遇海難的，再細分出船原因和船程遠近的統計分別如下：以船行事由來看，賦稅 21 件、經濟活動 15 件、賦稅兼經濟活動 4 件、朝貢 3 件、民生 1 件、公務來往 1 件、不詳 1 件。以船程遠近來看，琉球三十六島的來往 39 件、長程 4 件、近海沿岸 2 件、不詳 1 件。

琉球國的稅收採行實物制，各島嶼稅物需透過船隻運送至中山王府交納。以《中山傳信錄》的記載來看文本中常出現的琉球各島嶼之盛產物產，宮古島（昔稱太平山、麻姑山）頗富饒，產穀類和畜類，畜類以牛、馬為多，又出產棉布、麻布、草蓆、紅酒。久米島（昔稱姑米山），產穀類、畜類、魚類和貝類，畜類有雞、豬、牛、馬，又出產土綿、繭紬、白紙、蠟燭。八重山島，產五穀、畜類、林木、海產，畜類有牛、馬，

20. 另有 3 件琉球人風飄至臺灣的案例，僅記載受難人數和受難地，缺乏其餘資訊，故不列入本文探討對象。

林木有檳木、黑木、黃木、赤木，海產有螺石、玳瑁、珊瑚等，又出產麻布、棉布、紅酒、草蓆。²¹ 雖然各島盛產穀類、加工品、山林資源、海洋資源等多面向物產，但以物產的普遍性與保存性來看，稅物以穀類最多，於 25 件案例裡佔提及稅物總次數的 67%，亦有零星次數提及山林資源如柴木，加工品如草蓆、棉布、芭蕉線。由於賦稅為每年定期性的常態活動，故海難事件之事由以琉球三十六島間的繳納賦稅、催納賦稅所佔最多。此外，為善加利用船隻空間和減少空船狀態，賦稅去回程、催納賦稅過程，船上亦載有民生必需品、嗜好物等以進行貿易活動，或是搭載乘客。

（二）海難過程

海難可再細分為兩個過程，先是遭遇大風致使船隻毀損或是漂離航道，後是漂至臺灣附近觸岸或擱淺，致使船隻損壞更嚴重。

以遭風詞彙來看，「遭風」22 件、「遇颶風」10 件、「遇暴風」5 件、「遇狂風」4 件、「遇風浪」2 件、「遇大風」1 件、未記錄 2 件。過程中，大多造成船體的桅、舵損壞，其次是船員砍斷桅篷以降低船隻翻覆。由於船體本身受損或是操控船隻航向設備毀損，因而將船上的貨物、稅物等具有重量的物品拋至洋中，減輕船隻重量，以利隨風漂流。

根據朱瑪瓏的研究，颱風在二十五史和地方志裡的用詞為「颶風」，亦常見「大風」、「大風水」，以及較少見的「暴風」，但「大風」與一般意義下的大風需探究地點和季節以示區辨。²² 筆者將文本的遭風詞彙與出航月份整理成表 1，可得知船隻遭風遇難的季節以冬季最高，春季次之，夏、秋季較少；而由文本記錄的船隻破壞程度、偏離航道等遭風過程，推測夏、秋二季的颶風和暴風指涉颱風的機率較高，風則以一般的瞬間大風居多，其他詞彙數量太少，判斷較困難。值得注意的是，這些遭風詞彙是透過翻譯記錄下來，雖然琉球國屬於東亞文化圈的一環，但琉球國與中國對此用詞是否一致，則有待進一步的比較和探討。經由進一步的分析，海難發生的時間雖然與本系統後分類呈現的月份有所不同，與原先推測有落差，但發生在夏、秋二季的海難案例，確實由颱風導致的機率較高。

21. 徐葆光，《中山傳信錄》卷四，中國哲學書電子化計劃，<https://ctext.org/wiki.pl?if=gb&chapter=237251>，2021 年 2 月 22 日瀏覽。

22. 朱瑪瓏，〈近代颱風知識的轉變——以臺灣為中心的探討〉（臺北：國立臺灣大學歷史學系碩士論文，2000）頁 27。

表 1 遭風詞彙與出航月份統計表

農曆月份／ 描述詞彙	遭風	遇颶風	遇暴風	遇狂風	遇風浪	遇大風
一月	3	—	1	—	—	—
二月	—	1	—	—	—	—
三月	4	—	—	—	—	—
四月	4	—	1	2	—	—
五月	—	—	—	—	—	—
六月	—	—	—	—	1	—
七月	—	—	—	—	—	—
八月	1	1	—	—	—	—
九月	3	1	1	—	—	1
十月	3	5	1	1	1	—
十一月	3	1	1	—	—	—
十二月	1	1	—	1	—	—

資料來源：根據歷代寶案脈絡分析系統關鍵字「臺灣」135 筆資料整理而成。

再以船型來看，海船 10 件（其中 1 件帶有杉板小船）、小海船 9 件（其中 1 件帶有杉板小船、1 件帶有小腳船）、船 6 件、貢船 4 件、小船 3 件、小舟 1 件、獨木小船 2 件、不詳 11 件（其中 1 件帶有杉板小船）。如前所述，船隻以琉球三十六島間的往來居多，故未記載者可能為海船或小海船。至於船隻大小，由於文本中僅一件有記錄：「船一隻，船身長 8 丈 5 尺（27.2 公尺）、樑頭闊 2 丈 2 尺 5 寸（7.2 公尺）」²³，在歷代寶案系統中以「船身長」做搜尋觀察相似用途的船隻大小，海船長 4 丈 5 尺至 9 丈 6 尺（14.4 公尺 -30.72 公尺），闊 1 丈 5 尺至 2 丈 2 尺（4.8 公尺 -7.04 公尺）；小海船長 3 丈 6 尺至 5 丈 5 尺（11.52 公尺 -17.92 公尺），闊 1 丈 1 尺至 2 丈 1 尺（3.52 公尺 -6.72 公尺）。雖然船隻大小與乘船人數不見得為正成比，但通常搭載人數愈多，船隻規模愈大。

以飄臺地區來看，北部 21 件、南部 10 件、宜蘭 8 件、東部 2 件、東部或南部 2 件、宜蘭或東部 2 件、北部或宜蘭 1 件、澎湖 1 件、不詳 1 件（見圖 24），亦可從文本中的描述看出臺灣行政區劃在不同時期的進展。北臺灣以今北海岸、東北角和宜蘭一帶居多，南臺灣以今屏東縣東南部一帶為主，主要與洋流、風向有關。

23. 〈琉球國中山王尚為報明事〉，清乾隆 49 年（1784-01-22~1785-02-08），編號 pdsg-ntu-1088311-00266，歷代寶案脈絡分析系統。

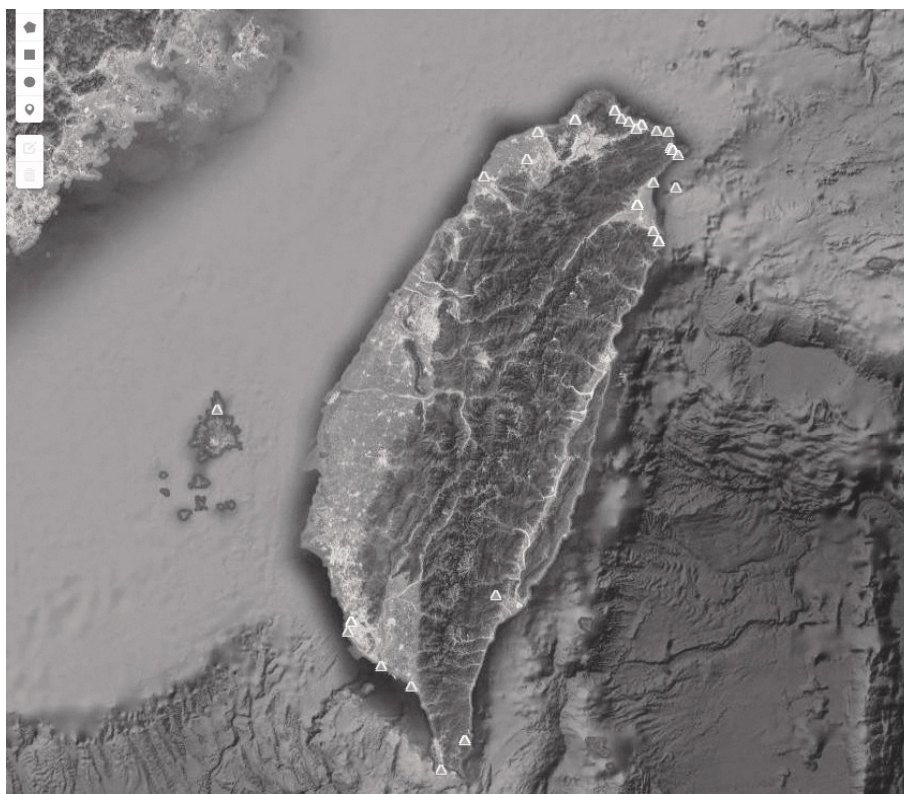


圖 24 琉球船隻飄臺地區示意圖

說明：部分案例雖然可判斷飄臺地區，但未說明明確地點供標示，因此本圖僅呈現 41 件案例。

資料來源：根據附表 1 以 DocuGIS 製作而成，<https://docusky.org.tw/docusky/docuTools/DocuGIS/>。

(三) 海難後的救助

大多數的案例，船隻在遭風時即受損害，飄臺觸陸後幾乎會面臨解體狀態，致使難民落水。因落水導致溺斃、行蹤不明、遭浪捲走有 7 件，約佔 15%。途中因病死亡有 9 件，約佔 20%。部分沿岸地區人煙稀少，難民需往內陸移動以尋求救助，在生番地遭生番殺害或行蹤不明有 3 件，約佔 7%。以救助者的身分來看，熟番和社丁 9 件、當地居民 7 件、漁船 5 件、中國人 3 件、兵船 2 件、粵人 1 件、澳甲 1 件、生番 1 件。再由地方官府提供難民糧食、衣物等急難救助，協助剩餘商品、船隻零件變賣，並安排船隻護送到省。

對於遭生番殺害者，中國官員表示「生番本係不通人性，且不知姓名、住址，未便遽令兵役徑入番地查拿，轉致驚擾。應飭臺灣文武各官，遴選妥幹兵役，協同屯兵、

社丁，在於番界留心查訪，務將正兇緝獲究辦」²⁴、「該難夷被害處所果否番境，剋日查明詳覆核辦，毋再遲延，速速」²⁵，然文本中不見臺灣地方官府的積極作為，再查 THDL 明清檔案文獻集，亦無更詳細的記載。²⁶ 因病滯留番地者，則派通事前往番社確查，該番社稱無人逃來，另有差役遍查難民上岸後的路徑，發現當地無難民逗留，但曾有不知姓名的患病者逗留未去，已相繼歿故，恰符合行蹤不明的人數。²⁷ 兩件案例分別發生於嘉慶、道光年間，可看到臺灣地方官府以消極模式處理生番地問題，主要與番界政策有關。

清朝對於難民的撫卹措施，安置柔遠驛期間，每人每日給米 1 升、鹽菜銀 6 厘，回國之日另給行糧 1 個月。乾隆 2 年（1737）以後，地方官府可由存公銀兩支付，核實報銷。然撫卹措施亦有例外的情況，乾隆 38 年（1773）有一海難受難者高達 111 人，地方官府認為此次人數眾多，且回國之期尚遠，若女性、兒童、嬰幼兒依例發給口糧，花費公款過於龐大，應請折半給賞。²⁸ 具外交性質的朝貢團人數雖多，但其身分不同於常民，因此待遇更優渥，正、副使每人日給廩銀 2 錢，其餘官員每人日給米 3 升、蔬薪銀 5 分 1 釐，跟伴、水梢日給米 1 升、鹽菜銀 1 分。²⁹ 此外，皇帝特別提到琉球國不需補進貢物，也不需追究、懲處朝貢人員未善盡其職一事。³⁰

多數船隻在海難後不堪使用，難民返國大多附搭船型規模較大的貢船。若難民人數過多則分批附搭貢船，³¹ 或募中國商船與貢船一同乘汛返國，³² 朝貢團則由官府賞銀 1,000 兩，用以租雇中國商船，船隻返還則待貢期到來開船歸還。³³ 僅二件原船返國的

24. 〈署福建等處承宣布政使司為恭迎敕書並接回使臣事〉，清嘉慶 17 年 5 月 4 日（1812-06-12），編號 pdsg-ntu-1088329-00085，歷代實案脈絡分析系統。

25. 〈署福建等處承宣布政使司為抄摺行知事〉，清道光 14 年 5 月 4 日（1834-06-10），編號 pdsg-ntu-1088316-00262，歷代實案脈絡分析系統。

26. 國立臺灣大學數位人文研究中心，臺灣歷史數位圖書館（THDL），<http://thdl.ntu.edu.tw/>。

27. 〈福建等處承宣布政使司為進貢事〉，清嘉慶 18 年 4 月 20 日（1813-05-20），編號 pdsg-ntu-1088329-00094，歷代實案脈絡分析系統。

28. 〈琉球國中山王尚為移知事〉，清乾隆 38 年 11 月 11 日（1773-12-24），編號 pdsg-ntu-1088311-00108，歷代實案脈絡分析系統。

29. 〈琉球國中山王尚為抄摺行知事〉，嘉慶 13 年 12 月 21 日（1809-02-05），編號 pdsg-ntu-1088329-00005，歷代實案脈絡分析系統。

30. 〈琉球國中山王尚為抄摺行知事〉，嘉慶 13 年 12 月 21 日（1809-02-05），編號 pdsg-ntu-1088329-00005，歷代實案脈絡分析系統。

31. 〈琉球國中山王尚為移知事〉，清乾隆 38 年 11 月 11 日（1773-12-24），編號 pdsg-ntu-1088311-00108，歷代實案脈絡分析系統。

32. 〈琉球國中山王尚為送還恩恩〉，清雍正 10 年 10 月 24 日（1732-12-11），編號 pdsg-ntu-1088309-00091，歷代實案脈絡分析系統。

33. 〈琉球國中山王尚為抄摺行知事〉，嘉慶 13 年 12 月 21 日（1809-02-05），編號 pdsg-ntu-1088329-00005，歷代實案脈絡分析系統。

案例，一為遭風但船隻無過多損毀的貢船，一為海船。

綜而言之，琉球國人因遭風飄至臺灣，基於對藩屬國的救助原則，臺灣地方官員須向有關單位報告此事並護送琉球難民到中國，間接與琉球國產生關連，因而被記錄在《歷代寶案》中。案例中的琉球人出航遭風之後，歷經數日到三個餘月漂至臺灣，花費二個月到近一年的時間安插到省，等待登船歸國的安置時間又耗費二個月到一年餘的時間，但相較在異地身亡者，可說是不幸中的大幸。

伍、結論

《歷代寶案》記錄琉球王國 400 多年與周遭國家的外交活動，反應 15 世紀琉球國和明朝建立宗藩關係與朝貢貿易之後，在政治、經濟、社會、文化等多層面的交流與影響，並延伸擴展至東亞各國，直到 19 世紀。此批沖繩的重要史料在日治時期的臺灣被保留下來，成為世界上保存最完整的《歷代寶案》版本，而臺灣為響應聯合國教科文組織的世界記憶計畫，故於 2018 年文化部文化資產局將其登錄為臺灣世界記憶國家名錄之一。³⁴

為了讓《歷代寶案》發揮更大的價值，針對文獻特性和使用者需求，本系統的建置呈現「多面向的鳥瞰文獻全貌」、「多重脈絡的探勘與分析」、「檢索結果的兩層式呈現」等三個重要特點。以便利使用者做為出發點，藉由系統功能將資料進行分類與分析，呈現數位化文獻的多元面向，提升使用者對於研究資料探勘的深度和廣度。

透過實作案例的探討，使用者可以迅速分析研究主題的多重脈絡關係，快速掌握研究主題的特點，亦可在系統中快速搜尋到研究主題的相關議題，但系統的建置有其邏輯和限制存在，更深入的探討仍需經由使用者進一步針對文本內容進行驗證與分析。最後，建置歷代寶案系統有賴多個單位跨領域的協力合作，期待本系統能成為使用者的好工具。

34. 琉球歷代寶案—世界記憶，文化資產局臺灣世界遺產潛力點，<https://twh.boch.gov.tw/world-memory/detail.aspx?id=96&t=nominate>，2021 年 3 月 1 日瀏覽。

參考文獻

小葉田淳 (1963)。歷代寶案について。史林，46(4)，523-539。

朱瑪瓏 (2000)。近代颱風知識的轉變——以臺灣為中心的探討。臺北：國立臺灣大學歷史學系碩士論文。

何孟興 (2009)。明代福州海防要地「竿塘山」之研究 (1368-1456 年)。止善，7，29-51。

呂青華 (2007)。琉球久米村人的民族學研究。臺北：國立政治大學民族研究所博士論文。

沖繩縣教育委員會 (2018)。歷代寶案の栞。沖繩：沖繩縣教育委員會。

徐葆光。中山傳信錄。中國哲學書電子化計劃。<https://ctext.org/wiki.pl?if=gb&res=378406>

梁嘉彬 (1987)。明清兩代福州那霸間風帆去程之研究。《第一屆中琉歷史關係國際學術會議論文集》。臺北：聯合報文化基金會國學文獻館，335-352。

國立臺灣大學數位人文研究中心。乾隆朝大清會典與則例對照檢索系統。<http://cspis.digital.ntu.edu.tw/>

——。歷代寶案脈絡分析系統。<http://lidaibaoan.digital.ntu.edu.tw/>

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

國立臺灣大學數位人文研究中心、資訊工程學系數位典藏與自動推論實驗室規劃，項潔教授主持，杜協昌博士設計開發。DocuSky 數位人文學術研究平台。

臺灣銀行經濟研究室編 (1964)。流求與雞籠山。臺北：臺灣銀行經濟研究室。

附錄 1 《歷代寶案》琉球人風飄臺灣案例

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
康熙 59 年 10 月 11 日	缺	缺	康熙 60 年 4 月 8 日	康熙 60 年 6 月 10 日	北山永郎部	中山	淡水金包裡社	前間等 6 人 (其中 2 人餓死)	經濟活動	pdsg-ntu-1088353-00197;pdsg-ntu-1088353-00200;pdsg-ntu-1088353-00206
雍正 9 年 11 月 24 日	雍正 9 年 11 月 24 日	雍正 10 年 1 月 27 日	雍正 10 年 4 月 11 日	雍正 10 年 7 月 7 日 (抵國)	那霸港	太平山	大鼻頭	石垣等 48 人 (其中 8 人病死)	賦稅	pdsg-ntu-1088309-00091;pdsg-ntu-1088309-00092;pdsg-ntu-1088309-00077;pdsg-ntu-1088309-00097
乾隆 4 年 10 月 23 日	缺	乾隆 4 年 11 月 16 日	乾隆 5 年 2 月 1 日	缺	永良部島	山原	三貂社	田澤實等 9 人	經濟活動	pdsg-ntu-1088309-00173;pdsg-ntu-1088309-00186
乾隆 15 年 2 月 12 日	乾隆 15 年 2 月 12 日	乾隆 15 年 2 月 25 日	乾隆 15 年 4 月 22 日	乾隆 15 年 7 月 24 日 *	馬齒山	無	淡水八尺門	慶留間等 4 人	經濟活動	pdsg-ntu-1088325-00012;pdsg-ntu-1088325-00027
乾隆 16 年 10 月 23 日	乾隆 16 年 11 月 6 日	乾隆 17 年 2 月 26 日	乾隆 17 年 4 月 25 日	缺	古米山	中山王府	淡水雞籠山	比屋定目指等 22 人	賦稅	pdsg-ntu-1088325-00117;pdsg-ntu-1088325-00113;pdsg-ntu-1088325-00053;pdsg-ntu-1088325-00077
缺	缺	缺	乾隆 19 年 6 月 20 日	缺	缺	缺	淡水奇直港	豐宇望等 19 人	缺	pdsg-ntu-1088325-00117;pdsg-ntu-1088325-00113
乾隆 38 年 8 月 23 日	乾隆 38 年 8 月 23 日	乾隆 38 年 8 月 27 日	乾隆 38 年 10 月 23 日	缺	宮古島	多良間	淡水南崁港	多良間等大小男婦 114 人 (其中 4 人病死, 婦人生 1 子)	民生	pdsg-ntu-1088311-00108;pdsg-ntu-1088311-00094;pdsg-ntu-1088311-00095

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
乾隆 41 年 6 月 14 日； 乾隆 41 年 11 月 13 日	乾隆 41 年 6 月 22 日； 乾隆 42 年 4 月 5 日	無； 乾隆 42 年 4 月 8 日	乾隆 41 年 8 月 24 日； 乾隆 42 年 8 月 15 日	乾隆 41 年 11 月 13 日； 乾隆 42 年 12 月 4 日	太平山；福建	中山王府；琉球國	浙江奉化縣； 鳳曼尾 / 鳳鼻尾	向宣烈等 34 人	賦稅	pdsg-ntu-1088311-00161;pdsg-ntu-1088311-00180;pdsg-ntu-1088311-00185;pdsg-ntu-1088311-00192
乾隆 41 年 11 月 23 日； 乾隆 42 年 4 月 25 日	乾隆 41 年 11 月 23 日； 乾隆 42 年 4 月 25 日	缺	缺	乾隆 42 年 12 月 4 日	姑米山；東洋山島	中山王府	三貂	田福等 10 人 (其中 4 人溺死)	賦稅	pdsg-ntu-1088311-00161;pdsg-ntu-1088311-00180;pdsg-ntu-1088311-00192
乾隆 44 年 9 月 19 日 *	乾隆 44 年 9 月 24 日 *	乾隆 44 年 10 月 6 日 *	乾隆 44 年 12 月 29 日、 乾隆 45 年 1 月 27 日	乾隆 45 年 1 月 25 日； 乾隆 45 年 2 月 10 日	久米島 *	中山王府	阿里 *	石原等 25 人	賦稅	pdsg-ntu-1088311-00185;pdsg-ntu-1088311-00192
乾隆 50 年 1 月 20 日	乾隆 50 年 1 月 21 日	乾隆 50 年 1 月 29 日	乾隆 50 年 5 月 29 日 *	乾隆 50 年 12 月 30 日	太平山	多良間	臺灣山後生番地界	向裔富濱等 20 人 (其中 4 人死亡、1 人病死)	賦稅	pdsg-ntu-1088311-00271;pdsg-ntu-1088311-00276
乾隆 50 年 11 月 11 日	缺	乾隆 50 年 12 月 8 日	乾隆 51 年 5 月 16 日	乾隆 51 年 12 月 30 日	那霸港	缺	臺灣山後	平田等 23 人 (其中 1 人病死)	經濟活動	pdsg-ntu-1088311-00308;pdsg-ntu-1088311-00290
乾隆 53 年 10 月 22 日	缺	乾隆 53 年 10 月 29 日	乾隆 54 年 3 月 24 日	乾隆 54 年 閏 5 月 2 日	中山王府	太平山	生番地	平良等 16 人 (其中 10 人遇難後行蹤不明)	賦稅	pdsg-ntu-1088327-00021;pdsg-ntu-1088311-00323

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
乾隆 55 年 10 月 8 日	缺	乾隆 55 年 11 月 27 日	乾隆 57 年 3 月 22 日	缺	那霸府	那姑呢	臺灣海邊	古波津等 16 人 (12 人落水後行蹤不明、1 人病死)	賦稅	pdsg-ntu-1088327-00054;pdsg-ntu-1088327-00087
嘉慶 4 年 9 月 28 日	嘉慶 4 年 10 月 2 日	缺	嘉慶 5 年 1 月 22 日	嘉慶 5 年 閏 4 月 27 日	那霸府	宮古島	淡水三貂	兼个段等 20 人 (其中 1 人遇浪淹死)	經濟活動	pdsg-ntu-1088313-00021;pdsg-ntu-1088313-00052
嘉慶 8 年 8 月 25 日	嘉慶 8 年 9 月 5 日	缺	缺	無	缺	中山王府	鳳山縣	馬齒山 (病死)	賦稅	pdsg-ntu-1088313-00186
嘉慶 7 年 10 月 15 日	嘉慶 7 年 10 月 16 日	缺	嘉慶 7 年 2 月 8 日	嘉慶 8 年 5 月 16 日	那霸港	福建	大武崙	二號貢船王成教等 80 人	朝貢	pdsg-ntu-1088313-00150;pdsg-ntu-1088313-00177;pdsg-ntu-1088313-00125;pdsg-ntu-1088313-00127;pdsg-ntu-1088313-00133;pdsg-ntu-1088313-00154;pdsg-ntu-1088313-00155;pdsg-ntu-1088313-00157;pdsg-ntu-1088313-00158;pdsg-ntu-1088313-00161;pdsg-ntu-1088313-00163;pdsg-ntu-1088313-00166;pdsg-ntu-1088313-00139;pdsg-ntu-1088313-00143;pdsg-ntu-1088313-00145;pdsg-ntu-1088313-00121

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
嘉慶 11 年 8 月 10 日； 嘉慶 11 年 10 月 9 日	缺	嘉慶 11 年 10 月 20 日	嘉慶 12 年 2 月 30 日、 3 月 8 日	嘉慶 12 年 10 月 6 日	那霸港；姑米山	閩	鳳山縣枋寮	頭號船楊克敦等 119 人	朝貢	pdsg-ntu-1088313-00218;pdsg-ntu-1088313-00219;pdsg-ntu-1088313-00220;pdsg-ntu-1088313-00221;pdsg-ntu-1088329-00004;pdsg-ntu-1088329-00005
嘉慶 11 年 8 月 10 日； 嘉慶 11 年 10 月 9 日	缺	嘉慶 11 年 10 月 14 日	嘉慶 11 年 6 月 7 日	嘉慶 12 年 6 月 6 日	那霸港；姑米山	閩	澎湖吉貝嶼	二號船梁躬等 82 人	朝貢	pdsg-ntu-1088313-00218;pdsg-ntu-1088313-00219;pdsg-ntu-1088313-00220;pdsg-ntu-1088313-00221;pdsg-ntu-1088329-00004;pdsg-ntu-1088329-00005;pdsg-ntu-1088329-00006;pdsg-ntu-1088329-00007;pdsg-ntu-1088329-00009;pdsg-ntu-1088329-00038;pdsg-ntu-1088329-00040
嘉慶 13 年 3 月 1 日	缺	嘉慶 13 年 4 月 15 日	缺	嘉慶 13 年 4 月 30 日	琉球絲滿	無	臺灣北路不識地名洋面	金城、三里、宮平等 3 人	經濟活動	pdsg-ntu-1088329-00019;pdsg-ntu-1088329-00013
嘉慶 15 年 10 月 17 日	嘉慶 15 年 10 月 20 日	嘉慶 15 年 10 月 23 日	嘉慶 15 年 11 月 7 日	嘉慶 17 年 5 月 3 日	中山	麻姑山	臺灣南路番地四浮鑾	建西表等 42 人 (其中 19 人病死、4 人病重行蹤不明、3 人被生番殺死、3 人渡溪溺死)	公務來往	pdsg-ntu-1088329-00094;pdsg-ntu-1088329-00085

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
嘉慶16年閏3月16日	缺	嘉慶16年4月4日	缺	嘉慶17年5月3日	琉球國泊村	八重山	淡水	具志堅等12人	賦稅	pdsg-ntu-1088329-00085
嘉慶18年4月10日	缺	嘉慶18年4月22日	嘉慶18年7月4日	嘉慶19年4月3日	那霸府	八重山	芝罘里海邊	樂永儀等9人	經濟活動	pdsg-ntu-1088329-00106;pdsg-ntu-1088329-00125
嘉慶18年4月14日	缺	嘉慶18年4月25日	嘉慶18年7月4日	嘉慶19年4月3日	姑米山	馬齒山	金雞貂海邊	伊良波等3人	經濟活動	pdsg-ntu-1088329-00106;pdsg-ntu-1088329-00125
嘉慶19年3月15日	缺	嘉慶19年4月1日	缺	嘉慶20年5月2日	那霸府	八重山	鳳山縣轄番社	宮城等7人	賦稅	pdsg-ntu-1088329-00133;pdsg-ntu-1088329-00138
嘉慶20年4月21日	缺	嘉慶20年5月3日	嘉慶20年8月27日、10月24日	嘉慶21年6月2日	芭蕉島哪吧村	八重山	鳳山縣	安里等13人	經濟活動	pdsg-ntu-1088329-00156
嘉慶20年4月8日	嘉慶20年4月9日	嘉慶20年4月24日	嘉慶20年9月9日	嘉慶21年5月3日;嘉慶22年4月28日	宮古島	那霸府	噶瑪蘭烏石港口	馬瑞慶山等9人	賦稅	pdsg-ntu-1088329-00153;pdsg-ntu-1088329-00159;pdsg-ntu-1088329-00167;pdsg-ntu-1088315-00014;pdsg-ntu-1088315-00021
嘉慶22年10月18日	嘉慶22年10月19日	嘉慶22年10月23日	嘉慶23年2月24日	嘉慶23年4月27日	久高島、鳥島	外島(未詳細說明)	艋舺金包里澳口	內間等7人(其中1人病死)	經濟活動	pdsg-ntu-1088315-00011;pdsg-ntu-1088315-00035;pdsg-ntu-1088315-00014;pdsg-ntu-1088315-00036
嘉慶24年8月2日	缺	缺	嘉慶24年11月4日	嘉慶25年4月26日	姑米山	那霸府	雞籠	志理真等9人	賦稅	pdsg-ntu-1088315-00047

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
道光 5 年 9 月 24 日	缺	道光 5 年 10 月 7 日	道光 6 年 3 月 24 日	道光 6 年 4 月 30 日	姑米山	那霸府	噶瑪蘭 蘇澳馬賽	上江洲 筑登之 等 30 人	賦稅	pdsg-ntu-1088315-00154;pdsg-ntu-1088315-00155;pdsg-ntu-1088316-00057;pdsg-ntu-1088315-00189;pdsg-ntu-1088315-00190;pdsg-ntu-1088315-00195
道光 8 年 9 月 24 日	道光 8 年 9 月 26 日	道光 8 年 10 月 7 日	道光 8 年 12 月 21 日	道光 9 年 5 月 2 日	那霸港	名護郡	琅嶠山 後	葉渡山 等 14 人	經濟活動	pdsg-ntu-1088316-00136;pdsg-ntu-1088316-00090;pdsg-ntu-1088316-00084;pdsg-ntu-1088316-00088
道光 12 年 11 月 4 日	缺	道光 12 年 11 月 24 日	道光 13 年 4 月 12 日	道光 13 年 5 月 1 日	那霸府	山北府	噶瑪蘭 龜山外洋	比嘉等 4 人	經濟活動	pdsg-ntu-1088316-00276;pdsg-ntu-1088316-00285;pdsg-ntu-1088316-00249;pdsg-ntu-1088316-00219
道光 13 年 1 月 5 日	道光 13 年 1 月 9 日	道光 13 年 1 月 23 日	道光 13 年 12 月 10 日	道光 14 年 5 月 2 日	那霸府	渡名喜島	南風澳 觸奇梨	知念等 10 人 (其中 6 人被 生番殺 死、1 人溺死)	經濟活動	pdsg-ntu-1088316-00286;pdsg-ntu-1088316-00262
道光 13 年 12 月 18 日	道光 13 年 12 月 18 日	道光 13 年 12 月 23 日	道光 14 年 7 月 18 日	道光 15 年 5 月 5 日	那霸府	八重山 島	噶瑪蘭 觸奇利	真章等 5 人 (其中 2 人病 死、1 人溺死)	賦稅	pdsg-ntu-1088316-00300;pdsg-ntu-1088317-00004
道光 16 年 1 月 13 日	缺	道光 16 年 1 月 19 日	道光 16 年 4 月 19 日	道光 16 年 5 月 1 日	姑米山 島	中山王 府	不識地 名洋面	嘉手苳 史地頭 等 17 人 (其中 11 人行 蹤不明)	賦稅	pdsg-ntu-1088317-00139;pdsg-ntu-1088317-00074;pdsg-ntu-1088317-00038;pdsg-ntu-1088317-00085
道光 16 年 9 月 10 日	道光 16 年 9 月 11 日	道光 16 年 9 月 19 日	道光 17 年 1 月 11 日	道光 17 年 5 月 8 日	久米島	那霸府	不識地 名外洋	玉城仁 屋等 36 人	賦稅	pdsg-ntu-1088317-00096;pdsg-ntu-1088317-00090

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
道光 18 年 1 月 13 日； 道光 18 年 1 月 21 日	道光 18 年 1 月 13 日； 道光 18 年 1 月 21 日	道光 18 年 2 月 8 日	道光 18 年 7 月 13 日	道光 19 年 5 月 3 日	不詳； 德之島	那霸府	琅嶠大 秀房莊	大城筑 登之等 5 人	賦稅	pdsg-ntu-1088317-00172;pdsg-ntu-1088317-00173;pdsg-ntu-1088317-00175;pdsg-ntu-1088317-00163;pdsg-ntu-1088317-00165;pdsg-ntu-1088317-00166
道光 18 年 3 月 5 日	道光 18 年 3 月 14 日	道光 18 年 3 月 18 日	道光 18 年 8 月 18 日	道光 19 年 5 月 3 日	那霸府	八重山 島	東港	鄭氏山 口等 14 人	賦稅； 經濟 活動	pdsg-ntu-1088317-00173;pdsg-ntu-1088317-00175;pdsg-ntu-1088317-00166
道光 20 年 4 月 8 日	道光 20 年 4 月 8 日	道光 20 年 4 月 19 日	道光 20 年 11 月 26 日	道光 21 年 5 月 15 日	那霸港	宮古島	鳳山縣 轄打鼓 汛岐	平氏友 寄等 9 人	賦稅； 經濟 活動	pdsg-ntu-1088317-00263;pdsg-ntu-1088317-00257
道光 20 年 11 月 14 日	道光 20 年 11 月 17 日	道光 20 年 12 月 6 日	道光 21 年 2 月 27 日	道光 21 年 4 月 15 日	喜界島	那霸府	傀儡番 界	平姓大 城筑登 之親雲 上等 26 人	賦稅	pdsg-ntu-1088317-00259;pdsg-ntu-1088317-00255;pdsg-ntu-1088317-00257
道光 24 年 4 月 12 日	道光 24 年 4 月 13 日	道光 24 年 4 月 15 日	道光 24 年 9 月 2 日	道光 25 年 5 月 5 日	宮古島	泊村	噶瑪蘭	具志堅 等 7 人	經濟 活動	pdsg-ntu-1088333-00095;pdsg-ntu-1088333-00099;pdsg-ntu-1088333-00106;pdsg-ntu-1088333-00110
道光 39 年 9 月 9 日； 道光 39 年 9 月 11 日	道光 29 年 9 月 11 日	道光 29 年 9 月 24 日	道光 30 年 2 月 15 日	道光 30 年 5 月 1 日	中山； 八重山 外山平 久保村 內洋	八重山 島	噶嗎蘭 廳	山長元 等 41 人 (其中 5 人溺死)	賦稅； 經濟 活動	pdsg-ntu-1088333-00277;pdsg-ntu-1088333-00279
咸豐 1 年 12 月 6 日	咸豐 1 年 12 月 6 日	咸豐 1 年 12 月 17 日	咸豐 1 年 3 月 7 日	咸豐 2 年 4 月 28 日	久米山	渡名喜 島	噶瑪蘭	宮平等 14 人	經濟 活動	pdsg-ntu-1088334-00040;pdsg-ntu-1088334-00042;pdsg-ntu-1088334-00067;pdsg-ntu-1088334-00069;pdsg-ntu-1088334-00071

出發時間	遭風時間	飄臺時間	到省安插時間	離驛登舟時間	出發地	預估地	遭難地	受難者	事由分類	編號
咸豐 1 年 10 月 7 日； 咸豐 1 年 10 月 27 日	咸豐 1 年 10 月 27 日	咸豐 1 年 11 月 7 日	咸豐 1 年 3 月 7 日	咸豐 2 年 4 月 28 日	久米山；慶良間	那霸府	馬鍊洋面	永束齊等 10 人	賦稅	pdsg-ntu-1088334-00040;pdsg-ntu-1088334-00044;pdsg-ntu-1088334-00067
咸豐 1 年 12 月 6 日	咸豐 1 年 12 月 7 日	咸豐 1 年 12 月 13 日	咸豐 1 年 3 月 7 日	咸豐 2 年 4 月 28 日	那霸港	久米山	八尺門外洋	喜久里等 13 人	賦稅；經濟活動	pdsg-ntu-1088334-00042;pdsg-ntu-1088334-00067;pdsg-ntu-1088334-00069;pdsg-ntu-1088334-00071
咸豐 7 年 6 月 10 日	咸豐 7 年 6 月 11 日	咸豐 7 年 6 月 18 日	咸豐 7 年 9 月 13 日、9 月 14 日	咸豐 8 年 5 月 4 日	那霸府	八重山島	淡水廳三貂保	新垣等 9 人	賦稅	pdsg-ntu-1088334-00196;pdsg-ntu-1088334-00222

說明：1、* 記號者為根據清代中琉關係檔案等史料補充填表。

2、琉球船隻離驛登舟後，需先至閩安鎮驗明人船貨物，後由閩安鎮撥兵護送至外洋乘風放洋長行回國。離驛登舟之日和乘風放洋長行回國之日不同，因前者記載較齊全，故採用前者。

資料來源：根據歷代寶案分析脈絡系統關鍵詞「臺灣」所得 135 筆資料製作而成。



陳志銘
Chih-Ming Chen

陳志銘，目前為國立政治大學圖書資訊與檔案學研究所教授兼圖書館館長、社會科學資料中心主任，過去曾擔任國立政治大學圖書資訊與檔案學研究所所長、政大圖書館副館長、政大電算中心副主任等行政工作，曾獲得科技部吳大猷先生紀念獎、政大特聘教授、政大學術研究特優獎、政大學術研究優良獎等學術殊榮，過去合計已發表超過 70 篇 SSCI 及 SCI 等國際期刊論文，也出版過「DSpace 開放源碼數位典藏系統建置理論與實務」及「創新數位學習模式與教學應用」二本專書，主要研究專長包括數位學習、數位閱讀、數位人文、數位圖書館、大數據資料探勘、人工智慧與智慧型網際網路系統。目前擔任政大圖書館及圖書資訊與檔案學研究所共同發行「圖資與檔案學刊」主編，以及國立臺灣大學圖書資訊學系發行之「圖書資訊學研究」(TSSCI 收錄) 編輯委員，Elsevier 發行之 Computers & Education: Artificial Intelligence 國際期刊編輯委員，也曾擔任過科技部資訊工程學門規劃委員，以及資訊教育學門、圖書資訊學門複審委員。

個人網頁：<http://dlii-info.dlii.nccu.edu.tw/>

Chih-Ming Chen is currently a Distinguished Professor in the Graduate Institute of Library, Information and Archival Studies at National Chengchi University, and he is also the University Librarian at National Chengchi University Library, Taipei, Taiwan. He received B.Sc. and M.Sc. degree from the Department of Industrial Education at National Taiwan Normal University in 1992 and 1997, and received Ph.D. degree from the Department of Electronic Engineering at National Taiwan University of Science and Technology in 2002. His research interests include digital library, digital humanities, e-learning, data mining, machine learning, and intelligent agents on the web. He has received many awards for his outstanding researches.



張鐘
Chung Chang

張鐘，國立政治大學圖書資訊與檔案學研究所碩士，目前為華人文化主體性研究中心資訊工程師，研究興趣為數位圖書館、數位典藏、數位人文、數位學習，並發表多篇數位人文相關期刊與研討會論文，過去曾參與過羅家倫數位人文研究平台、明代文集數位人文研究平台、馬來西亞人物誌數位人文平台等建置及多項數位人文工具開發工作。

個人網頁：[http://www.ccstw.nccu.edu.tw/team/m/ 張鐘 /](http://www.ccstw.nccu.edu.tw/team/m/張鐘/)

Chung Chang received B.Sc. degree from the Department of Information and Library Science at Fu Jen Catholic University in 2016, and obtained his master degree from the Graduate Institute of Library, Information and Archival Studies at National Chengchi University in 2018. He is currently an IT engineer of the Research Center for Chinese Cultural Subjectivity at National Chengchi University, Taipei, Taiwan. His research interests include Digital Humanities, Digital Library, and E-learning.



徐志帆
Chih-Fan Hsu

徐志帆，國立政治大學圖書資訊學與檔案學研究所碩士，目前為國立政治大學圖書資訊與檔案學研究所專任助理，具有開發數位人文、數位學習平台經驗。

個人網頁：<https://billxu0521.github.io/>

Chih-Fan Hsu received his master degree from the Graduate Institute of Library, Information and Archival Studies at National Chengchi University. His research interests include Digital Humanities, Digital Library, and E-learning.



林嘉培
Chia pei Lim

林嘉培，馬來西亞華人，國立政治大學哲學系碩士，拉曼大學中華研究院中文系學士。目前為國立政治大學華人文化主體性研究中心計劃專任助理。研究興趣：數位人文、數位典藏、海洋史、沃格林研究（種族）

Chia Pei Lim received his Master degree from the Department of Philosophy at National Chengchi University in 2019, and received Bachelor's degree from the Institute of Chinese Studies at University Tunku Abdul Rahman in 2014. He is currently a Project Assistant in the Research Center for Chinese Cultural Subjectivity at National Chengchi University, Taiwan. His research interests included Digital Humanities, Digital Archive, Maritime History, and Eric Voegelin's on Race Issues.

馬來西亞人物誌數位人文平台之建置與應用

Development and Application of Digital Humanities Research Platform for Biographies of Malaysia Personalities

陳志銘

Chih-Ming Chen

國立政治大學圖書資訊與檔案學研究所教授兼圖書館館長

Professor & University Librarian, Graduate Institute of Library, Information and Archival
Studies at National Chengchi University

張鐘

Chung Chang

國立政治大學華人文化主體性研究中心資訊工程師

IT Engineer, Research Center for Chinese Cultural Subjectivity
at National Chengchi University

徐志帆

Chih-Fan Hsu

國立政治大學圖書資訊與檔案學研究所專任助理

Research Assistant, Graduate Institute of Library, Information and Archival Studies
at National Chengchi University

林嘉培

Chia pei Lim

國立政治大學華人文化主體性研究中心專任助理

Research Assistant, Research Center for Chinese Cultural Subjectivity
at National Chengchi University

摘要

馬來西亞人物誌數位人文平台為國立政治大學華人文化主體性研究中心與馬來西亞興安會館、馬來西亞馬來亞大學華人研究中心合作建置，此一平台以「馬來西亞興化群賢錄」作為典藏內容，採用開放原始碼的數位典藏系統 Omeka S 為基礎發展系統功能，結合近讀 (Close Reading) 及遠讀 (Distant Reading) 兩個面向功能，作為本平台發展數位工具

的基礎。近讀為透過細讀文本全文解讀內容，能對文本內容進行深入的理解，因此本平台提供友善圖文閱讀介面，並搭配多個輔助文本閱讀工具，例如實體命名 (Name Entity Recognition) 等功能，幫助研究者理解人物誌中的內容；遠讀則是透過文本分析方法，幫助研究者從宏觀的角度解讀文本，除關鍵詞分析外，另提供地理資訊、社會網絡分析等功能輔助使用者掌握人物誌中人物關係與地理位置分布狀況，有助於加速解讀文本效率，探討隱含在人物誌中的脈絡。本文介紹了開發馬來西亞人物誌數位人文平台的第一階段成果。除了優化該平台的功能和使用界面外，我們未來的工作還將繼續開發更多新功能，例如個人資源檔案，分層主題分析工具，觀點遷移和年代劃分的分析工具。並嘗試應用此數位人文研究平台來協助數位人文科學的教學。

關鍵詞：數位人文、Omeka S、馬來西亞人物誌、人物社會網絡分析、地理資訊系統

Abstract

The Research Center for Chinese Cultural Subjectivity of National Chengchi University co-developed a digital humanities research platform for Biographies of Malaysia Personalities (DHRPBMP) with Malaysian Chinese Research Center of Malaya University. We adopted "The Malaysia Henghua Personalities" as basic corpus for the DHRPBMP and chose Omeka S, which is a new generation open source of institutional repository system, as the foundation for the development of DHRPBMP. The DHRPBMP provides several basic functions including a convenient and user-friendly reading interface with the support of an automatic text annotation tool based on name entity recognition (NER), linked data (LD), metadata search, full-text search, geographic information system (GIS), and character social network analysis to support digital humanities research. The developed DHRPBMP aims to assist digital humanists to efficiently read or analyze the contents of The Malaysia Henghua Personalities through distant reading and close reading. This paper presents the first stage outcomes of developing the DHRPBMP. In addition to optimizing the functions and user interface of this platform, our future work will also keep on developing or integrating more new functions, such as personal resource archives, hierarchical topic analysis tool, observation tool of viewpoint migration and age division (OTVMAD) as well as try to apply this digital humanities research platform to assist digital humanities' instruction.

Keywords: Digital Humanities, Omeka S, Biographies of Malaysia Personalities, Character Social Network Analysis, Geographic Information System

壹、緒論

隨著網路與資訊科技的快速發展，數位化已經成為全世界的發展潮流與趨勢，而隨著數位科技的進步，已逐漸改變了傳統人文領域的研究方法，為人文研究創造新的思維。在過去，人文學者多半以紙本的形式進行人文研究，現今則透過紙本數位化後建立各種不同的數位典藏資料庫，再加上發展可以分析數位典藏資料庫文本的數位工具輔助研究，已對人文研究領域的研究模式與方法產生巨大的改變與影響。換言之，透過資料庫輔以進行人文研究，已逐漸成為人文學者考量的研究模式之一 (Hockey, 2004)，也逐漸發展為「數位人文學」之新興研究領域。面對此一數位時代，人文研究者也開始積極思考，要如何運用資訊科技輔以人文研究，來幫助尋找在前數位時代中難以觀察的現象、無法想像的議題與無法進行的研究 (項潔、涂豐恩, 2011)。

數位人文學為數位技術與人文學科融合的新興研究領域 (Drucker, 2013)，它包括有系統的使用人文學科的數位內容和工具，並使新的教學和研究成為可能。在數位人文領域的研究中，Correll(2014) 指出細讀 (Close Reading) 的鏈接是支持遠讀 (Distant Reading) 假設的重要方法，近讀係透過細讀文本全文來解讀內容，能對文本內容有著更深一層的理解；遠讀則是透過文本分析方法將文本進行抽象化之概念抽取，幫助人文學者從宏觀的角度解讀文本，例如檢索結果分布之視覺化，可以幫助人文學者掌握重要關鍵詞彙隨著時代、體裁及文章等的分布狀況觀察，有助於加速解讀文本效率，探討隱含在文本中的脈絡。在 Jänicke, Franzini, Cheema 與 Scheuermann(2017) 等人的研究中也提及，在使用遠讀視覺化工具時，若可以直接鏈結到來源文本，對於人文學者來說非常重要。目前所發展的數位人文分析工具的準確度仍難達成百分之百，因此數位人文分析工具應該要扮演的角色是幫助人文學者研究，而不是取代它。因此，數位人文分析工具應該有可以顯示內容摘要或者抽象層次的遠讀，同時也要有能夠鏈結遠讀細讀的功能，如此會更便於研究人員對於分析文本的解讀 (Moretti, 2016)。

社會網絡分析為遠讀的分析方式之一，能協助人文研究者找出隱含在大量史料、文本中複雜的網絡關係。哈佛大學與中央研究院歷史語言研究所及北京大學所合作之中國歷代人物傳記資料庫 (CBDB)，就是透過累積大量的中國歷代人物傳記資訊，利用群體傳記學 (prosopography) 的方式輔以人文研究 (中國歷代人物傳記資料庫, 2020)。目前已有許多國內外學術機構、學術圖書館陸續開發出支援不同研究取向的數位人研究平台及數位工具。例如國家圖書館與國立政治大學合作發展之明人文集古籍數位人文平台，提供視覺化分析、自動標註工具和社會網絡分析功能 (Chen & Chang, 2019)；國立政治大學圖書館於

開發之「羅家倫先生文存數位人文研究平台」，除提供視覺化分析、文本標註等工具外，亦根據「羅家倫先生文存」文本之特性，並結合概念史和資訊技術發展觀點變遷與年代劃分工具 (陳志銘、張鐘與徐志帆，2020)；中央研究院數位文化中心開發之數位人文研究平台允許研究者上傳文本與權威詞、手動與自動標記文本等功能，並有統計圖表、文字雲、社會網絡分析圖等視覺化分析工具 (中央研究院數位文化中心，2020)；臺灣大學數位人文研究中心「DocuSky」平台，能讓使用者上傳文本，建立個人化的雲端資料庫，並提供了文本標記、文本探勘、視覺化、時空分析等可輔助人文研究者進行文本解讀之數位工具 (杜協昌，2018)。

馬來西亞為除兩岸外，華人族群比例最高的國家之一，2019 年國立政治華人文化主體性研究中心研究成員組團至馬來西亞進行訪查，並與馬來西亞馬來亞大學華人研究中心於 2020 年合作開發「馬來西亞興化群賢錄數位人文研究平台」，平台中支援繁體中文、簡體中文檢索，提供近讀及遠讀兩種模式工具，如文本標註、時空分析及社會網絡分析等，能幫助使用者快速理解馬來西亞華人的發展脈動。

貳、數位人文研究平台發展與功能現況

要藉由數位科技輔助人文學者進行文本研究與分析，首要條件是需具備數位化之文本資料，其次是數位人文分析工具。因此，數位人文的發展可以分為兩個階段，第一個階段是「數位典藏」(Digital Archive)，即是將原先的紙本進行全文數位化；第二階段即是發展數位人文分析工具。

數位典藏計畫除了將人文資料進行永久保存外，也在網路上開放學術利用，為各領域的人文研究者帶來易取得而豐富的研究資料，從而提高了人文研究者的研究效能。此外，「中國歷代人物傳記資料庫 (CBDB)」係由哈佛大學、臺灣中央研究院歷史語言研究所及北京大學三個機構合作建置，該資料庫收錄超過 470,000 筆中國歷代人物傳記和譜系資料，研究者可以透過此資料庫檢索歷史人物之籍貫、官職等資料，並提供 API 介接的方式讓數位人文研究者使用數位工具取得其中的資料 (中國歷代人物傳記資料庫，2020)。然而，現代的數位人文研究者已不再滿足於建立單純只具典藏功能的資料庫，而是開始建置各式各樣的數位人文研究工具，法鼓山的「CBETA 數位研究平台」除了提供漢文佛經的線上閱讀介面之外，亦提供「對讀文本編輯」、「部類、年代、作者與譯者統計分析」、「詞彙前後綴詞、分布地圖」等功能，供佛學研究者參考 (Tu, Hung, & Lin, 2012)。DocuSky 數位人文學術研究平台是由

國立臺灣大學數位人文研究中心與資訊工程學系數位典藏與自動推論實驗室所開發，係一個針對數位人文研究需要所開發的平台。此一平台由「機構導向資料庫」轉向「個人導向資料庫」，讓研究者能透過工具自行建立資料庫 (翁稷安，2016；謝博宇，2016)。研究者除了可以上傳自身所典藏的文本外，平台亦提供數位工具讓研究者能夠對建構的文字庫進行統計分析，幫助其查找資料、進行內文比對，以及進行字詞相關統計等，增加研究者對於文本內容的掌握 (杜協昌，2018)。此外，「羅家倫先生文存數位人文研究平台」則是由為國立政治大學社會科學資料中心所發展，主要典藏「羅家倫先生文存」合計共十二冊之數位化版本，並且為了因應研究需求，平台亦提供掃描檔與對照全文、全文檢索與後分類、自動文本標註 (Chen, Chen, & Liu, 2019)、觀點變遷分析，以及階層式主題分析等功能，以輔助使用者進行文本閱讀與多面向的文本分析。而「通用型古籍數位人文研究平台」則除了提供文本閱讀介面、全文檢索、雙詞檢索與自動標註等功能外，亦提供人物關係分析圖，可以自動從文本中識別人名，並透過視覺化方式呈現人物與人物之間的關係脈絡，協助人文學者理解文本中的人物社會網絡關係 (Chen & Chang, 2019)。

綜合以上所述，人文學者已皆越來越仰賴於「工具」的輔助，並運用數位工具來解決過去的人文研究問題，而一個支援數位人文研究需具備近讀及遠讀兩種模式的工具來輔助人文學者進行研究，並且人機合作才能進行準確度高與有效產出之數位人文研究模式。

參、實體命名與社會網絡分析

命名實體識別 (Named Entity Recognition, NER) 是標註功能中常使用到的技術之一，可以識別文本中具有特定意義或代表性的實體，包括人名、地名、組織名、時間日期等專有名詞，亦廣泛應用於地理資訊系統 (Geographic Information System, GIS) 的地名識別、社會網絡分析 (Social Network Analysis, SNA) 的人物名稱識別上。在不同語種的文本分析上，英文的命名實體識別相關研究發展甚早，但由於只需考慮詞本身的特徵而不涉及分詞問題，因此實現難度相對較低。而在中文的分析上，因其內在的特殊性促使在文本處理時必須先進行詞法分析，相較於英文命名實體識別具備更高的挑戰性 (Sun & Wang, 2010)。目前已有許多數位人文工具應用實體命名技術發展相關輔助數位人文應用，例如荷蘭萊頓大學的 MARKUS 文本分析和閱讀平台即是一個線上文本標註工具，允許使用者上傳文本並選取所需的標註類型，系統會針對人名、地名、官名等專有詞彙進行自動標註，並提供使用者維基百科、中國歷代人物傳記資料庫 (CBDB)、中國歷史地理地名資料庫 (TGAZ) 與漢典 (ZDict) 等資源之外部參照 (Ho & Hilde, 2014)。Yohan(2014) 等人提出了一個基於規則的命名實體

識別和分類系統，可以自動識別文本中的命名實體，並為它們提供對應於泰盧固語 (Telugu language) 的適當類別。Kestemont 和 Gussem (2017) 則在深度學習領域應用了分層神經網絡架構，以自動解決兩個序列標記的任務，包括詞性標記 (part-of-speech tagging) 和詞形還原 (lemmatization)。其研究結果顯示，在可接受的成效上與先前發表的研究成果相當。

人文研究應用社會網絡分析多半為進行文本脈絡探究，將社會網絡理論應用於文本的情節分析中，利用社會網絡分析將敘事性的文本中具有空間與時間結構、核心角色等關係轉換成網絡結構圖，隨著情節的發展，網絡圖也會有所轉變，能讓研究者輕易了解文本中關係與情節的變化 (Moretti, 2011)。Moretti(2011) 藉由莎士比亞的名著《哈姆雷特》的社會網絡圖找出關鍵人物 Horatio，並從 Horatio 的社會網絡關係探討故事的脈絡。趙葳 (2016) 也用此概念研究李劫的《大波》三部曲中的人物關係，以及故事的情節變化。許多國內外經典或著名文學，例如《聖經》、《魔戒》、《三國演義》等，都有許多學者運用社會網絡分析對文本中的人物關係、故事情節進行探討 (Duling, 2000；Ribeiro, 2015)。

網絡分析結合多來源文本也能呈現人物的生平交往或遷移過程，以及整個大環境時空脈絡。So 和 Hoyt(2013) 曾分析跨太平洋的文學社團與文學活動，從中整理出 1920 年代現代主義詩歌在全球傳播的路徑；劉吉軒 (2012) 分析臺灣海外左派刊物，用社會網絡圖呈現出海外臺灣人士的政治思想輪廓；金觀濤等人 (2016) 研究《新青年》雜誌中群眾觀念的變化，分析重要關鍵詞變化，並利用概念網絡圖呈現《新青年》雜誌從自由主義轉向馬列主義的過程。史丹福大學的「Mapping the Republic of Letters」網站呈現啟蒙時代知識份子透過書信交流所形成的社會網絡 (Stanford University, 2013)。Chen 等人利用社會網絡分析建構明代藏書家之間的社會關係模型，從中探索藏書家社群的核心人物及呈現社群中不同時期的傳承性和關聯性 (Chen, Chen, & Huang, 2018)。

目前國內外已有許多數位人文研究平台能提供相關社會網絡數據或分析功能協助研究者進行社會網絡分析，例如「中國歷代人物傳記資料庫」提供開放版本的資料庫供研究者下載，研究者能將資料匯入社會網絡分析軟體進行分析，以了解各人物間的社會關係，圖 1 為透過「中國歷代人物傳記資料庫」資料所繪製福建莆田於 1200 至 1250 年間取得進士學位者的社會網路關係 (中國歷代人物傳記資料庫，2020)；中研院數位人文研究平台內提供文本社會網絡分析功能，能讓研究者調整社會網絡圖中關聯強度、群間距離、關係距離長度；國家圖書館通用型古籍數位人文研究平台呈現明代文集的史料人物關係圖，並透過人機互動方式，讓使用者能修正圖中的人物節點與網絡關係，進而提高史料人物關係圖的完整性及正

確性 (Chen & Chang, 2019)；「臺灣歷史人物傳記資料庫 (TBDB)」則是在，專注於臺灣人物類別、屬性與關聯資訊建置，開發人物資料檢索、勘考工具及網絡分析工具 (TBDB 臺灣歷史人物傳記資料庫，2020)，謝順宏等 (2018) 即使用此一資料庫內收錄的彰化縣人物傳記全文，配合自動分群技術探索人物的社會網絡關係，並產出關係網絡圖。綜合上述，社會網絡分析能幫助研究者釐清文本之中的人物關係、情境脈絡等，進而了解當時的時空背景，協助人文學者進行更深入的人物關係解讀。

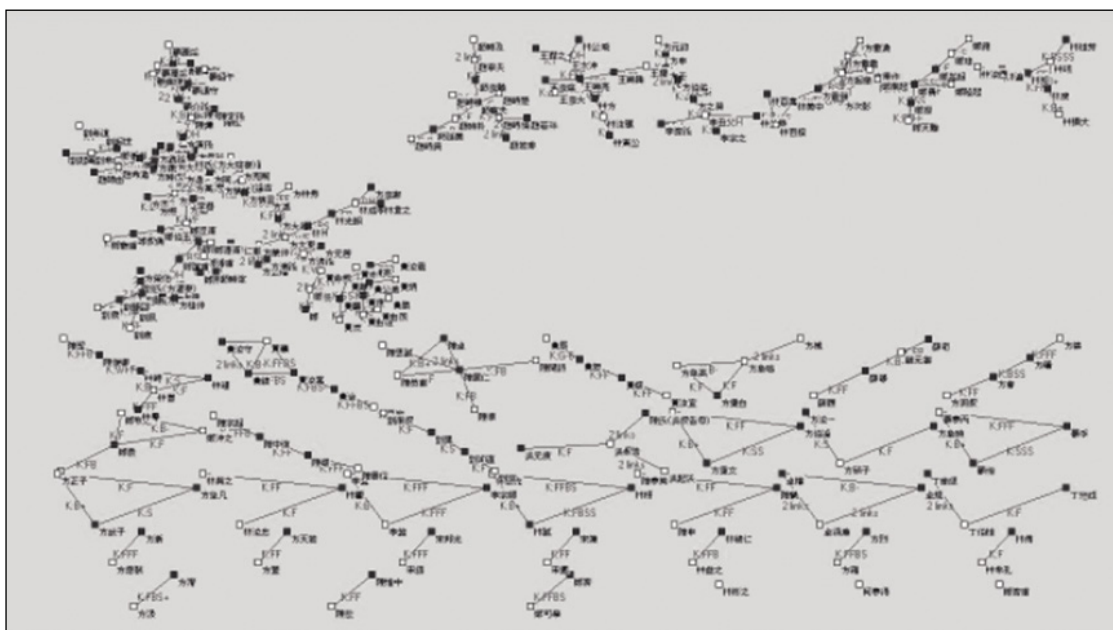


圖 1 福建莆田於 1200 至 1250 年間取得進士學位者的社會網路關係

資料來源：中國歷代人物傳記資料庫 (2020)

肆、馬來西亞華人發展與馬來西亞興化群賢錄

馬來西亞華人社會形成於十九世紀英殖民統治在馬來亞半島與北婆羅洲確立時期，大量華人因晚清中國國內政局動盪，被迫從南方省份經由英屬香港遠赴南洋謀生。馬來西亞華人社會網絡的形塑，最初係以草根階級（如農民、漁民）為主的移民為主，當地華人社團亦是由民間所組成，連同華文教育與華文報章被形容為馬來西亞華人社群的三大支柱。諸如興安會館等宗鄉親會館組織，除了以籍貫、地區作為地緣性組織之外，亦有以姓氏為主的血緣性會館組織。早期的會館組織部分為秘密結社組織而成，在華人移民社會初期凝聚同鄉的作用，負責建設與管理寺廟，舉辦祭祀活動，並且也設立私塾以進行教育，承繼華人文化傳統及其形式並將之紮根於南方。在英殖民政府逐漸加強對各類社團的管制以後，包含會館組織在內的華人社團也隨之適應環境的變化，在各不同時期作出相對應的調整，並且也起著代表

民間華人的訴求與當權政府的相關機構進行對話角色，以為華人爭取其權益，其中涉及包括經濟、社會和華文教育等議題。故會館組織及其人物發展軌跡，均能充份反映馬來西亞華人社會網絡是如何透過個別人物之間的個人關係、其職業身份、事件與行為之間，產生相互的關聯。

《馬來西亞興化群賢錄》一書由「馬來亞大學馬來西亞華人研究中心」與「馬來西亞興安總會」進行之馬來西亞興化人研究計畫成果之一，書中記載的八十五位興化民系人物介紹，涵蓋了政治、經濟、文教的數個領域，從他們的成長軌跡裡大多可以窺見到興化人家的文化印記，並反映馬來西亞華人當代史的各個時段發展脈絡。

伍、馬來西亞興化群賢錄數位人文平台

一般而言，支援數位人文研究的數位人文平台中，平台需要同時具備閱讀介面、資訊檢索、詞頻分析與資料探勘等面向功能，才能有效支援數位人文研究 (Widlöcher, 2015)。本研究發展之馬來西亞興化群賢錄數位人文平台也以上述方向為發展目標，主要開發四個輔助數位人文研究功能模組，包含資訊檢索與後設資料模組、閱讀介面模組、地理資訊模組及社會模組網絡分析模組。「馬來西亞興化群賢錄數位人文平台」採用開放原始碼的數位典藏系統 Omeka S 為基礎發展而成。Omeka S 已具有非常完整之資料的典藏、資料管理與數位策展功能，但是 Omeka S 缺乏友善、易用的閱讀介面，讓使用者難以同時閱讀掃描圖檔與對應之全文；文本檢索分析方面則對於中文語言的支援不佳，因此馬來西亞興化群賢錄數位人文平台針對 Omeka S 不足的地方進行改善及優化，並提供更進階的人物社會網絡分析功能，以輔助人文學者能藉由近讀及遠讀兩種解讀途徑，探討隱含在馬來西亞人物誌中的人物關係脈絡，讓此一平台使用上能夠更貼近人文研究者的需求。

平台首頁如圖 2 所示，系統架構如圖 3 所示，以下依序介紹各模組之功能與操作介面：



圖 2 馬來西亞興化群賢錄數位人文平台首頁

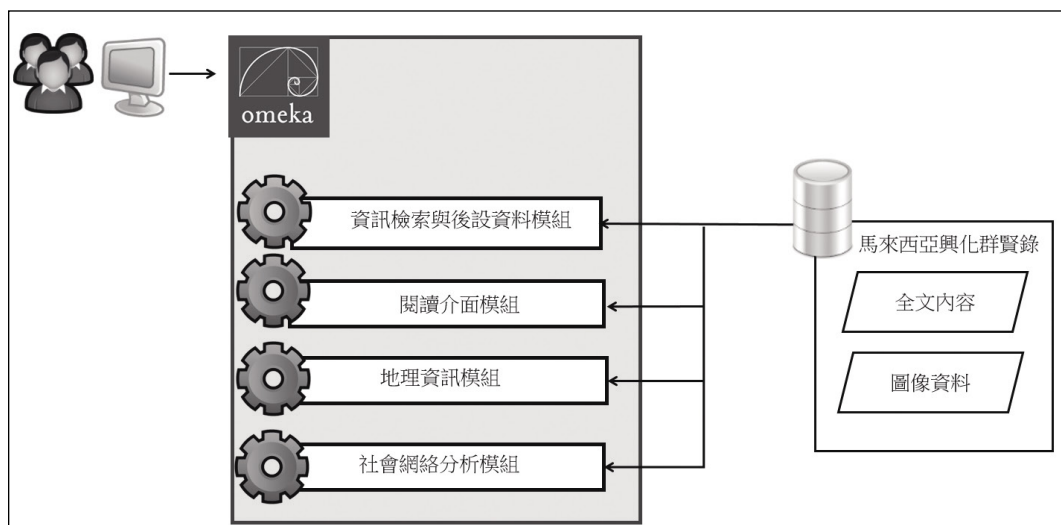


圖 3 馬來西亞興化群賢錄數位人文平台系統架構圖

一、資訊檢索與後設資料模組

「馬來西亞興化群賢錄數位人文平台」參照《馬來西亞興化群賢錄》一書中條目 (item) 的編排方式，係將書中每一位人物分別獨立成一個條目，以都柏林核心集作為主要後設資料欄位，如表 1 所示。資訊檢索部分針對中文語系檢索部分進行優化，將原本平台內建使用 MySQL 的檢索功能替換成 SOLR 檢索引擎，並支援簡體中文與繁體中文全文檢索和後分類分析功能，能加速使用者解讀效率，並能嘗試從不同面向分析檢索結果，目前平台提供稱

謂、關鍵字、作者等面向的檢索後分類功能，如圖 4 所示；本平台經熟悉馬來西亞文化之專家將書中的人物條目所屬之領域進行分類，建立各界群賢分類瀏覽功能，使用者可依據此一平台事先定義好之專業領域，進入該領域瀏覽相關的人物條目，目前本平台將書中的人物分為軍事、警界、法律、演藝、藝術、媒體、政治、經濟、文教、宗教等十大領域，如圖 5 所示。在資料交換上，本平台支援以 RESTful API 方式，提供 JSON、IIIF 等資料交換格式，可與其他數位典藏系統進行文字或圖像等資料交換，如圖 6 所示。後設資料模組方面，後設資料介面如圖 7 所示，點擊後設資料介面內關鍵詞，可查詢具有相同關鍵詞之條目。

表 1 後設資料範例

標題	林金樹
頭銜	拿督 DMSM、AMN、太平局紳 JP
頁面開始	002
頁面結束	007
關鍵字	馬六甲、興安、興化、興安會館
生卒年	1919-2001
出生地	中國福建福清新唐鎮塘頭村
祖籍	中國福建福清新唐鎮塘頭村
所屬會館	馬來西亞興安會館
資料來源	馬來西亞興化群賢錄 The Malaysia Henghua Personalities
撰寫者	林德順

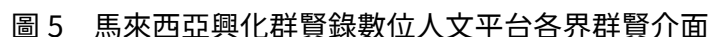
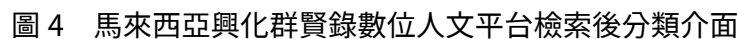


圖 6 RESTful API 格式

貢獻者 (Dublin Core)	林德順
主筆 (Dublin Core)	兴化
時間跨度 (Dublin Core)	1924-2013
頁面開始 (Bibliographic Ontology)	142
頁面結束 (Bibliographic Ontology)	147
頁數 (Bibliographic Ontology)	6
關鍵字 (A vocabulary for biographical information)	馬共 萬豐隆 中共 革命 義勇 抗日 武裝鬥爭 露露 英國 和平協議
祖籍 (Friend of a Friend)	福建省福清县江兜
所屬團體 (Friend of a Friend)	馬來亞共产党
引證 (Dublin Core)	馬來西亞興化群賢錄 The Malaysia Henghua Personalities

圖 7 後設資料介面

二、閱讀介面模組

「馬來西亞興化群賢錄數位人文研究平台」希望使用者在使用本平台閱讀文本的過程中，能像是閱讀真正的書籍一般便利與舒適，故採用掃描影像檔與文本共同顯示之閱讀介面，如圖 8 所示。左邊頁面為掃描之影像檔，右邊為與影像檔對應之全文，讓使用者於解讀過程中能進行圖文交互參照閱讀，可使用滑鼠滾輪或卷軸直覺性的下拉進行閱讀或頁面切換。此外，閱讀介面中也提供多項功能輔助使用者解讀文本，圖 9 為閱讀介面功能選單，排版設定支援使用者根據自身需求自由地調整字型大小、字距寬度、行距高度等；圖片設定允許使用者切換影像檔之顯示狀況；最後是本平台根據「馬來西亞興化群賢錄」之文本特性所發展的文本人名自動識別功能。

開門選單

開啟選單



陈平 (王文華, 1924-2013)

陈平 (Chin Peng) 原名王文華 (Ong Boon Hua)，福建省福清县江兜乡人王声彪之子。1924年10月21日出生于马来联邦霹靂州一个以福州人为主的华人城镇实兆远 (Sitiawan)。母亲 关芹治是新加坡梧槽区 (Rochor) 兴化大户 关家之女。王文華家有11名兄弟姊妹，他排行第二，下有5名弟弟，3名妹妹。其父把首4名儿子，分别取名文荣、文华、文富、文贵 (荣华富贵)，接下来的3名儿子则分别取名文全、文钦、文忠。

父亲从事兴化人传统行当 王声彪自幼离乡出海到新加坡投奔同村乡长王禄梓，在他的万丰隆脚车店当学徒。万丰隆创立于1915年，位于惹兰勿剎 (Jalan Besar) 和美瑶街 (Mayo Street) 的交接处。王声彪的生意渐见起色后，不断提携乡亲，支持他们到东

陈平 (Chin Peng) 原名王文華 (Ong Boon Hua)，福建省福清县江兜乡人王声彪之子。1924年10月21日出生于马来联邦霹靂州一个以福州人为主的华人城镇实兆远 (Sitiawan)。母亲 关芹治是新加坡梧槽区 (Rochor) 兴化大户 关家之女。王文華家有11名兄弟姊妹，他排行第二，下有5名弟弟，3名妹妹。其父把首4名儿子，分别取名文荣、文华、文富、文贵 (荣华富贵)，接下来的3名儿子则分别取名文全、文钦、文忠。

父亲从事兴化人传统行当 王声彪自幼离乡出海到新加坡投奔同村乡长王禄梓，在他的万丰隆脚车店当学徒。万丰隆创立于1915年，位于惹兰勿剎 (Jalan Besar) 和美瑶街 (Mayo Street) 的交接处。王声彪的生意渐见起色后，不断提携乡亲，支持他们到东

圖 8 文本閱讀介面



圖 9 文本閱讀介面功能選單

「馬來西亞興化群賢錄」書中所收錄條目以人物為主體，並記錄該人物之生平記事，本書中提及之人物在馬來西亞華人社會具有相當程度重要性，而重要人物不單僅限於書中所收錄之 85 位人士，亦有可能隱藏於描述此 85 位人物的內文之中，因此有必要從非結構化的文本中抽取潛在的人物資訊。依據以上目的，本平台特別開發文本人名自動識別功能，處理流程如圖 10 所示。首先將文本透過 BERT 將文本自然語言向量化，並透過實體命名 (NER) 辨識技術進行人物實體命名分析，將辨識出之人名，以顏色直接標註於文本介面中，點擊人物可查看該人物於其它文本中所出現之段落，進行文本內部參照，加速使用者掌握該人物於平台內的分布狀況或整體脈絡資訊，標註人物查詢如圖 11 所示。

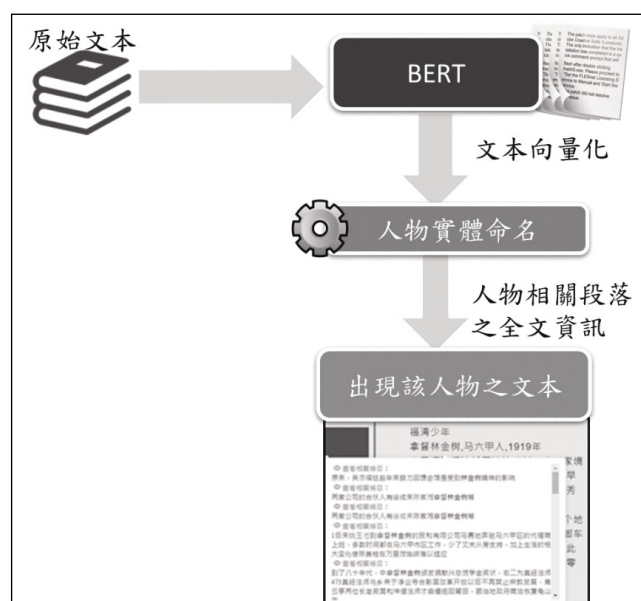


圖 10 文本人名自動識別功能流程

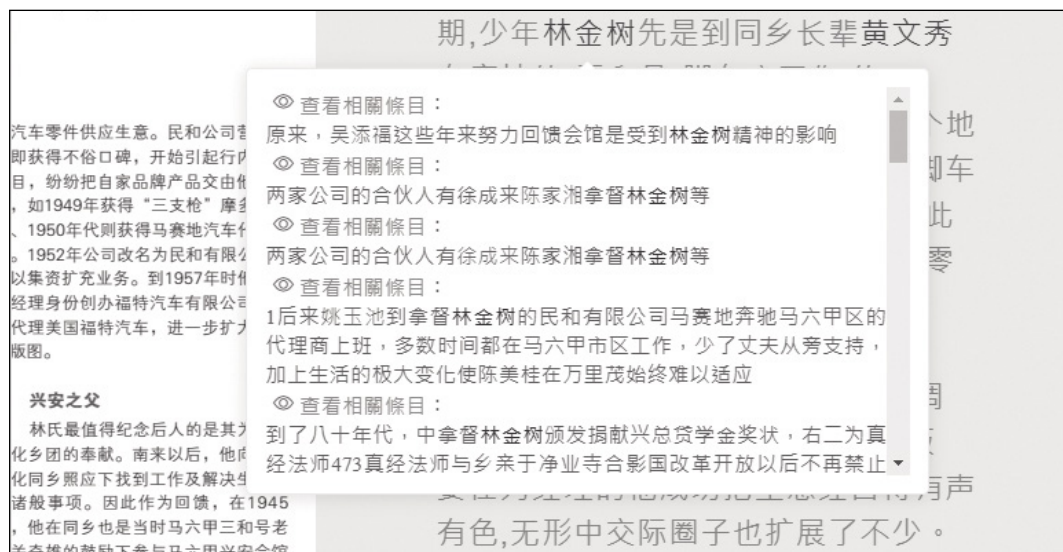


圖 11 標註人物查詢

三、地理資訊模組

近年廣泛應用於各個層面的地理資訊系統 (Geographic Information System, GIS)，可以打破純粹的文本分析限制，得以對文本進行「空間分析」(spatial analysis)，以利於進行基於空間資訊之觀察、發掘、思考與分析社會現象 (鄧志松、裘怡婷，2009)。「馬來西亞興化群賢錄」一書中，詳細記載關於每一位人物詳細之地理資訊，本平台除了將人物之籍貫、出生地等資訊記錄在後設資料欄位外，也結合地理資訊系統，將該人物之地理資訊以視覺化方式呈現，如圖 12 所示。使用者亦可以自由切換不同圖層樣式進行空間資訊解讀，此一功能亦使用於群賢瀏覽功能中的各領域頁面下，可以幫助使用者快速掌握該領域人士之地理資訊分布，如圖 13 所示。



圖 12 地理資訊模組介面



圖 13 群賢瀏覽政治領域相關人物之籍貫地理資訊分布

四、社會網絡分析模組

如前述段落所提及之「馬來西亞興化群賢錄」中的人物實體資訊，對於使用者解讀文本是相當重要的線索，因此如果能更進一步判斷這些人物實體之間是否具有潛在的關係屬性，

進而產生人物關係網絡，將更具輔助人物關係解讀應用價值。因此，本平台發展社會網絡分析模組，建立「馬來西亞興化群賢錄」中的人物網絡，並從人物實體間的文本內容推斷其關係屬性為何，最後以視覺化方式呈現「馬來西亞興化群賢錄」中的人物社會網絡關係圖，功能流程如圖 14 所示。本功能採用流水線 (Pipeline) 方式抽取人物關係，首先與閱讀介面中的人名識別功能流程相同，將文本透過 BERT 轉換成詞向量後，進行人物實體命名辨識，作為人物關係抽取的前置作業，接著將文本內辨識出的人名實體，成對輸入本平台預先訓練好的人物關係分類器，進行人物關係預測，目前分類器將人物關係初步分為夫妻、手足、親子、朋友、師生、親戚、工作及“無關係”共八類，若預測結果非“無關係”類別，則將結果存入資料庫，並以視覺化方式呈現於介面中。社會網絡分析關係介面如圖 15 所示，圖中所呈現的資訊為該條目下抽取出的人物關係網絡，使用者可點擊網絡圖中的人物節點，直接進行人物全文搜尋，此一功能希望輔助使用者透過快速掌握整體「馬來西亞興化群賢錄」中的人物關係，發掘隱藏於其中的人物關係脈絡。

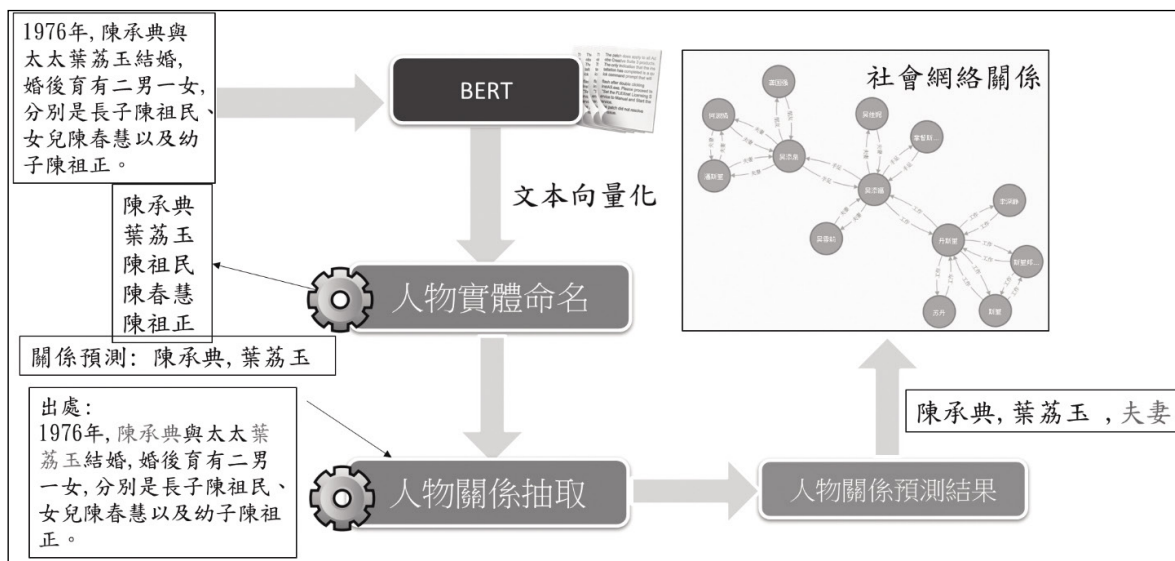


圖 14 馬來西亞興化群賢錄社會網絡分析流程圖

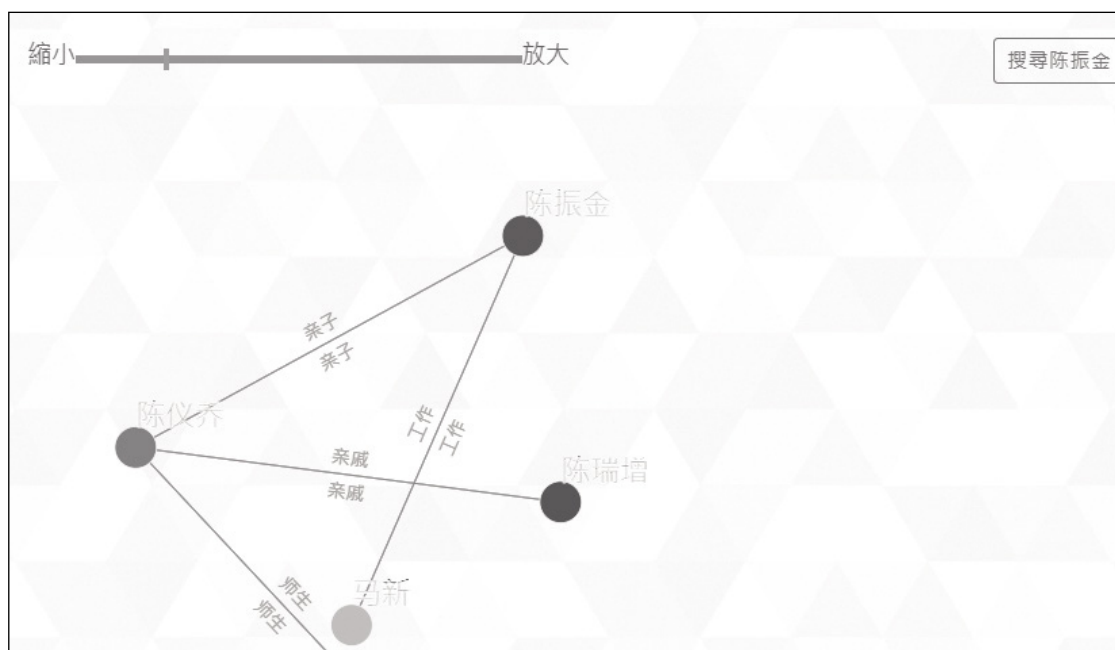


圖 15 人物社會網絡介面

陸、馬來西亞興化群賢錄數位人文平台研究案例分享

本節將分享其中一位研究者使用本平台進行探討馬來西亞華人發展，作為本平台解讀說明案例，該作者為某國立大學哲學系碩士學歷，研究興趣為馬來西亞華人政治發展探討，以下簡述該研究者之解讀過程。

該研究者首先透過平台中的各界群賢的分類功能進入軍事分類，該分類下有陳平、涂耐冰等人，使用者閱讀文本內容發現陳平與涂耐冰二人都與抗日事業和中共有關。

從陳平條目下人物關係社會網絡圖，研究者知悉陳平、馬共與中共領袖鄧小平之間存在一定關聯，如圖 16 所示，經細讀文本及外部查詢相關得知，陳平為馬來西亞共產黨之重要人物，並擔任過馬共總書記。約 1960 年代，毛澤東掌權時期，陳平受到時任總書記的鄧小平鼓勵，給予資金及武裝援助，對抗馬來西亞政府，而至約 1980 年代，鄧小平掌權，因「戰略轉移」為由，宣布停止輸出革命，停止對東南亞共產游擊隊的一切支援，馬共最後缺乏中共的支援和國際冷戰結束之際，於 1989 年停止所有軍事行動和平解散，陳平也流亡海外。從陳平的條目中，可以一窺中共對馬共從事游擊戰的支持到轉而冷漠之過程。

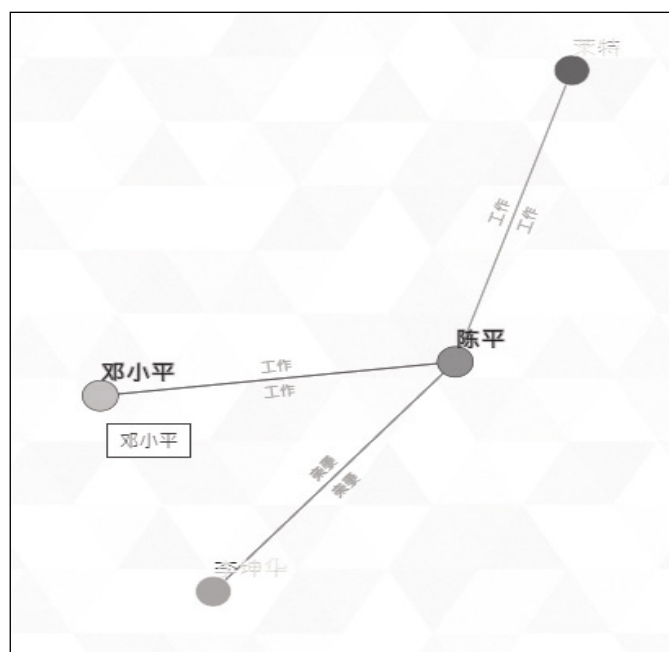


圖 16 陳平與鄧小平之人物關係網絡

至於涂耐冰，研究者一開始先透過人物社會網絡圖，發現他與二戰前後時期著名的陳嘉庚有關聯，如圖 17 所示，從涂耐冰條目的全文得知，1938 年時任南洋總會陳嘉庚會長、涂耐冰則是擔任南洋總會的秘書長一職。此外人物社會網絡圖中出現一位名叫黃慶昌人士，和陳嘉庚、涂耐冰相連，經細讀文本後，1939 年，陳嘉庚發出南洋總會第六號公告，號召成立南僑機工團回大陸服務，主要的計畫執行就是涂耐冰、黃慶昌兩人，而兩人之間的關係於砂勞越日報的創刊及古晉興安會館的成立上亦有重要的影響力。

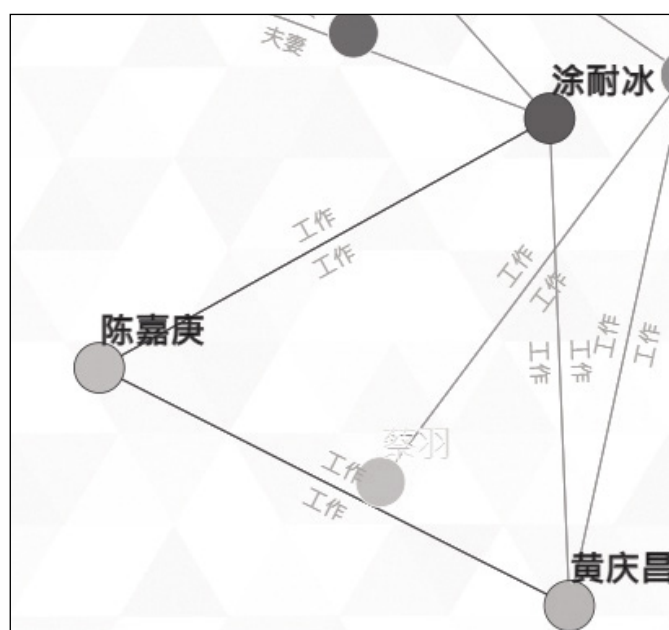


圖 17 涂耐冰、陳嘉庚與黃慶昌之人物關係網絡

透過此研究案例可以看出本平台所提供之功能，能幫助研究者找出隱藏於文本中之脈絡，如黃慶昌雖不是條目中的人物，但經由社會網絡分析功能，發現此人在南洋抗日的歷史上，也是相當重要的角色。而本平台也有助於研究者從南洋的角度出發，從各方面探討南洋華人發展與近代中國之間的關係。

除上述研究案例外，仍有許多潛在的研究議題，從本地研究的角度，可以關注個別籍貫人士所從事的經濟活動進行彙整，以對其他籍貫的人士做比較，這涉及馬來西亞獨立前後華人經濟活動與職業分配。如腳踏車行業、礦業等，多數商業活動都是由華人所從事。宗教方面，多數禪師活躍於清末民初時期下南洋，對各位禪師活動經歷的觀察與整理，透過地理資訊模組及社會網絡分析模組輔助，也許可梳理出佛教在南洋傳承與發展的路徑，而陳秉忠傳教士的帶領大批興化地區教友經新加坡至馬來西亞的詩巫地區開墾作業，及作為新馬地區移民的先驅者之一，他如何於扎根並融入當地，這些都是值得探索之議題。

柒、結論與未來方向

「馬來西亞興化群賢錄數位人文研究平台」除了提供友善的閱讀介面、輔助文本解讀工具，亦可藉由地理資訊模組或社會網絡分析模組的輔助，從宏觀的角度，以兼具時空及人物關係方式探索文本與人物關係脈絡。此外，本文亦透過平台研究案例分享提出應用此一平台進行人物關係之可能解讀模式，也提出幾點可進一步探討的馬來西亞華人人物研究議題。未來將持續蒐集、擴充馬來西亞華人相關史料，也進一步強化以下平台之功能發展：

一、鏈結資料形式串聯平台內部及外部資源

若能將數位人文平台中內部所典藏龐大、複雜、分散且具關聯的文本內容進行相互鏈結，並從中計算與該文本具有相似概念的內容資訊，將具有較高相似度的文本內容以鏈結方式，呈現給使用者，亦是輔助數位人文研究之重要技術發展議題。若能如此，不僅能幫助使用者取用相關聯的文本資訊，亦能節省其尋找關聯資訊所花費的時間。

此外，透過研究解讀案例可以發現，人文研究者在解讀資料時，時常需要參照其他的相關資源，以協助其理解文本中的脈絡或進行驗證。因此，未來平台將整合 Chen 和 Liu(2019) 所發展之自動標註工具，以鏈結資料 (Linked Data) 方式，自動的從多個外部來源資料庫取得資料，讓使用者在閱讀時，能即時參考針對閱讀文本之不同來源補充資訊，讓使用者更快速理解文本的內容。

二、知識圖譜之建立

在大量的史料與文本內容中，往往蘊含著複雜的實體網絡關係，這些實體可能包含人物、機構、事件與地點等。在過去研究者欲釐清其中的脈絡，往往需要耗費極大的時間與精力。目前本平台只初步進行人物與人物關係之間的抽取，未來預計將實體命名的類型拓展至機構、事件、時間等，並抽取各實體彼此之間的潛在關係，整合多維的網絡關係，並透過資訊視覺化技術建立完善的知識圖譜，並且隨著文本中時序的推移，知識圖譜中所呈現的內容也會有所變化。

三、具個人化之平台收藏庫與分析功能

本平台預計發展個人化之平台收藏庫與分析功能，開發我的收藏庫 (My Collection) 功能，克服文本內容與工具綁死之限制，允許使用者將感興趣之條目加入個人收藏庫，同時支援上傳自身所典藏之文本，結合平台本身已具有的豐富內容，以及使用者個人擁有的內容，建立個人化的文獻集，並可根據自身需求，佈建合適分析個人收藏庫的數位分析工具，如此將能發展符合出人文學者需求且具彈性之數位人文平台。

致謝

此平台建置係由教育部高教深耕計畫特色領域中心之「華人文化主體性研究中心」經費補助 (經費代碼 109H21)

參考文獻

- 中央研究院數位文化中心 (2020)。數位人文研究平台。上網日期：109 年 10 月 11 日，檢自：<https://dh.ascdc.sinica.edu.tw/member/>
- 中國歷代人物傳記資料庫。CBDB querying and reporting system -online。上網日期：109 年 9 月 18 日，檢自：<http://db1.ihp.sinica.edu.tw/cbdb/help/systemintro.html>
- 杜協昌 (2018)。DocuSky：個人文字資料庫的建構與分析平臺。數位典藏與數位人文，2, 71-90。
- 金觀濤、邱偉雲、梁穎誼、陳柏聿、沈錫坤、劉青峰 (2016)。觀念群變化的數位人文研究 -- 以《新青年》為例。數位人文：在過去、現在和未來之間。
- 翁稷安 (2016)。從「機構導向資料庫」到「個人導向資料庫」：數位人文下一階段的可能發展。第七屆數位典藏與數位人文國際研討會，臺北。
- 陳志銘、張鐘、徐志帆 (2020)。羅家倫先生文存數位人文研究平臺之建置與應用。數位典藏與數位人文，5，73-115。

項潔、涂豐恩 (2011)。導論—什麼是數位人文。從保存到創造：開啟數位人文研究，9-28。

趙薇 (2016)。「社會網路分析」在現代漢語歷史小說研究中的應用初探——以李劫人的《大波》三部曲為例。數位人文：在過去、現在和未來之間。

劉吉軒、柯雲娥、張惠真、譚修雯、黃瑞期、甯格致 (2012)。以文本分析呈現海外史料政治思想輪廓。數位人文要義：尋找類型與軌跡。頁 83-114。

鄧志松、裘怡婷 (2009)。GIS 中國經社指標網路資料庫：建置與推廣。地理資訊系統季刊，3(4)，15-22。

謝博宇 (2016)。以 DocuSky 為核心的工具開發與建置。第七屆數位典藏與數位人文國際研討會，臺北。

謝順宏、柯皓仁、張素玢 (2018)。自動分群應用於傳記人物關係建立。「第九屆數位典藏與數位人文國際研討會」發表之論文，法鼓文理學院。

Aparecido Ribeiro, M., Vosgerau, R. A., Larissa Pereira Andruchiw, M., & Ely de Souza Pinto, S. (2015). The complex social network from The Lord of The Rings. ArXiv E-Prints, 1606, arXiv:1606.02610.

Chen, C. M. & Chang, C. (2019). A Chinese ancient book digital humanities research platform to support digital humanities research. The Electronic Library, 37(2), 314-336.

Chen, C. M., Chen, Y. T., & Liu, C. Y. (2019). Development and evaluation of an automatic text annotation system for supporting digital humanities research. Library Hi Tech, 37(3), 436-455.

Chen, K. C., & Chen, B. Y., & Huang, C. C. (2018). Who was the key figure? A social network analysis of Suzhou book collector groups in Ming Dynasty through a digital humanities approach. Journal of Educational Media & Library Sciences, 55(3), 243-284.

Correll, M., Alexander, E., Albers, D., Sarikaya, A., Gleicher, M. (2014). Navigating reductionism and holism in evaluation. In Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization, pp. 23-26.

Drucker, J. (2013). Intro to digital humanities: Introduction. UCLA Center for Digital Humanities. Web available at http://dh101.humanities.ucla.edu/?page_id=13, Retrieved January 28, 2018.

Duling, D.C.(2000). The Jesus Movement and Social Network Analysis: (Part II. The Social Network). Biblical Theology Bulletin. Vol 30, Issue 1, pp. 3 -14

Franco Moretti. (2011). Network theory, plot analysis. New Left Review.

Ho, H. I. B. & Hilde, D. W. (2014). MARKUS text analysis and reading platform. Retrieved September 11, 2020, from <https://dh.chinese-empires.eu/markus/beta/>

Hockey, S. (2004). The history of humanities computing. A Companion to Digital Humanities, 3-19.

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2017). Visual text analysis in digital humanities: Visual text analysis in digital humanities. Computer Graphics Forum, 36(6), 226-

250.

Kestemont, M. and Gussem, J.D. (2017). Integrated sequence tagging for medieval Latin using deep representation learning. *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages. available at: <https://jdmdh.episciences.org/3835>

Moretti, G., Sprugnoli, R., Menini, S., & Tonelli, S. (2016). ALCIDE: Extracting and visualising content from large document collections to support humanities studies. *Knowledge-Based Systems*, 111, 100-112.

So, R. J. & Hoyt L. (2013). Network Analysis and the Sociology of Modernity, *boundary 2*, 40(2), 147-182.

Stanford University (2013). Retrieved September 16, 2020, from <http://republicofletters.stanford.edu/>

Sun, Z. & Wang, H. (2010). Overview on the advance of the research on named entity recognition. *New Technology of Library and Information Service*, 26, 6, 42-47.

TBDB 臺灣歷史人物傳記資料庫 (2020)。TBDB 臺灣歷史人物傳記資料庫。上網日期：109 年 9 月 11 日，檢自：<http://tbdb.ntnu.edu.tw/index.jsp#about>

Tu, A., Hung, J.-J., & Lin, Y.-H. (2012). Building a text analysis platform for Chinese Buddhist text -An example based on CBETA and Tripitaka Catalog Projects. Presented at the PNC 2012 Annual Conference and Joint Meetings, UC Berkeley School of Information, USA. Retrieved from <http://joeyhung.info/publications/>

Yohan, P.M., Sasidhar, B., Basha, S., A.H. and Govardhan, A. (2014). Automatic named entity identification and classification using heuristic based approach for Telugu. *International Journal of Computer Science Issues*, Vol. 11 No. 1, pp. 173-180.

【子題三】

圖書館之數位人文創新服務建置與應用

Session III

Innovative Services for Digital Humanities in Libraries



肖珑
Long Xiao

肖珑，北京大学图书馆研究馆员，山西大学图书馆馆长，研究生导师，兼任国际图书馆协会联合会（IFLA）知识管理专业委员会委员、山西图书馆学会副理事长、中国图书馆学会学术委员会用户与研究专委会副主任委员、中国社科情报学会常务理事，《图书情报工作》、《大学图书馆学报》等多个核心期刊 / 专业期刊编委。【主要研究领域】信息资源建设与共用，图书馆服务与管理，数字图书馆建设及标准规范，图书馆空间布局等。已出版专著 / 教材 16 部，发表中英文论文 110 余篇，完成 1 项国家标准。

Long Xiao is Professor of Peking University Library, Director of Shanxi University Library and Standing Committee member of IFLA Knowledge Management Section. Her main research areas focus on the following: information resource development and sharing; user services; library management; digital library standards and criteria; knowledge organization and management; library spaces. She has achieved several important researching projects of National Natural Science Foundation of China, Ministry of Science & Technology, Ministry of Education, published 16 books and more than 110 papers in Chinese and English.

古文献數位化中的知識體系建設——兼論數位人文之應用

The Construction of Knowledge System in the Digitization of Ancient Documentation -- With a Study on the Application of Digital Humanities

肖瓏

Long Xiao

北京大學圖書館研究館員

山西大學圖書館館長

Professor, Peking University Library

Director, Shanxi University Library

摘要

古文献数字化，是指基于古籍再生性保护和传播的目标，采用电脑技术，将古籍、拓片等古代文献进行数字化加工、处理，从而制成古文献目录数据库、全文数据库、图像数据库等，并揭示其内容信息、系统性建立古文献相关知识体系的工作。

本文在回顾“中华古籍保护计划”十多年建设基础上，论述古文献知识体系的重要性，从历史纪年、语言学、文字学、书史、版本学、编目学、地理资讯系统等方面，讨论在古文献数字化过程中知识体系的建设目标和方法，兼述数字人文在其中发挥的作用。

关键词：古文献数字化；知识体系；数字人文

Abstract

Basing on the goal of regenerative protection and dissemination of ancient documents, the digitization of ancient documents refers to digitize and process ancient collection (such as rare books, rubbings, etc.), build catalog database, full-text database, image database, and reveal their content information, systematically establish a related knowledge system.

On the basis of reviewing the developments of the "Conservation of Ancient Book Collection" for more than ten years, this paper clarifies the importance of the knowledge system of ancient documents. Specially from various aspects such as historical chronology, linguistics, philology, book history, edition studies, cataloging, geographic information

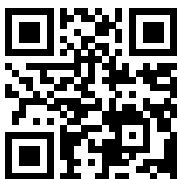
system, etc., it discusses the construction goals and methods of the knowledge system in the digitization of ancient documents, as well as describes the role of digital humanities in it.

Keywords: Ancient Collection Digitization, Knowledge System, Digital Humanities

※ 本場次演講，作者只提供摘要

預錄及簡報檔

<https://pse.is/3e37pp>





林信成
Sinn-Cheng Lin

林信成，臺灣大學電機工程博士，現為淡江大學資訊與圖書館學系（簡稱資圖系）教授，曾任淡江大學資圖系主任、文學院院長。致力於數位典藏、數位人文領域之教學與研究，為臺灣棒球維基館、臺灣籃球維基館、療癒閱讀維基館、淡水維基館等數位共筆網站之創辦人。

Sinn-Cheng Lin, Ph.D in Electrical Engineering, National Taiwan University, now a Professor in the Department of Information and Library Science, Tamkang University. He used to be the Dean of College of Liberal Arts, Chairman of Department of Information and Library Science and Director of Publishing Center, Tamkang University. His research fields focus on Digital Humanities, Digital Archives, Digital Preservation, Digital Collaboration and Collective Intelligent, etc. He is also the founder of digital collaboration websites such as Tamsui Wiki, Taiwan Baseball Wiki, Taiwan Basketball Wiki, Healing Reading Wiki, etc.



范凱婷
Kai-Ting Fan

范凱婷，淡江大學中國文學系博士，現為淡江大學資訊與圖書館學系博士後研究員。

Kai-Ting Fan, Ph.D. in Chinese literature, Tamkang University. Now she is a Postdoctoral Researcher in the department of Information and Library Science, Tamkang University .

地方學數位共筆平台之資料度用與數位策展

Data and Digital Curation for the Digital Collaboration Platform of Local Historical Studies

林信成

Sinn-Cheng Lin

淡江大學資訊與圖書館學系教授

Professor, Department of Information and Library Science, Tamkang University

范凱婷

Kai-Ting Fan

淡江大學資訊與圖書館學系博士後研究員

Postdoctoral Researcher, Department of Information and Library Science,
Tamkang University

摘要

人類文明因著時代、自然條件、地區等因素而有不同的演進，並發展出多元且具地域差異的文化特色。圖書館、博物館或檔案館作為一個典藏機構，長期以來都擔負著文化的保存、收藏、展示予大眾的義務與責任。然而因應當前數位化的時代，大眾對文化的傳遞與期待不再只是典藏與被動的查找，在眾多的資料中如何利用數位工具讓讀者能更快汲取資訊，都在考驗當前的文化記憶機構。

本文試圖以「地方知識學數位共筆平台—淡水維基館」的建構與展覽經驗為例，從作為一個收集、組織、整理淡水地方知識學的數位博物館中，思考數位人文如何能幫助地方文化創生，並且更貼近社會大眾，以更有效的方式傳遞大淡水在地知識，以達人文科學教育普及之目標。為了解其中之關鍵，本文以淡水維基館所涵蓋的大淡水地方學之資料度用為思考進路，從中規劃 VR 虛擬實境的數位度用與策展，並輔以行動故事車等實體展覽活動，以此發展出可供參考之人文科學普及推廣之經驗。

關鍵字：淡水地方知識學、淡水維基館、資料度用、數位人文、虛實策展

Abstract

Human civilization has evolved differently due to factors such as times, natural conditions, and regions, and has developed diverse cultural characteristics with regional differences. For a long time, libraries, museums or archives, as collection institutions, have been responsible for cultural preservation, collection, and exhibition to the public. However, in the digital age, the people's transmission and expectation of the culture is no longer just collection and passive search. Therefore, how to use digital tools to allow readers to acquire information more quickly in numerous materials will test the current cultural memory institutions.

This article attempts to take the construction and exhibition experience of the "Tamsui Wiki" as a digital curation example. Thinking about how to go from being a digital museum for collecting, organizing, and organizing Great-Tamsui local knowledge to helping Regional Revitalization with digital humanities, and get closer to the general public, spread the knowledge of Great-Tamsui area in a more effective way to achieve the goal of popularizing humanities education. We uses the data curation of the "Tamsui Wiki" as a starting point, thereby planning the digital curation with Virtual Reality, and supplemented by physical exhibition activities such as mobile story car, such that the experience of popularization and promotion of humanities can be developed for reference.

Keywords: Tamsui Local Historical Studies, Tamsui Wiki, Data Curation, Digital Humanities, Virtual Reality Curation

壹、前言

地方學是一門古老的學問，先秦時代《周禮·地官》中提到：「掌道方志，以詔觀事；掌道方輿，以詔辟忌，以知地俗。¹」即指：掌握地方的地理可以知道地方樣貌，掌握其災害及變化可以知道地方風俗。「地方志」將地方歷史、地理、人物、自然、經濟發展等內容紀錄造冊，一直是過去行政官員了解地方事務，以及治理地方的重要工作。明清以後地方的鄉志或鎮志撰寫，以地方官統籌召集地方文人仕紳撰寫的模式為主，例如清代《淡水廳志》由淡水廳同知陳培桂邀請侯官舉人楊浚代為纂輯，集合過去竹塹進士鄭用錫的《淡水志初稿》與金門舉人林豪的《淡水續志稿》於同治十年（1871 年）完成刊行²。「地方志」這樣半官方記錄區域人事、地物、歷史的傳統延續到了現代，遂成為地方學研究的知識型態。

1980 年代後期的臺灣，史地相關學門的研究者意識到地方鄉土探索的意義、區域研究的價值，紛紛開始出現「地方學」的概念³，而在臺灣地方學的研究中「淡水學」是相對較早被提出的地方學範疇⁴。淡水是一個承載著豐富人文知識的地區，從淡水地區出海的淡水河作為北臺灣重要的水系而言，淡水不僅在水文上匯流了發源自新竹、桃園、雙北、基隆的大漢溪、南勢溪、新店溪、基隆河，更在北臺灣開發歷史上佔有舉足輕重的地位。早在十七世紀各國海權國家看重北臺灣的地理位置，西班牙人、荷蘭人都曾在淡水與基隆經營貿易與傳教的據點。明鄭時期，鄭經派兵北上攻打荷蘭人，結束了荷蘭人在臺灣的佔領，但也因為此時與西班牙、荷蘭、日本、中國等國的往來貿易，臺灣已然躍上國際舞臺。清咸豐八年（1858 年）清朝與英、法兩國簽訂《天津條約》開放淡水港通商後，淡水成為國際商船往來頻繁的國際商港。商船自淡水河口登陸後，沿著河岸深入臺北盆地中心，例如艋舺、大龍峒、大稻埕等地區，當時北臺灣居民透過淡水河水系的水路交通從事農業、商業貿易行為，進而形成生活聚落。除此之外，光緒十一年（1884 年）清朝與法國因為安南主權問題而在臺海引發戰役，一路從中國東南沿海打到雞籠、滬尾，直到今日都還可以看到當時戰爭所留下的影響。淡水河的發展直到淡水河港、艋舺河岸淤積，日治時期北臺灣航運重心才逐漸移往基隆港。有了這些故事，使得淡水成為一個迷人的城市。

-
1. 〔漢〕鄭玄注、〔唐〕賈公彥疏：《周禮注疏》，臺北：臺灣古籍出版，2001 年 12 月初版，影印嘉慶二十年江西南昌府學刻本。
 2. 〔清〕陳培桂：《淡水廳志》，臺灣文獻叢刊第 172 種，臺北：臺灣銀行經濟研究室。
 3. 洪健榮，〈大臺北地方學研究的回顧與展望（1990-2013）——以地方志書與學位論文為中心〉，《輔仁歷史學報》，第三十六期，2015 年 3 月，頁 285-333。
 4. 洪健榮，〈大臺北地方學研究的回顧與展望（1990-2013）——以地方志書與學位論文為中心〉，《輔仁歷史學報》，第三十六期，2015 年 3 月，頁 285-333。

地方學在經過近四十年的發展，其中研究除了從學院內臺灣史區域研究中分支出不同區域的地方學外，另一方面由於「大眾史學」風氣漸盛，地方文史工作者、社區團體與社區大學等組織亦同樣進行地方文史資料搜集與保存工作，兩者皆給予地方學相當的養份。然而，學界與民間地方學的資料統整，長時間卻有賴於地方政府的補助與主導，彼此之間的關聯及利用程度並不高，且礙於地方政府行政區的劃分，地方學的研究始終無法擴及更大的範圍，而錯失區域整合的研究，再者，地方學研究雖有地方文史工作者的加入，然而研究成果卻無法真正普及於一般大眾。王御風先生(2011)在考察各地地方學發展的現況後提到：「唯有整合各方力量，如大學與社區大學合作，統合學院內學者與地方文史工作者對地方知識的看法，才能讓地方學跳脫現有的格局，持續發展。⁵」因此，我們認為：地方學的研究除了文史地方考察外，更為重要的是能有一個資訊的整合平台。歷史學者林素玢教授(2020)提出：以「數位共筆」方式取代傳統方志的聲音時有所聞。這種融合數位科技、田野調查研究、社群經營、影像敘事等等元素建置數位資料庫的方式，在透過數位協作平臺由社群共筆，亦可由全民一起來寫方志，目前較成氣候的有「淡水維基館」。⁶其中所提及的「淡水維基館」是由本文作者林信成教授於2013年所創建，以Wiki協作系統(Wiki Collaboration System)為主的一個書寫平台。Wiki是一種以Web為基礎，可自由擴展且相互連結的資料庫系統，每個頁面皆能開放使用者編輯，只要藉由Web瀏覽器即可進行協同創作(Collaboration)⁷。「淡水維基館」由於具開放性特色，提供了地方學研究者一個可以互通彼此資訊的平台，以達到共筆的目的；同時開放式系統也多少解決過去地方政府僅能以行政區為框架來書寫地方志的限制。因此，本文在「淡水維基館」作為地方學數位共筆平台的現有基礎上，進一步進行主題式資料度用與數位策展，以期了解數位共筆平台資料的加值與再利用可能得到什麼樣的成效？

貳、文獻探討

一、從數位典藏到數位共筆

隨著網路科技的發展，資源分享的途徑與管道亦日益增加，而以資源共享(Resource Sharing)為核心價值的數位典藏，更藉由網路平台得以迅速擴大。數位典藏概念意即：將具有蒐藏價值的實體或非實體資料，運用數位技術加以保存及應用，進而透過網路加以傳

5. 王御風，〈地方學的發展與挑戰〉，《思與言》，第49卷第四期，2011年12月，頁31-55。

6. 林素玢：〈傳統與創生：文化進程中的方志纂修〉，《臺灣史研究》27卷1期2020年3月，頁155-185。

7. Bo Leuf and Ward Cunningham, The Wiki Way: Quick Collaboration on the Web, (Addison Wesley, 2001).

播⁸。更具體來說，數位典藏是將各式各樣具有文化保存價值的藏品或活動，包括器物、書畫、文件、典籍、建築、標本、藝術表演、民俗活動……等，轉化為數位內容形式，如：文字、圖片、影片、聲音、動畫……等，儲存於數位典藏系統中，以利使用者管理、取用、展示、傳遞各類型的數位典藏品，有助於人們了解先民的風俗習慣、生活樣態、社會組織、文化工藝、生命演化等，達到教育、學習、娛樂與研究目的。⁹ 過去二十年來有許多國家發展數位典藏技術，其目的大多以透過科技將具有教育與文化意義的典藏品數位化方式進行建檔與保存，除了保存，更希望能藉此讓使用典藏品的使用者能有後續加值利用的機會。

而在數位加值利用上，「數位共筆」(Digital Collaboration) 首要符合了數位典藏的概念外，再者，其重要的核心精神在於群體智慧 (Collective Intelligence) 的展現。Wiki 協作系統 (Wiki Collaboration System) 是常見的數位共筆平台，Wiki 系統最早是 1995 年由 Ward Cunningham 所設立的「維基維基網」(WikiWikiWeb)^{10, 11}，其後 Jimmy Walse 於 2001 年應用 Wiki 創建了為人所熟知的維基百科 (Wikipedia)¹²。本文作者林信成教授所創建的「淡水維基館」，亦是藉由 Wiki 系統建立的全開放共筆平台，以提供對淡水知識有興趣的所有大眾使用，「淡水維基館」的使用者可以自由地查找與大淡水地區有關的知識頁面，同時也能書寫編輯增補系統上的資料，以此展現群體智慧與大眾史學精神。

二、資料度用與數位策展

資料度用 (Data Curation) 在伊利諾大學圖書館與資訊科學研究院的定義中指出：「資料度用 (Data Curation) 是一種動態的資料管理過程，是對使用者感興趣和有用的資料，依其生命週期 (life cycle) 加以進行積極和持續的管理。¹³」從字面上而言，Data 指的是資料，Curation 是度用；而 Curation 這個字源自於拉丁文 cura，有照料 (to take care) 之意，而在資料度用的概念下，度用便是一種動態管理的歷程，為的是讓使用者感到興趣以及自覺發現資料的有用。而與 Data Curation 相似的另一個詞為 Digital Curation，則被譯為數位策展，其概念在 Digital Curation Centre (DCC) 的定義是：「在數位資料的

8. 蔡永橙、邱志義、黃國倫等著：《數位典藏技術導論》(臺北市：國立臺灣大學出版中心、中央研究院，2007 年)。

9. 林信成：《開放式數位典藏系統之研究》(臺北：文華圖書管理，2010 年)，頁 6。

10. Ward Cunningham, "WikiWikiWeb," <http://wiki.c2.com/?WikiWikiWeb> (accessed Nov. 29, 2020).

11. Ward Cunningham, "Portland Pattern Repository," <http://wiki.c2.com/?PortlandPatternRepository> (accessed Nov. 29, 2020).

12. Wikipedia, "Welcome to Wikipedia," <http://en.wikipedia.org> (accessed Nov. 29, 2020).

13. Illinois School of Information Science, "Data Curation," <https://ischool.illinois.edu/research/areas/data-curation> (accessed Nov. 29, 2020).

生命週期中，對其進行管理、保存和加值利用。¹⁴」意即在數位資料的生命週期中除了有「保存資料」的管理過程外，同時對資料作「加值利用」直到資料的生命週期結束為止。兩相比較可發現，數位策展的概念其實也包含了部分資料皮用的意義，因此在對數位資料管理的過程中，從數位典藏到數位策展即是利用現代數位科技加以保存、展示，並與他人分享、交流的過程，以達到「加值利用」的目標，同時能達到資料生命週期的延長與被再利用的可能。

然而不論是「皮用」或是「策展」都牽涉到一個更為關鍵的問題：他們所保存與加值利用的內容為何？是如何被決定的？策展人（Curator）即是其中關鍵的角色。過去歐洲對於策展人的解釋認為：策展人任務兼負保存守護、研究論述以及教育宣導三項要務，專業策展人具典藏管理、統籌督導和傳播教育之責。¹⁵而黃靖斐與陳志銘（2018）則是從現當代數位化的時代出發，認為策展人廣義而言是指在網路上不斷查找、篩選、分類、組織和分享特定問題最相關內容的資訊處理者，然而策展人不能只是資料的收集與編輯者，因此策展人還需要針對特定主題，通過查找、組織、詮釋以達到共享有意義、且相關、高品質的內容。¹⁶因此，策展人的行動更牽動著資料的管理收藏、詮釋教育、知識普及甚而有文化傳承的工作。

參、《淡水維基館》之資料皮用與虛實策展

一、推動協同合作共建數位內容

《淡水維基館》建置的目的在於提供一處線上的網路社群協作平台，讓大眾共同記錄淡水河流域、海域及其週邊地區的歷史文化、地景地貌、古蹟建築、城鄉聚落、風土民情……等，以協助地方學和臺灣史之學術研究與人文普及教育，達到文化資產數位保存、行銷推廣與加值利用¹⁷。平台的內容建置以淡江大學「文五合 e」跨領域研究團隊為主¹⁸，整合淡江大學社會責任計畫中之地方社區、各級學校與政府機關等合作單位，藉由多元課程與研究計畫整合，如：數位人文、歷史研究、田野踏查、編輯出版、程式設計、電子書製作、新媒體設

14. Digital Curation Centre, "What is digital curation?", <https://www.dcc.ac.uk/about/digital-curation> (accessed Nov. 29, 2020).

15. 林廷宜：〈專業策展人才養成：以美國藝術育才策略為例〉，《設計學報》第 18 期第 4 卷（2013 年 12 月），頁 23-40。

16. 黃靖斐、陳志銘：〈數位內容策展在大學校史館的協作與應用〉，《圖書與檔案學刊》第 10 卷第 2 期（2018 年，12 月），頁 75-106。

17. 〈淡水維基館首頁・關於〉<https://tamsui.dils.tku.edu.tw/wiki/index.php/%E9%A6%96%E9%A0%81>（瀏覽日期：2020/12/19）

18. 淡江大學「文五合 e」研究團隊，初期由文學院資圖、歷史、中文、大傳、資傳等五系師生所組成，後來逐漸擴及不同學院。

計等類型的課程，共同產出大淡水相關之數位內容，並將資訊科技應用於地方文化之知識建構與探索，達到藉由數位工具輔助在地人文、歷史、文化之典藏、教學、傳播與研究，並進行 e 化整合之目的。

自 2013 年開站以來至 2020 年 10 月為止，《淡水維基館》的使用者數量即呈現逐步上升趨勢，而網站的總瀏覽量目前已達到 226 萬餘次，使用者總數約有 90 多萬人，分布於全球 145 個國家。¹⁹

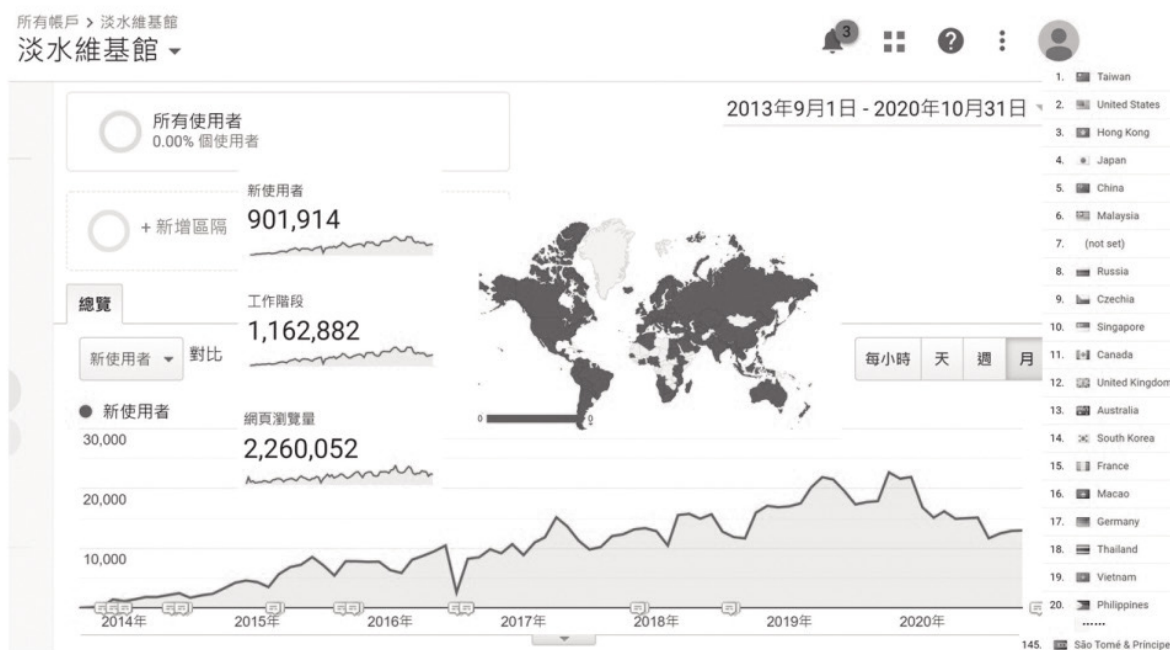


圖 1 本系統自 2013 年 9 月至 2020 年 10 月為止之使用概況

由此可見，藉由數位科技的輔助，《淡水維基館》已有效的將淡水地方知識學推向全世界。然而《淡水維基館》這樣的數位共筆系統雖然相較於實體圖書館、博物館等實體典藏機構，可不限時間空間限制的服務使用者，卻與實體典藏機構面臨相同的困境，皆是被動等待使用者上門。

因此，《淡水維基館》數位典藏淡水地方知識的下一步便是從資料的度藏中走向加值再利用的可能，以其設立的目標為依歸，為了能讓淡水地方學和臺灣史之學術研究成果能更貼近大眾，以及推廣地方文史人文研究知識的普及教育，《淡水維基館》自 2019 年起，從平台上累積的 1 千 8 百多頁數位內容與 3 千 6 百多幅圖片，由本文作者林信成教授擔任策展人，策畫 VR 園區定點展覽及行動故事車巡迴展覽，將淡水地方知識學資料適度利用，結合現代新興科技，以期能加長加廣淡水地方知識學資料的生命週期。

19. 本數據根據 Google Analytics 進行使用者大數據分析，瀏覽日期：2020/10/31。

二、開發 VR 園區進行虛實策展

虛擬實境 (Virtual Reality, 簡稱 VR) 是近年來的新興科技，藉由 3D 建模、環場全景、頭戴式顯示器等設備，令觀眾有如親臨現場的感覺。本文所稱「虛實策展」的概念，是由「數位策展」加上「虛擬實境」延伸而來，兩者相同之處在於策展概念同樣以數位工具為展覽方式之一，藉由策展人的選擇與安排將其所欲展覽的內容展示於觀眾面前。然而虛實策展不同於數位策展的是在於展覽模式上，「虛實」策展除了是將典藏資料再利用外，其中強調虛擬的線上展覽內容與真實情境的展場規劃，需要有一定程度的配合展示，在實際的展示內容中再延伸出虛擬展覽，以實體展覽整合虛擬展覽達到相輔相成之效。

而「淡水維基館 VR 園區」便是一個數位、實體兼具的虛實策展計畫，採用 VR 技術將《淡水維基館》數以千計的數位內容，擇定主題進行虛實整合與數位策展，讓更多人了解當地的歷史脈絡和文化底蘊。自 2019 年起至今，已完成兩檔 VR 程式的開發及展覽：(1) 淡金海岸石滬群 VR 特展、(2) 河海山城老淡水：蔡坤煌攝影 VR 展。分述如下：

(一) 《淡金海岸石滬群 VR 特展》

第 1 檔《淡金海岸石滬群 VR 特展》，旨在促進石滬文化資產之保存、積累與運用。石滬是早期維持漁民生計的重要設施，漁民在沿海區域，依循潮流方向堆築石塊為滬，利用魚類逐浪特性與潮間變化，以「守滬待魚」方式捕撈：漲潮時，魚群隨著潮汐游入滬中覓食；退潮時，魚群困於海水退去的石滬中，漁民不用駕船出海，只要進到石滬內，使用簡單漁具即可輕鬆捕捉漁獲。因此，石滬不但是早期漁民生活的主要經濟來源，也是人類漁業史上最早的生態工法，更是先人智慧的結晶，具備人文、歷史、建築、生態、觀光之多元價值，非常值得有組織、有系統的研究、保存、推廣與應用。澎湖周邊島嶼是目前所知使用石滬捕魚最多的區域，臺灣四面環海，但保存下來的並不多。淡水舊稱滬尾，早年因漁源豐富，在淡水河口、沙崙一帶，甚至沿著北海岸一直到石門地區，皆建有石滬，規模相當龐大。然而，隨著科技進步，動力漁船取代舊式捕魚方法，石滬漁法逐漸乏人問津；加上地方開發和破壞、沿岸漁源衰竭、漁民日漸年老無人承繼等因素，淡金海岸石滬漸漸因年代久遠、乏人維護而荒廢坍塌，如今僅剩幾處石滬遺跡。

有鑒於此，由淡江大學資圖系林信成教授、資傳系賴惠如副教授、歷史系李其霖副教授組成的跨域研究團隊，自 2014 年起以 3 年時間，應用數位科技將淡金海岸一帶具有歷史價值之石滬文化，藉由船拍、空拍、GPS 定位、紀錄片拍攝及耆老訪談等方式，

並透過軟體、硬體、資訊、內容、創意、加值的有效整合，於《淡水維基館》中建構「淡金海岸石滬探勘暨影像資料庫」，以促進石滬文化的累積、傳承與利用。如下圖所示。



圖 2 《淡水維基館》中之「淡金海岸石滬探勘暨影像資料庫」

到了 2019 年，進一步開發 VR 程式並以虛實整合策展方式，結合 VR 虛擬實境、實體海報、現場解說，於 2019/05/20~2019/08/20 在淡江大學海事博物館，以《淡金海岸石滬群 VR 特展》為題進行展覽，讓更多人能接觸並了解淡金海岸古老的石滬文化。



圖 3 《淡金海岸石滬群 VR 特展》活動

過去大家都知道澎湖有石滬，尤其雙心石滬更是聞名遐邇，卻鮮少有人注意到北臺灣的漁民亦有使用石滬捕魚的方法，於是著手策展《淡金海岸石滬群 VR 特展》，成為第一檔《淡水維基館》資料皮用與虛實策展的主題。

（二）《河海山城老淡水—蔡坤煌攝影 VR 展》

第 2 檔《河海山城老淡水—蔡坤煌攝影 VR 展》，主要將蔡坤煌醫師的攝影作品以 VR 方式展示，重現早期的淡水小鎮意象。

淡水的開發相對早於北臺灣整體發展，從大航海時代開始即有各國商船在此往來。而淡水小鎮上的生活與文化早已沁入各個歷史時期的記憶，例如明清時期福建移民所帶來的家鄉信仰，清水祖師、媽祖、觀音、王爺信仰便在此落腳；西班牙、荷蘭與英國人在此貿易留下了貿易據點紅毛城與英商洋行；淡水河上往來的漁船、商船在觀音山下穿梭。這些都成了淡水小鎮最迷人的文化資產，而早期紀錄工具選擇不多的情況下，攝影可能是較為容易的方式之一。以此，本檔展覽以展出早期攝影作品，讓觀眾能藉此重回 1960 年代的淡水，體驗老淡水的風景與文化氛圍，於是第二檔展覽選定《淡水維基館》上：「蔡坤煌先生攝影作品」為策展內容。

蔡坤煌醫師生於 1922 年，卒於 1994 年，畢生熱愛淡水與攝影，遺留下許多珍貴老照片，紀錄了五、六十年代的淡水，很多場景如今已不復存在，相當珍貴！蔡醫師獲有中國攝影學會榮譽博學會士、臺灣省攝影學會博學會士、臺北市攝影學會博學會士、美國紐約攝影學會博學會士……等榮銜。2017 年底，在蔡醫師過世 23 年後，淡水紅毛城曾舉辦「心所傾慕，阮疼的淡水—蔡坤煌醫師鏡下的淡水」攝影展，透過鏡頭可回憶五、六十年代淡水純樸風情，呈現多樣的人文地貌光景。可惜實體展場空間、展期皆有所侷限，一旦撤展，向隅者便無緣觀看，殊為可惜。數位策展便無此限制，藉由數位科技可不受時空限制，重建在地記憶、再現淡水風華，讓使用者能以更全面的方式，看待歷史文化的傳承、地方風物的遷移、地景地貌的演變 … 等。下圖所示為蔡坤煌醫師所攝淡水老照片的一部分，多年前即已取得蔡醫師之子，前淡水區長、新北市議員蔡葉偉先生同意與授權，以數位化方式，於淡水維基館網站上呈現。



圖 4 蔡坤煌醫師所攝淡水老照片

而《河海山城老淡水－蔡坤煌攝影 VR 展》，則進一步以虛實整合策展方式，結合 VR 虛擬實境、實體海報、現場解說，於 2020/3/1 ~ 2020/4/30 在淡江大學海事博物館展出，再度勾起人們對河海山城老淡水的記憶。



圖 5 《河海山城老淡水－蔡坤煌攝影 VR 展》活動

三、以 VR 行動故事車巡迴策展

在《淡水維基館》VR 園區兩檔展覽分別實施定點展覽之後，為了能更主動地接觸使用者，另以車輛搭載 VR 設備與器材，組成「VR 行動故事館」，藉此機動地巡迴於合作夥伴學校或文教機構，進行虛實策展，訴說在地故事，達到寓教於樂的目的。



圖 6 《淡水維基館》行動故事車

「《淡水維基館》行動故事車」至今已洽妥之合作單位有百齡高中、滬尾偕醫館、淡江大學 USR 辦公室、臺北基督學院、淡江大學外語學院 …… 等，並於 2019 年至 2020 年之間，陸續舉辦了 6 場 VR 巡迴策展活動：

- (一) 2019/12/26：與臺北市立百齡高中合作，結合高一（108 上）多元選修課程，於該校圖書館舉行。
- (二) 2020/07/02：與臺北市立百齡高中合作，結合高一（108 下）多元選修課程，於該校圖書館舉行。
- (三) 2020/10/18：與淡水教會、淡江歷史系、淡江資傳系合作，結合賴清德副總統參訪淡水教會行程和淡江大學 USR 計畫成果展，於滬尾偕醫館舉行。
- (四) 2020/11/03：與淡江大學 USR 辦公室合作，結合淡江大學 70 週年校慶 USR 成果展，於守謙國際會議中心大廳舉行。
- (五) 2020/12/01：與臺北基督學院圖書館合作，正式將兩檔展覽於臺北基督學院校內同時展出。
- (六) 2020/12/17：與淡江大學外語學院合作，結合 USR-HUB 計畫在外語學院大廳展出 VR 體驗虛實策展。

此 6 次巡迴虛實策展展覽之主題展覽各有差異，以下簡單列表說明：

展覽時間	展覽地點	展覽主題
2019/12/26	臺北市立百齡高中	淡金海岸石滬群 VR 特展
2020/07/02	臺北市立百齡高中	淡金海岸石滬群 VR 特展
2020/10/18	淡水教會滬尾偕醫館	河海山城老淡水：蔡坤煌攝影 VR 展
2020/11/03	淡江大學 守謙國際會議中心	河海山城老淡水：蔡坤煌攝影 VR 展
2020/12/01-02	臺北基督學院	淡金海岸石滬群 VR 特展 河海山城老淡水：蔡坤煌攝影 VR 展
2020/12/17	淡江大學外語學院	河海山城老淡水：蔡坤煌攝影 VR 展

展覽內容以實體展架說明配合虛擬 VR 園區，展區及展示過程如下圖：



圖 7 實體海報展示



圖 8 觀展者了解展覽主題



圖 9 VR 體驗

肆、VR 觀展者問卷調查與分析

《淡水維基館》VR 園區作為一個致力於人文科學普及化的策展內容，特意強調實體展場與虛擬展場整合之效用，本文認為有其必要性。一者是數位策展所使用之數位技術有時對於觀展民眾而言是陌生的，當科技工具無法有效輔助民眾了解展覽內容，展覽將有其侷限，而使得數位策展之內容無法有效的轉換典藏物件與民眾。二者是 VR 展覽目前大多以個人體驗為展覽主流觀看模式，若只有數位展示似乎無法讓觀展群眾有共感的觀展體驗，策展人亦

無法有效掌握觀展者的體驗與策展目標是否一致。有鑑於此，數位策展時為了要能達到傳遞知識，普及人文教育的目的，本文特意延伸出「虛實策展」之模式。

本研究欲分析「《淡水維基館》VR 園區」之資料皮用與虛實策展策略，是否達到策展之初「協助地方學和臺灣史之學術研究與人文普及教育，達到文化資產數位保存、行銷推廣與加值利用」的目標，因此透過「《淡水維基館》VR 行動故事館巡迴策展」的方式輔以問卷調查法以了解其成效。

問卷調查法是一種常見的調查法，藉由提問受試者數個相關的量化選擇問題，以此了解受試者對研究主題的看法。而本研究施以問卷調查，主要目的即是探討 VR 觀展的感受問題，因此在巡迴展演前設計了若干問題（詳見附錄），問卷發放的場次以淡江大學守謙國際會議中心、臺北基督學院與淡江大學外語學院 3 場巡迴展為主。發放問卷前觀眾會先有觀展體驗，其中包含導覽人員解說展覽內容，以及觀賞 VR 展覽的操作說明，並且從旁協助觀眾全程之 VR 展覽體驗。

本次問卷共回收 102 份。首先，受試者基本資料方面，年齡分布以 20-29 歲之比例最高，共 50 份占 49%，次高為 30-39 歲有 20 份占 19.6%。在教育程度上 60 份 58.8% 為大學專科學歷為最高比例，次高為碩士有 24 份 23.5%，就讀科系或服務學院中有 76 份 74.5% 為人文科學（包含：藝術、文學、哲學、歷史、語言），社會學科（包含：人類學）有 15 份占 14.7%。

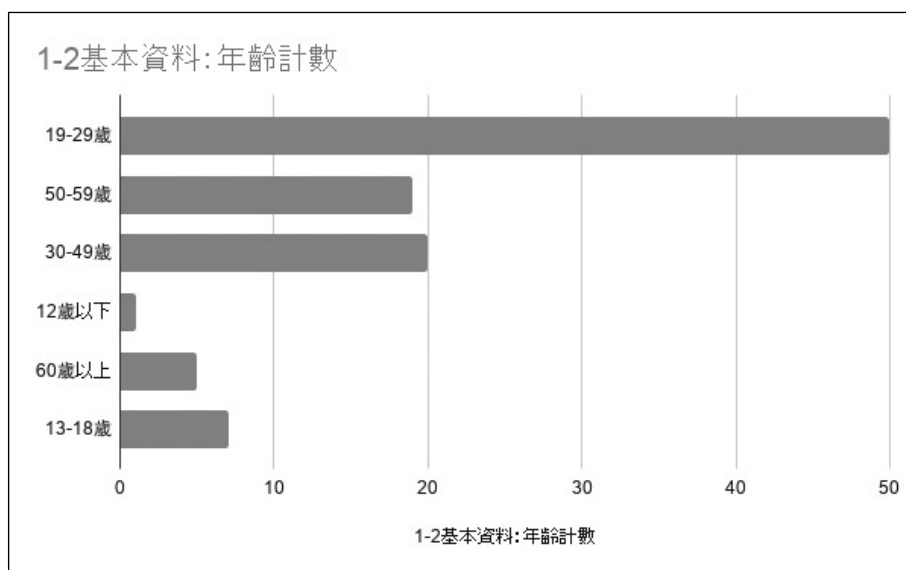


圖 10 觀眾之背景資料：年齡

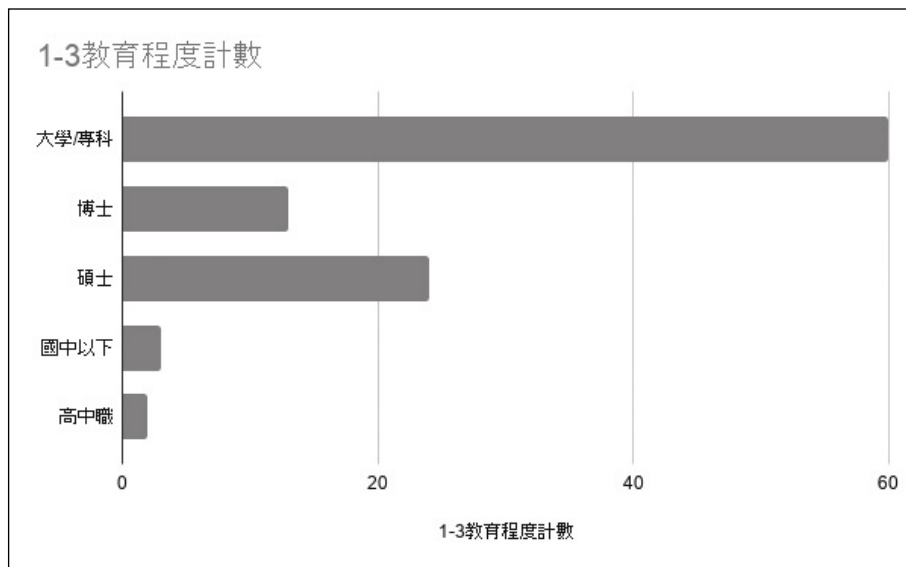


圖 11 觀眾之背景資料：教育程度

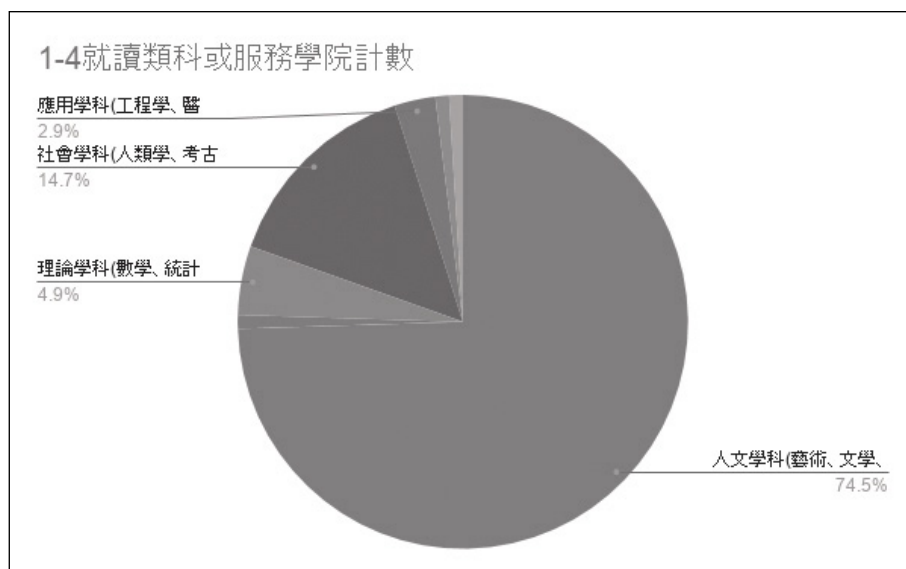


圖 12 觀眾之背景資料：就讀科系及服務學院

由基本資料可推測由於此次策展場地以大專院校為主，因此年齡以及教育程度人數比例大致吻合觀展人的基本背景。而在服務或就讀學院上亦是如此，不論是外語學院或是基督學院亦大致與問卷數據相似。

在 VR 設備的經驗上，有 97 人 95.1% 的觀眾都曾經聽過 VR 虛擬實境技術，僅有 5 人 4.9% 從未聽說。但只有 44 人 43.1% 的觀眾曾使用過 1-3 次的 VR 裝置，有 40 人 39.2% 是從未使用過的，僅有 5 人 4.9% 的觀眾有使用 10 次以上的經驗。由此可推測 VR 虛擬實境的裝置普及率不高，仍是一種新興型態的數位策展裝置。

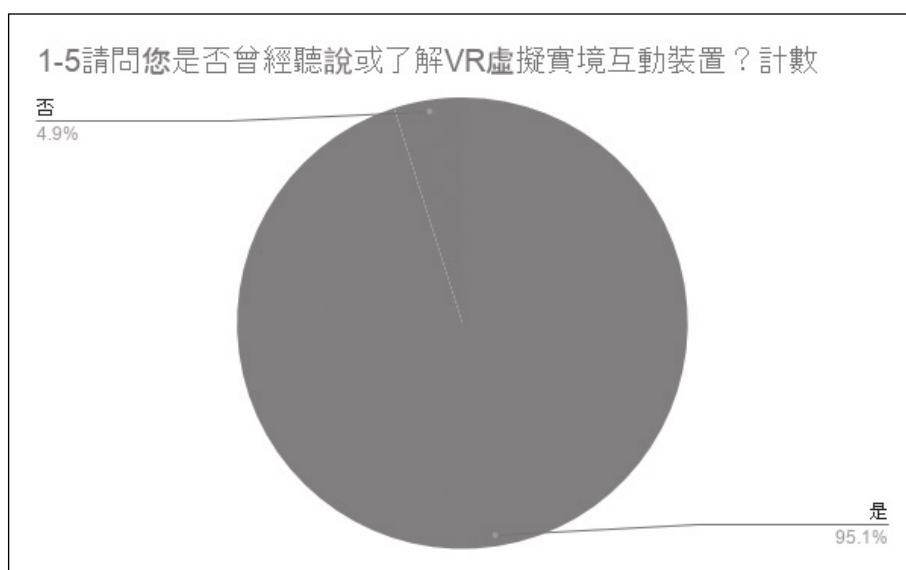


圖 13 觀眾對於 VR 的基本認知與使用頻率

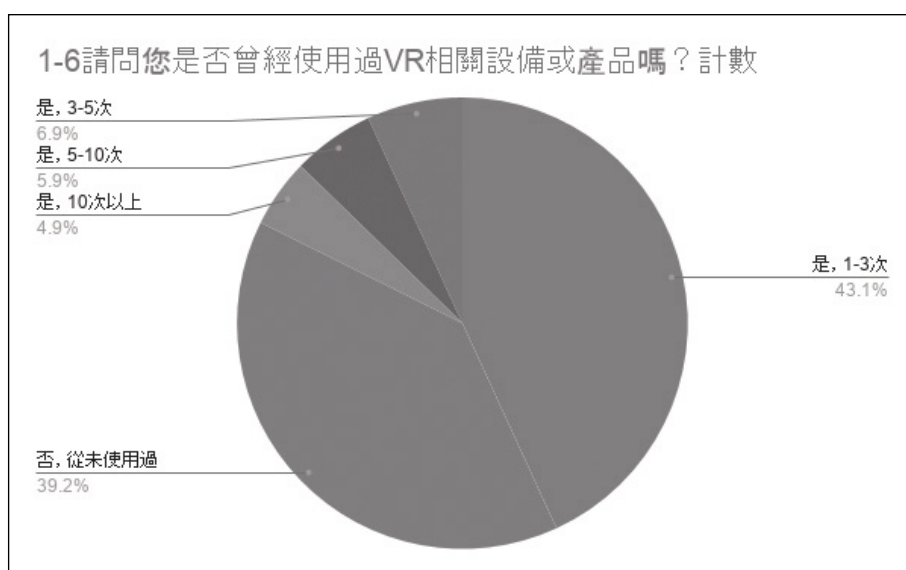
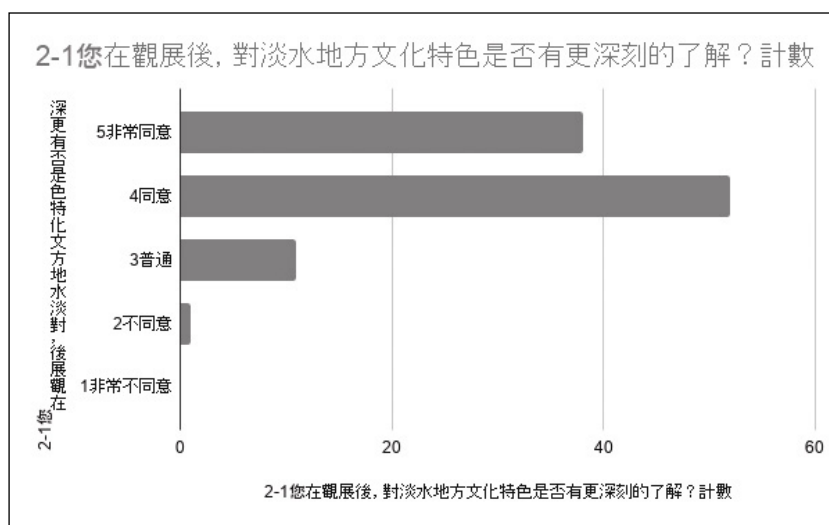
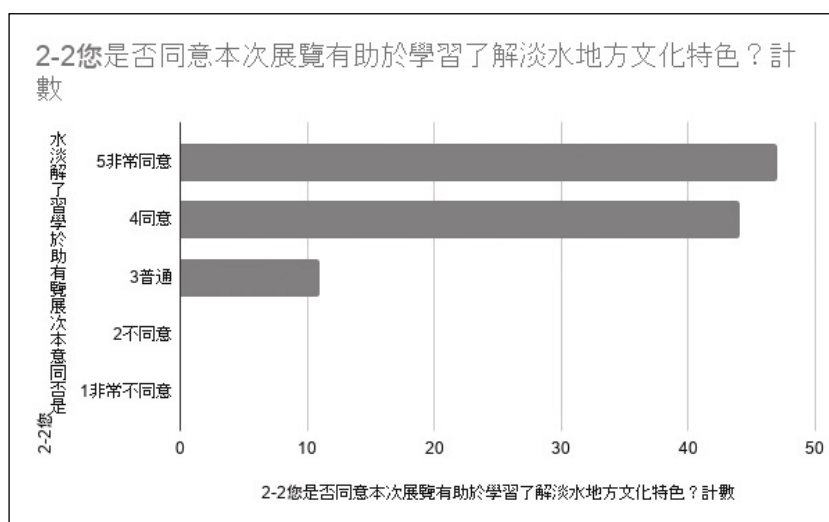


圖 14 觀眾對於 VR 的基本認知與使用頻率

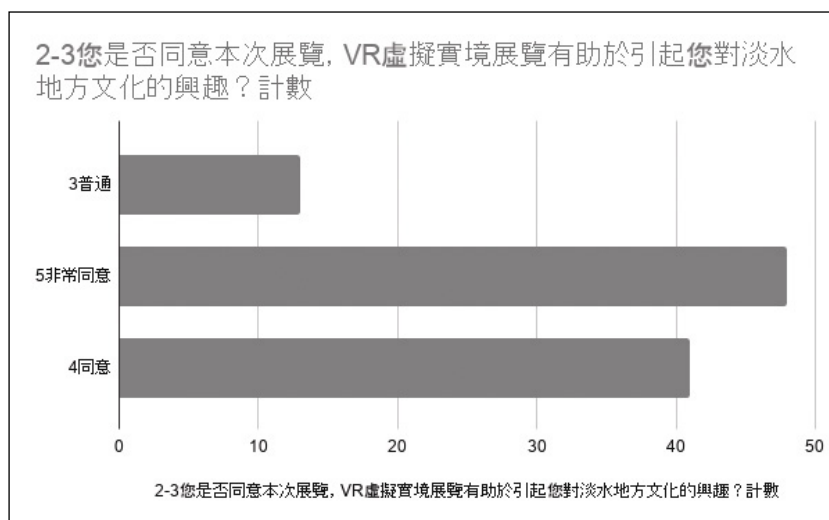
在展覽主題上，近九成的觀眾認為展覽內容有助於他們了解淡水的地方文化與特色，並認同《淡水維基館》的展覽是有助於學習在地知識。同時，亦有九成以上的觀眾認同 VR 裝置以及其中影音互動的過程能讓觀眾對主題更加了解，亦符合觀展前對展覽的期待。



(a)



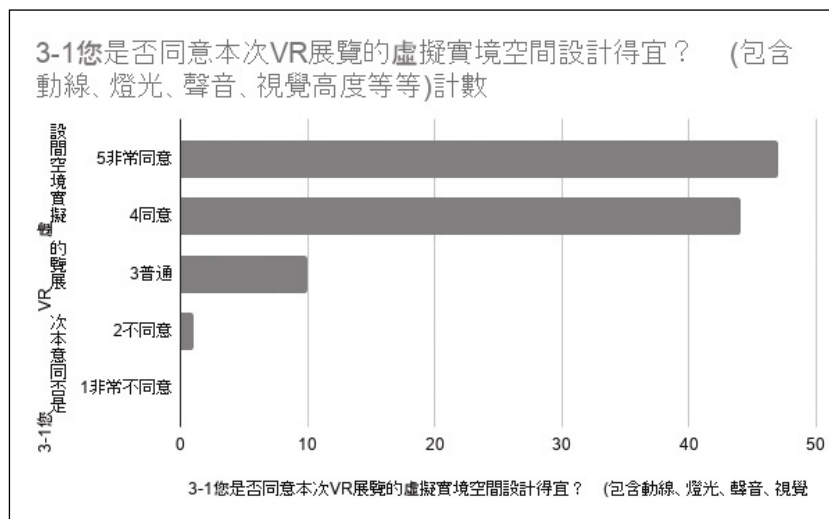
(b)



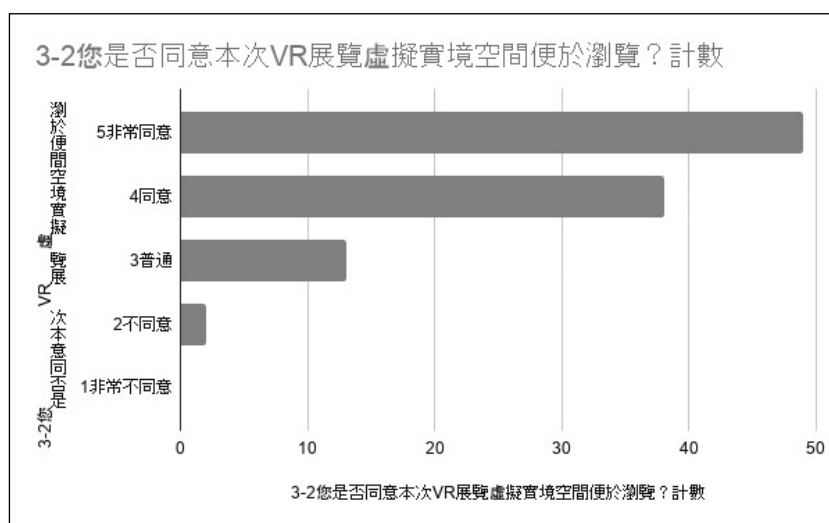
(c)

圖 15 觀眾對展覽主題與展覽方式的滿意程度

另外，在虛擬展場的觀展感受中，觀眾對於 VR 展場的空間、光線、視覺高度等觀展體驗上，感受普通與同意（最高分為非常同意）的觀眾有 54 人 53.1%；虛擬展場的動線規劃亦有 51 人 52.0% 的觀眾認為有改善空間。這也說明了虛擬展場目前的困境，若是民眾不熟悉虛擬互動裝置的情況下，觀展感受會受到一定程度的影響。



(a)



(b)

圖 16 VR 觀眾對於裝置的觀展感受

最後，在開放問卷的回饋上，部分觀眾對於展覽中印象深刻的部分是虛擬實境的體驗，因為 VR 裝置對大部分的觀眾而言都是新嘗試，但也因為對裝置的不熟悉，在虛擬展場的觀賞過程中操作不順利、移動方向受限、裝置的穿戴重量都影響著觀眾的觀展感受，更多觀眾在觀展後表示 VR 目前的移動方式無法讓自己手眼協調操作，造成觀展障礙。這也是 VR 裝

置目前較難以突破的限制，若是沒有實際導覽人員在旁協助，對 VR 裝置不熟悉甚至抗拒的觀眾可能無法完整了解展覽的內容。另外，在 VR 裝置的虛實展場的幫助上，部分觀眾提及數位化的展覽內容可以補足展牆海報上無法看清楚的画面，加上虛擬的情境營造，彷彿走向另一個世界，若有機會願意將展覽介紹給更多親友來參觀。

伍、結語

傳統地方志與近現代地方學的研究是一門與地方文化、土地、自然……等面向息息相關的學問，研究課題及對象時常有賴研究者走入地方，進行田野調查與蒐集資料，然而，這些過程在過去非數位化的時代，收集、整合、分析、統整都不是一件容易的事。當地方學走向數位化的過程，除了將研究資料及成果以數位典藏的方式保存外，典藏之後如何再次利用成為值得思考的問題。本文以地方學數位共筆平台《淡水維基館》為實驗場域，將過去地方學數位典藏於此的資料加以整理再次利用，並嘗試尋找出一套可供參考的策展模式，以達到典藏資料數位皮用的目標。而在本文的實驗中提出「虛實策展」的概念，進階於數位策展中以數位工具為主的展覽方式，提出實體展覽與數位展覽同時並呈的概念，以行動故事車的模式進行，將展覽主動接近社會大眾，以達更高效率的知識普及與傳播。

致謝

本研究感謝科技部編號 MOST 108-2410-H-032-048-MY2 研究計畫、淡江大學 108-109 重點研究計畫經費補助；「文五合 e」研究團隊成員賴惠如副教授、李其霖副教授，過去幾年在淡水石滬研究上的合作；前新北市議員蔡葉偉先生提供蔡坤煌醫師攝影老照片；淡江大學海事博物館、臺北市立百齡高中、淡水教會滬尾偕醫館、淡江大學 USR 辦公室、臺北基督學院、淡江大學外語學院等，提供展覽場地協助展出，使得本研究得以順利進行，在此一併致謝。

附錄

淡水維基館 VR 園區展覽研究問卷

您好，本問卷為學術研究問卷，目的以了解您在觀賞本次虛擬實境互動裝置展出後之使用評估與感受，感謝您耐心地回答與協助！本問卷採匿名形式填答，所有資料僅供學術研究用途，絕不對外公開，請您安心填答，再次感謝您的配合！

淡江大學資圖系
數位典藏與數位人文研究室
指導老師：林信成 教授
研究員：范凱婷

1-1 基本資料：性別 *

☐女 ☐男

1-2 基本資料：年齡 *

☐ 12 歲以下 ☐ 13-18 歲 ☐ 19-29 歲 ☐ 30-49 歲 ☐ 50-59 歲 ☐ 60 歲以上

1-3 教育程度 *

☐國中以下 ☐高中職 ☐大學 / 專科 ☐碩士 ☐博士

1-4 就讀類科或服務學院 *

- ☐人文學科 (藝術、文學、哲學、歷史、語言)
- ☐社會學科 (人類學、考古學、人文地理學、經濟商學、政治學、心理學、社會學、教育學)
- ☐自然學科 (物理學、化學、生物學、地球科學、太空科學)
- ☐應用學科 (工程學、醫學、健康科學、建築學)
- ☐理論學科 (數學、統計學、計算機科學、系統科學)
- ☐其他

1-5 請問您是否曾經聽說或了解 VR 虛擬實境互動裝置？ *

☐是 ☐否

1-6 請問您是否曾經使用過 VR 相關設備或產品嗎？ *

- ☐是，10 次以上 ☐是，5-10 次 ☐是，3-5 次 ☐是，1-3 次
☐否，從未使用過

2-1 您在觀展後，對淡水地方文化特色是否有更深刻的了解？ *

- 非常不同意 非常同意
 1 2 3 4 5

2-2 您是否同意本次展覽有助於學習了解淡水地方文化特色？ *

- 非常不同意 非常同意
 1 2 3 4 5

2-3 您是否同意本次展覽，VR 虛擬實境展覽有助於引起您對淡水地方文化的興趣？ *

- 非常不同意 非常同意
 1 2 3 4 5

2-4 您是否同意展覽中，VR 虛擬實境展覽的「影音互動裝置」有助於進一步了解淡水地方文化特色？

- 非常不同意 非常同意
 1 2 3 4 5

2-5 您是否認同這次的 VR 展覽有符合您原先的期待？ *

- 非常不同意 非常同意
 1 2 3 4 5

3-1 您是否同意本次 VR 展覽的虛擬實境空間設計得宜？（包含動線、燈光、聲音、視覺高度等等）*

- 非常不同意 非常同意
 1 2 3 4 5

3-2 您是否同意本次 VR 展覽虛擬實境空間便於瀏覽？ *

- 非常不同意 非常同意
 1 2 3 4 5

3-3 我對本次的 VR 展覽整體規劃感到滿意 *

非常不同意

非常同意

1 2 3 4 5

3-4 我對這次的 VR 展覽參觀經驗是愉快的 *

非常不同意

非常同意

1 2 3 4 5

4-1 本次展覽中，您印象最深刻的是什麼部分？ *

4-2 您在 VR 裝置操作上，最困擾的是什麼部分？ *

4-3 您會願意推薦本次 VR 展覽給其他人嗎？請簡述原因。

4-4 若淡水維基館 VR 園區規劃展覽內容，您會希望看到什麼類型的淡水故事？ (例如：歷史發展、自然生態、人物志、淡水大事件)*



洪振洲 Jen-Jou Hung

洪振洲，2006 年於臺灣科技大學資訊管理系取得博士學位，目前任教於法鼓文理學院佛教學系，為專任副教授，並兼任圖書資訊館館長一職。他目前參與許多由法鼓法鼓文理學院執行的數位典藏計畫。他的研究興趣包含漢譯佛典作譯者分析、數位典藏專案建構、數位人文研究資源開發與數位文字處理議題。

Jen-Jou (aka “Joey”) Hung was born in Taiwan in 1976. He received his Ph.D. in 2006 from the National Taiwan University of Science and Technology, Taipei. Now, he is Associate Professor in the Department of Buddhist Studies and Director of the Library and Information Center at Dharma Drum Institute of Liberal Arts (DILA), Taiwan. He is currently engaged in a variety of digital archives projects at DILA. His research interests include authorship attribution for ancient Buddhist literature, the construction of digital archives, and digital text processing.

開放資料服務於漢籍數位人文研究資源之發展與應用

The Development and Application of Open Data Services in Classical Chinese Digital Humanities Resources

洪振洲

Jen-Jou Hung

法鼓文理學院佛教學系副教授兼圖書資訊館館長

Associate Professor, Department of Buddhist Studies, Dharma Institute of Liberal Arts

摘要

開放資料 (Open Data) 一詞，是近年來於科學界與政府機關的發展討論中，經常被提到的重要議題。其主要目的是希望促進各界將具有應用與研究價值的原始資料，開放給社會公眾取用，而不受其使用目的的限制。對於學術研究應用而言，開放資料將帶來多項優點，包括：增進研究品質與效率、快速的產生研究新想法、促進新創應用等優點。而對於十分依賴研究資料的數位人文研究領域來說，資料的取得與累積是本就是研究的關鍵之一。因此開放資料服務的普及與否，將可能對於數位人文的研究進展造成重大的影響。實際上，於我們熟悉的漢籍資料研究環境中，已經陸續有許多的數位研究資源工具，注意到開放資料的重要性，進而願意提供完整的原始資料下載，或甚至建置方便其他系統進行資料交換的開放性應用程式界面 (Open API)。此外，我們也注意到，已經有數個數位人文工具，進一步利用各自所設置的開放資料服務，建立跨系統連結，儼然形成一個小的數位人文研究網絡。這樣的發展，可以說是形成了大型整合數位人文環境的一種雛型樣貌。但細觀這些發展，將會發現各系統間的資料開放與整合連接方式並尚未有統一的標準規範，也就是說，各數位資源皆採用了不同的方式建置開放資料服務。因此，本論文的主要目的，便是針對幾個知名的數位人文系統所建置的開放資料服務，進行整理說明與分析比較，一方面讓讀者以系統化的方式了解到現行漢籍數位人文系統開放資料服務的發展現況，同時也可以作為新數位資源於建置開放資料服務時之參考。

關鍵字：開放資料、漢籍數位資源、開放性應用程式界面、數位人文

Abstract

The term Open Data is an important topic frequently mentioned in the discussions on development between the scientific community and government agencies in recent years. Its main purpose is to promote the application and research value of the original materials from all walks of life, and make them available to the public, without restrictions placed on their use. For academic research applications, open data will bring many advantages, including: improving the quality and efficiency of research, generating research ideas more quickly, and promoting the creation of new forms of innovative applications. For the research field of the digital humanities that relies heavily on digital data, the acquisition and accumulation of data is one of the keys to this research. Therefore, the popularity of open data services may have a major impact on the progress of digital humanities research. In fact, in the Chinese data research environment with which we are familiar, there is an increasing number of digital research resource tools, indicating the importance of open data, and a willingness to provide complete original data download, or provide an open application programming interface (Open API) that facilitates data exchange with other systems. There are even several digital humanities tools that have further used their respective open data services to establish cross-system links, forming a small digital humanities research network. This type of development can be said to be a prototype of the development of an integrated digital humanities environment. However, because there is no unified standard specification for the open data and integration connection methods between various systems, all digital resources currently adopt different methods to build open data services. Therefore, the main purpose of this paper is to sort out, explain, analyze and compare the open data services built by several well-known digital humanities systems; on the one hand, it provides an overview and allows readers to understand in a systematic way the development status of the current Chinese digital humanities' open data services, and at the same time, it can also be used as a guide for the establishment of open data services or the development of a common open data service implementation strategy.

Keywords: Open Data, Chinese Digital Research Environment, Open API, Digital Humanities

壹、簡介

將資訊科技應用於整理人文研究資料、探索人文知識與解決人文的研究問題，是數位人文的主要核心概念。因此，數位人文研究領域的進步，一直以來便與數位內容的產出及分析工具的發展有著密不可分的關係。近年來隨著資訊科技的快速進步，與數位人文研究概念的蓬勃發展，因此有為數不少的組織團體積極投入數位人文工作項目，也連帶許多以漢籍文獻資料為主的傑出數位研究資源，諸如：中國歷代人物資料庫 (China Biographical Database Project, 簡稱 CBDB)¹、CBETA 數位人文平台 (CBETA Research Platform)²、The SAT Daizōkyō Text Database³、中國哲學書電子化計劃 (Chinese Text Project, 簡稱 CText)⁴、漢リポ (Kanseki Repository, 簡稱 kanripo)⁵、MARKUS 半自動標記平台⁶、Docusky 數位人文研究平台⁷、中央研究院數位人文研究平台⁸等，都可以說是其中的佼佼者。

觀察目前的較為成功的數位人文成果，我們也不難發現到，大部分成功的數位範例，其內容多半是以單一的核心功能為主⁹，少有能概括完整人文研究需求的工具。這是因為在一個完整的人文研究的過程當中，研究者必須經歷資料蒐集、資料整理、內容擷取、資訊分析等多個步驟。而上述的每一個步驟的目的都不同，執行上的需求也不同，再加上資料本身的異質性以及後續分析方法的多樣性，導致要想製作單一數位人文工具來滿足研究者所有需求，實有其難以突破的困難之處。於此情況下，應運而生的是各種為了完成特殊工作項目的工具與資源。但是大量出現的工具資源，不僅可能造成使用者的選擇困難，也可能會讓開發者花費過多心力於提供重複的基礎功能。因此近年來，在許多相關會議與討論場合中，當提及數位工具與資源的發展時，都會出現希望見到能整合現有的工具與資源，以創造一個更好的數位人文研究環境為主題的相關討論。以目前現有的成功整合範例來看，由歐洲研究基礎建設策略論壇 (the European Strategy Forum on Research Infrastructures, 簡稱

1. 有關 CBDB 的細節，詳情請參閱：Harvard University, Academia Sinica, and Peking University, China Biographical Database (CBDB) (2018), <https://projects.iq.harvard.edu/cbdb>, (2020/10/25 上網)。

2. 「CBETA 數位研究平台」，網址為：<http://cbeta-rp.dila.edu.tw>, (2020/10/25 上網)。

3. 「SAT 大正新脩大藏經テキストデータベース」，網址為：<http://21dzk.l.u-tokyo.ac.jp/SAT/> (2020/10/25 上網)。

4. 「中國哲學書電子化計劃 (英文：Chinese Text Project)」，網址為：<https://ctext.org/zh>, (2020/10/25 上網)。

5. 「Kanripo 漢籍リポジトリ」，網址為：<https://www.kanripo.org/>, (2020/10/25 上網)。

6. 「MARKUS」，網址為：<https://dh.chinese-empires.eu/markus/beta/>, (2020/10/25 上網)。

7. 「Docusky 數位人文學術研究平台」網址為：<http://docusky.digital.ntu.edu.tw/>, (2020/10/25 上網)。

8. 「中央研究院數位人文研究平台」，網址為：<https://dh.ascdc.sinica.edu.tw/>, (2020/10/25 上網)。

9. 例如以提供大量高品質的數位資料給研究者的「資料提供者」，多半僅提供少數開發者預想之資料整理與分析功能。而另一方面，也有許多「數位工具平台」系統是以提供進行資料處理與分析為主的。當然這些平台工具通常本身並沒有先預先載入資料，而希望讓使用者能以自己的資料，搭配平台的功能來運行。

ESFRI)，發起的 DARIAH 計畫，可說一個相當成功的數位人文資源整合案例，該計畫中由考慮資料的徵集、數位化、分析、出版等完整的程序，並提供整合良好的數位資源與工具供使用者使用。但該計畫規模龐大，也需要大量人力與經費資源的挹注才得以實現¹⁰。

在現實情況下，要進行大型整合計畫並非易事，但於我們熟悉的漢籍資料研究環境中，我們觀察到另一種由下而上的發展樣貌。更仔細的說，我們發現到已經有多個數位研究資源，建置了相當完整的開放資料服務，包括：提供完整原始資料的完整下載，或進一步提供具有存取原始資料的開放性應用程式界面，公其他系統取用全部或部份的原始資料。不僅如此，我們也觀察到幾個數位人文工具已經透過這些資料開放服務，進行實質上的系統間資料交換與服務整合，建立了跨系統連結。也就是說，在這些已經相互連結的數位資源中，使用者可以將所需研究的數位資料由一個資源傳送至另一個工具平台中，以進行更進一步的處理。這樣的發展，可以說是發展整合型數位人文環境的一種雛型樣貌，其未來發展相當令人期待。不過，但這樣的分散式發展，也連產生了規範不一致的問題。實際上，目前這些提供開放資料服務的漢籍數位資源，皆採用了不同的方式來建置其開放資料服務。這對於欲使用這些開放資源的使用者而言，可能有較難掌握全貌之感，另外對於要新設開放資料服務的數位資源來說，也可能不易找到適用的發展樣板。因此，本論文的主要目的，便是將幾個知名的數位人文系統所建置的開放資料服務進行整理說明與進行分析比較，一方面讓讀者可以有系統的了解到現行漢籍數位人文系統開放資料的發展策略概觀，也同時可以進行各家策略的優劣比較。此外，本論文之整理結果，也可以作為新系統發展資料開放策略之參考，或用於發展共通的資料分享策略時之指引。

貳、開放資料服務

一、開放資料 (Open Data)

開放資料一詞，是近年來經常於科學界與政府機關的發展討論中，經常被提到的重要議題。其主要目的是希望促進相關各界將具有應用與研究價值的原始資料，開放給社會公眾取用，而不需受其使用目的的限制。開放資料的背後的理念，可以追溯到 Merton(1973) 所倡導之「所有研究者應該將其研究成果全面性的免費公開，並放棄智慧財產權，以促進知識發

10. 目前 DARIAH 計畫，共有 19 個歐盟會員國參與，並且該計畫以建設工具、資料、知識的交流平台、教育與訓練平台、以及其他推進數位人文研究的活動為主軸。詳請參閱 DARIAH (2020)

展的概念」¹¹。由此概念開始，漸漸擴展為應該開放研究來源資料、學術研究成果、學術論文、程式原始碼及政府公開觀測資料等等的全面開放科學 (Open Science) 運動。

而對於科學研究應用而言，開放資料的策略將帶來多項優點，包括：增進研究效率、增加對研究方法的意識、增進研究品質、更快速的產生研究想法、促進產生新型態的新創應用等優點 (Open Science and Research Initiative, 2014, p.3)。在數位人文研究的應用中，資料的取得與累積是非常重要的環節，透過快速取得與大量資料的累積，數位研究平臺所提供的知識，將可能讓研究者達到博學的程度 (劉昭麟，2018，頁 14)。因此資源間開放資料服務的普及與否，對於數位人文的研究進展將會有重大的影響。至於要提供怎樣的資料，才符合何謂開放資料的定義，在 Open Definition 網站¹²中，給予開放資料一個較為明確的定義：

- (一) 開放資料必須以開放式授權的方式公開發行。
- (二) 開放資料必須可由網路下載。
- (三) 開放資料必須為可機讀格式。
- (四) 開放資料必須為開放格式。

由以上的定義，我們不難看出，上述對於開放資料的要求，其實著重於讓資料可以被其他系統再次運用的可行性。也就是希望藉由資料開放的方式，能讓資料在不同系統間流動，進而產生各種不同的應用。以此目的來看，單單僅把原始資料打包後公開於網路上提供下載，並不一定能構成開放資料的最好策略，因此開放性應用程式界面 (Open API) 便應運而生，成為提供開放資料的另一個良好管道。

二、鏈結開放資料 (Linked Open Data)

在提到開放資料時，鏈結開放資料是一個我們無法忽略的重要發展分支。該技術是由全球資訊網 (World Wide Web) 發明者提姆柏納 - 李 (Tim Berners-Lee) 於 2001 年在《科學人》雜誌專文提出語意網 (Semantic Web) 的概念 (Berners-Lee, Tim, et al. 2001)。他也在 Berners-Lee(2006) 中提到鏈結資料 (Linked Data) 技術的概略樣貌。該技術希望藉由一個簡單的設計，建立開放的、外顯的、容易理解的方式來描述資料間的關係，其最終目的是要將現有的數位資料串連，以產生一個讓電腦能夠直接或間接處理的資料網絡 (Web of

11. 原文為：each researcher must contribute to the “common pot” and give up intellectual property rights to allow knowledge to move forward.

12. Open Definition 為開放知識基金會 (Open Knowledge Foundation) 所設立之網站，網址為：<https://opendefinition.org/>，(2020/10/25 上網)。

Data)，以幫助電腦正確解讀人類的知識內容。並且他在 2009 年 TED talk 以「下一代的網路」為題¹³，呼籲資訊界重視「語意網」與鏈結資料的發展，以推進網絡的發展方向。同時，他也進一步將鏈結資料的概念定義為符合以下要求的資料內容：

- (一) 每一個表述的資源必須具有唯一 URI (Uniform Resource Identifier，統一資源識別碼)，用以表達與存取資料。
- (二) 資源的存取方式是透過 HTTP 協定 (Hyper Text Transfer Protocol，超連結傳輸協定) 與該資源的 URI 而達成。
- (三) 資源必須藉由開放標準來提供存取方式，例如 RDF 表示方式、SPARQL 查詢語言。
- (四) 當資料發佈於網路上時，必須提供與網路上其他資料集合的連結。

後續鏈結資料的概念，已經被 W3C 正式收錄為網路資料建置的標準項目之一，近年來也陸續訂定了許多與鏈結資料相關的網路標準。由於鏈結資料概念具有相當的突破性與重要性，因此近年來鏈結資料的概念得到許多大機構的支援，紛紛推出具有鏈結資料的概念之資料集合與相關服務，諸如 Depdia¹⁴、Microsoft Concept Graph¹⁵、Google Knowledge Graph¹⁶ 都是相當知名且成功的範例。

三、Open API

所謂的 API (Application Programming Interface) 意指一個應用程式所保留給其他應用程式進行互動的界面，這個界面通常規定了函式間如何呼叫、如何傳送需求參數、如何解讀回應資料與資料的詳細格式等等。簡而言之，API 就是程式之間的互動方式，可以讓各自獨立的程式於運行的過程中達到相互溝通、協同作業的管道。一般來說，API 多數存在於程式開發過程中使用到的軟體開發套件 (Software Develop Kit, 簡稱 SDK) 中所提供的底層函式庫中。這些 API 設計的目的，也多半用於簡化複雜但經常會利用到的功能，例如溝通機器

13. Tim Berners-Lee 於 2009 年在 TED 網站上所發表之演講，原文講題為：The next web. 網址為：https://www.ted.com/talks/tim_berniers_lee_the_next_web，(2020/10/25 上網)。

14. 由位於德國萊比錫的自由大學的團隊於 2007 年建立，該網站上所有資料都利用開放式鏈結資料形式來提供。其內容主要是將維基百科 (Wikipedia) 的內容經過電腦系統進行自動關聯匹配與連結而產生。網站除提供線上服務外，也提供資料下載。在目前最新資料版本 2016-04 內，共有約 4,233,000 個概念在此資料庫，且彼此間有相當綿密的聯結。(Auer et. al, 2007)

15. 微軟 (Microsoft) 公司也發表了「Microsoft Concept Graph」，其內容紀錄超過 540 萬條核心概念，與 8400 萬筆的概念關係資訊。「Microsoft Concept Graph」，其網址為：<http://concept.research.microsoft.com>。(2017/10/24 上網)

16. 在 Google 進行搜尋時，若所搜尋的目標為一個名詞，此時右方會出現的一個「名詞說明欄」(Knowledge Graph information)，其背後的技術，被 Google 稱為，Google Knowledge Graph，其後端原理即是透過處理鏈結資料而達成。Google Knowledge Graph，其網址為：<https://googleblog.blogspot.tw/2012/05/introducing-knowledge-graph-things-not.html>。(2017/10/24 上網)

硬體的元件（如印表機列印函式），或複雜計算函式（如統計函式、矩陣運算函式）等等。

而近年來隨著網頁應用程式與雲端大數據等應用的興起，API 的概念開始延升應用到了網頁環境中，而衍生出了新的應用模型，這個模型被稱為 WEB API。顧名思義，所謂的 WEB API 指的是網頁應用環境中的程式溝通界面，此一概念大致上出現於西元 2000 年左右，由 ebay, amazon, salesforce 等商業軟體廠商，不約而同同時推出利用網頁環境進行系統交換的 WEB API 服務¹⁷。而此後 WEB API 的概念持續熱門，Google 也於 2006 年提供公開的 API 以存取 Google MAP 的資料¹⁸。根據 Programmable Web 網站的統計¹⁹，網路上所提供 WEB API 服務的數量由 2005 年的 105 種，提昇至 2013 年的 9000 種，並於今日已達超過 20000 種的規模。

四、漢籍數位人文資源開放資料服務

由下一章節開始，本文將針對幾個漢籍數位人文研究工具資源所提供的開放資料服務，進行整理說明與差異比較。由於各數位資源與工具的屬性具有相當的差異性，而這些性質上的差異，也影響到它們對於資料開放方式的選擇。因此我們將欲討論的漢籍數位人文研究工具資源項目，簡單分類為：「名錄型數位資源」、「全文型數位資源」與「純數位研究工具」，等三個類型，再分別針對各資源所提供的資料開放服務進行整理比對。而比較的重點，則集中在所各資源提供的開放資料範圍、檔案格式、授權方式、更新頻率，以及是否提供 Open API、與該 API 的涵蓋範圍、函式數量等等項目。

參、名錄型數位資源的開放資料服務

所謂的「名錄型數位資源」泛指以提供人物、地點、組織、家族或其他重要實體的相關資料之數位資源。此類資源所收錄的物件資料，通常可以看作為一筆具有多個屬性值的資料序列。而資源內容除了紀錄各物件的基本屬性資料之外，有時更重視紀錄物件之間的關係。而相較於第肆章將介紹的「全文型數位資源」，名錄型資源內所紀錄的資料，通常是較為精練的內容，是已經過人類閱讀後所產生的結構化資料的紀錄。

17. 參考自 History of APIs, <https://apievangelist.com/2012/12/20/history-of-apis/>, (2020/11/03 上網)

18. 實際上 Google 於 2006 年公開的 MAP API 版本為 v2，而 Google maps 的正式服務上線時間為 2005 年，因此推測其 v1 版本的 API，應是在 2005-2006 間推出。(Google Maps API Blogs, 2006).

19. 詳情參閱：Programmableweb (2013)

一、中國歷代人物傳記資料庫 (CBDB)

本資料庫為郝若貝教授 (Robert M. Hartwell) 開始發起的資料庫建置計畫。目前由哈佛大學費正清中國研究中心、中央研究院歷史語言研究所及北京大學中國古代史研究中心等三個機構合作進行維護工作。該資料庫廣泛的收集中國歷史上重要人物之傳記資料，並將其結構化，提供使用者進行查詢。資料庫內所收集的人物主要出自七世紀至十九世紀，除人物基本背景資料外，內容也詳加紀錄其社會關係與親屬關係。根據其網站資訊，截至 2020 年 5 月為止，該資料庫共收錄約 470,000 人的傳記資料與期間的關係訊息，形成一個巨大的中國歷代人物關係網絡，也成為近代學者研究中國歷史人物關係，不可或缺的重要參考資料。而 CBDB 相當重視提供開放資料，提供了許多不同的資料存取方式。

(一) 原始資料下載

CBDB 主要是以傳統關聯式資料庫的方式來進行資料的整理與儲存，並主要提供單機離線版資料庫的給研究者進行下載並使用²⁰。該資料庫使用 Microsoft Access 軟體製成，內容並無設定讀取權限限制或也無經過加密，也因此讀者可以輕易匯出所有資料內容，也可說是一種完整原始資料的下載的形式。不過經由 Microsoft Access 軟體製成的 mdb 資料庫格式實際上為非開放格式，而使用者也必須有擁有對應的軟體才能開啟運行。或許是為了彌補此缺憾，CBDB 另外提供了開放式資料庫 SQLite 格式的原始資料檔案下載，以便讓使用者也可以透過開放授權軟體取得資料庫的完整內容。而在資料更新的頻率方面，在本文撰寫時，網站上所陳述的最新一次更新為 2020 年 5 月，距離前一次更新約為 1 年的間隔。而在資料授權方面，根據網站資訊，CBDB 的內容以「創用授權 非商業性使用 - 相同方式共享 4.0 國際許可協議 (CC BY-NC-SA 4.0)」。因此對於非商業性的應用系統來說，可以直接方便的取用 CBDB 的資料內容。

(二) 鏈結數據服務

除了關聯式資料庫的服務之外，CBDB 製作小組也於 2018 年時，公開與上海圖書館合作建置之鏈結數據服務網站²¹。該網站除提供線上查詢 CBDB 資料的界面外，也提供 SPARQL 的查詢界面，讓進階使用者可以更進一步的依個人需求，運行其需要的查詢，取回對應的資料。另外，該網站也提供 CBDB 人物傳記的鏈結本體模型說明文件，讓在應用該鏈結數據資料時，能得到資料模型本體方面的完整參考資訊。

20. 「下載 CBDB」，網址為：<https://projects.iq.harvard.edu/chinese/cbdb/> 下載 CBDB 單機版，(2020/11/03 上網)

21. 「CBDB 關聯數據」，網址為：<https://cbdb.library.sh.cn>，(2020/11/03 上網)

(三) Open API

為了方便其他應用程式的使用，CBDB 團隊也同時提供公開的 WEB API 服務²²，提供第三方系統透過網址呼叫取得相關人物資料內容。該 API 基本上包含一個服務網址，與 3 個參數，提供使用者取回所查詢的人物相關資料。該服務透過不同參數的設定，可以讓使用者選擇以 CBDB ID 或人物名稱等兩種方式進行資料查詢，並且提供另一個參數來讓使用者控制資料輸出格式，該 API 可能的回應資料格式包括有：HTML、XML 與 JSON。雖然提供的資料查詢面向不多，但因該資料庫的主要使用需求便是以取得完整人物資料，因此雖然 API 數量不多，也應該可以滿足大部分的使用需求。

二、臺灣歷史人物傳記資料庫 (TBDB)²³

有鑑於 CBDB 的成功與臺灣歷史研究的需求，臺灣師範大學臺灣史研究所與該校圖書館於 2017 年起，進行「臺灣歷史人物傳記資料庫」的開發計畫。該資料庫以 CBDB 為模範，希冀於 CBDB 的成功經驗之上，建立屬於臺灣歷史人物的資料庫。資料庫內容以主要以臺灣近年來新修訂的各縣市志為基礎²⁴，抽取文獻內容對於人物的背景與社會、親屬等關係的描述，經結構化後匯集成為「臺灣歷史人物傳記資料庫」。

雖然 TBDB 專案主要是以 CBDB 為樣版，但在成果發行與資料開放的方式上，兩者間有相當大的差異。相較於 CBDB 採用離線版關聯式資料庫作為其主要傳播媒體，TBDB 則改用時下較為流行的網頁資料庫的形式，作為提供使用者查詢資料的主要界面。但由於 TBDB 並未提供單機版的資料庫下載，因此使用者便無法使用以單機版資料庫來取得原始資料。而 TBDB 網站為了彌補使用者無法自由利用資料庫內容進行進階統計分析的缺憾，因此於其網站上也製作了不少統計分析、網絡視覺化等功能。但對於開放資料的部份，便著墨較少。根據謝順宏 (2020) 的說明，該資料庫目前採用的資料紀錄模型是大量參考 CDOC-CRM 模型的紀錄規範，並依鏈結關聯數據的精神所建立的資料表格設計。因此筆者預估該專案成果，未來有可能於以鏈結關聯數據模型的方式提供開放資料分享服務。另外，TBDB 為預留未來與介接外部資源的彈性，而建立了幾個資料存取的 API 服務。包含：回傳卒年分佈資料、回傳指定年份區間之年齡分佈、取得人物、詩社與地域關係等服務²⁵。

22. 「CBDB API」，網址為：<https://projects.iq.harvard.edu/chinese/cbdb/cbdb-api>，(2020/11/03 上網)

23. 「臺灣歷史人物傳記資料庫」，網址為：<http://tbdb.ntnu.edu.tw/>，(2020/10/25 上網)。

24. 目前共收錄有：續修南投縣志、續修澎湖縣志、續修澎湖縣志、續修臺北市志、新修彰化縣志等內容，共收錄有 1911 位人物的詳細資料。

25. 資料內容提到，該論文描述之 API 僅為部份 API，似乎意旨有更多 API 函式，但目前無法找到更多資料 (謝宏順，2020，頁 45-46)。

三、臺灣傳記知識本體資料庫 (TBIO)²⁶。

臺灣傳記知識本體資料庫專案計畫，由 Táňa Dluhošová 於 2017 年著手建置，其內容以紀錄臺灣精英人物之資料傳記為主。截至 2020 年為止，資料庫內容供收錄約 27,479 位人物、45,671 個組織與約 2,740 不同職位的資訊。

有別於前述的 CBDB 與 TBDB，此資料庫捨棄了以傳統關聯式資料庫的建構方式，而直接採用鏈結數據作為資料的後端基本資料的建置。也因此，本專案也透過提供符合鏈結資料運作方式的 SPARQL 查詢界面與 Endpoint 連接點，來提供其開放資料服務。另外，網站上也同時提供的許多的查詢範例供使用者參考。使用者可以將查詢結果以 Gephi 相容或 TSV 的格式匯出，方便使用者應用於後續的網絡視覺化與分析作業。

四、佛學規範資料庫²⁷。

佛學規範資料庫是為法鼓文理學院所建立的佛學參考資源之一，其下分為時間、地點、人物與佛經目錄等四個子資料庫。此資料庫建立之目的，就是要提供佛學文獻中所提及的人名、地名、時間、經典等實體的基本背景資料，並賦予每個實體一組不同的規範編碼用以區別不同實體。該專案十分重視資料的開放運用，因此建置了也多樣化的開放資料服務，簡述如下。

(一) 原始資料下載

規範資料庫網站提供非常完整的資料庫匯出檔案²⁸，人物與地點資料庫的匯出格式為符合 TEI 格式的 XML 檔案，而時間資料庫則提供資料庫共通的 SQL 資料格式匯出資料。同時，這些匯出資料也透過系統程式，同步掛載於該資料庫專案於 GitHub 網站內的資料倉庫中²⁹。而匯出資料的更新頻率也相當頻繁，人物與地點的資料因變動較多，因此該團隊每月會進行一次資料庫內容的自動打包與更新。同時網站上也提供了時間規範資料庫的關聯分析圖 (E-R Diagram) 與資料庫綱要 (Database Schema) 等文件，供他人應用這些匯出資料時參考之用。而資料授權部份，這些原始資料的匯出，則是採用創用 CC 授權 BY-NC-SA 2.5 的條款釋出。

26. 「臺灣傳記知識本體資料庫專案計畫」，網站，<http://tbio.orient.cas.cz/>，(2020/10/25 上網)

27. 「佛學規範資料庫」，網站，<https://authority.dila.edu.tw>，(2020/11/15 上網)。

28. 「佛學規範資料庫 Open Content Project」，網址：http://authority.dila.edu.tw/docs/open_content/，(2020/11/15 上網)。

29. GitHub 網站為目前全世界最大的程式碼存放網站和開源社群集中地，放置於該網站的開放專案內容，可以讓全世界的程式設計師以極方便的方式來取得。佛學規範資料庫於 GitHub 上的資料倉庫網址為：<https://github.com/DILA-edu/Authority-Databases>，(2020/11/15 上網)。

(二) Open API

為了方便其他網站工具動態介接資料庫，因此該資料庫也提供 Web API 界面，供其他系統取所需之參考資料³⁰。而該資料庫針對人物、時間、地點等資料庫內容，都分別提供一隻獨立的 API 函式。每個函式的設計，都是以實體的規範碼作為參數，並將傳回該實體的相關細節內容。

(三) 鏈結數據服務

為了提供更多元的資料服務，該團隊也於 2020 年秋天，將該規範資料庫內的人物與地點資料，轉換為開放鏈結數據資料的格式，並將轉換結果匯入鏈結資料專屬之資料庫 (Triplestore)，建立了名為「DILA 鏈結開放資料平台」的開放鏈結數據服務³¹。該系統主要提供三大功能：資料搜尋功能、資料集下載與 SPARQL 查詢界面。同時間系統也提供完整的 Ontology 模型說明，以方便其他系統使用。

五、中央研究院數位文化中心鏈結開放資料平台³²

此資料開放平台由中研院數位文化中心的鏈結開放資料實驗室所製作。該實驗室於 2018 年成立後，便積極與中研院院內各資料庫洽談合作，進行資料庫內容之鏈結開放資料格式的轉換，並提供以鏈結開放資料為基底的資料開放服務。截至筆者撰寫論文之日為止，該資料庫已經完成 13 個資料集，總和超過 60 萬筆資料的鏈結開放資料的資料轉置工作。內容含括生物學、人類學、宗教、藝術、影音、歷史等面向，十分精彩多元。

由於該平台本身主打開放資料的名號，自然會相當專注於開放資料功能的建置。因此所有在平台上開放的資料集內容，無論是整個資料集，或單一資料內容，皆可以完整打包下載，其下載格式也包含 JSON、RDF、Turtle、N-Triples 的鏈結資料相關格式下載，可說是非常完整。同時該系統也建立了開放性的 SPARQL Endpoint，並提供許多範例來引導使用者撰寫 SPARQL 查詢。不過該網站並未提供各資料集所定義的資料本體，是美中不足之處。而資料授權部份，則是採用「創用 CC 姓名標示 4.0 國際授權條款授權」的方式來釋出。

30. 「Data Service API」，網址為：<http://authority.dila.edu.tw/docs/services/>，（2020/11/15 上網）。

31. 「DILA 鏈結開放資料平台」，網站，<https://lod.dila.edu.tw/>，（2020/11/15 上網）。

32. 「中央研究院數位文化中心 鏈結開放資料平台」，網址為：<https://data.ascdc.tw/>，（2020/10/25 上網）。

六、小結

茲將上述提及之五個名錄型漢籍數位資源之資料開放策略，整理如下表 1。

表 1、五個名錄型的漢籍數位資源之資料開放服務比較

	中國歷代人物傳記資料庫 (CBDB)	臺灣歷史人物傳記資料庫 (TBDB)	臺灣傳記知識本體資料庫 (TBIO)	佛學規範資料庫	中央研究院數位文化中心鏈結開放資料平台
提供原始資料下載	Yes	No	No	Yes	Yes
下載資料格式	SQL lite	N/A	N/A	XML, SQL	RDF
更新頻率	約 1 年 1 次	N/A	N/A	約每月 1 次	無資訊
提供鏈結數據存取服務	Yes	未來有機會提供	Yes	Yes	Yes
Open API 數目	1 個	2 個	無	3 個	無
資料開放授權	創用 CC 授權 BY-NC-SA 4.0	未提供說明	未提供說明	創用 CC 授權 BY-NC-SA 2.5	創用 CC 姓名標示 4.0 國際授權條款授權

由表 1，我們不難發現到，目前在名錄型的漢籍數位人文資源所建置的資料開放服務中，幾乎每個資源都（預計）提供以鏈結資料格式為基礎的資料開放服務。我們認為這跟名錄型資料庫的內容多半以提供人物、時間、地點、組織、書目、實體物件等實體的概念性資料為主，其資源內容的不僅相當結構化，也十分著重於實體之間關係的梳理。因此這些資料很符合鏈結資料的概念，容易轉換為鏈結資料格式。而另一方面，我們也發現到一些專案項目以十分積極的方式在經營資料開放服務，如 CBDB 與佛學規範資料庫，兩者皆幾乎提供了上述所有資料分享的可能性。或許這樣的策略，是導致我們現在能在許多數位工具中，都能看見這些資料作為重要輔助內容的原因。

肆、全文型漢籍數位資源的開放資料服務

所謂的全文型漢籍數位資源，泛指內容主要由漢文典籍全文資料所組成的資源。相較於名錄型資料多半為具有相同結構的資料序列，全文資料通常具有較為複雜的階層式結構。例如一篇文章可能向上匯總成為書冊，而書冊可能再向上匯總成為叢書的集合。或於一篇文章

的內容包含有多層的章節內容，而章節之下還可細分為段落，段落內包含多個句子等等。因此，全文型的數位資源中，樹狀階層式結構的資料模型反而較容易用於紀錄資料樣貌，而非經常用於表達名錄型資料的表格狀模型。而這樣的差別，也導致全文型與名錄型數位資源在資料分享的策略上，有著相當不同的選擇。以下說明並比較三個經常被提及的漢籍全文數位資源之資料分享策略的異同。

一、中國哲學書電子化計劃 (Chinese Text Project, CTEXT)

中國哲學書電子化計劃資料庫，由 Donarld Sturgeon 製作，原先資料庫的設計重點為提供先秦兩漢諸子百家論著之電子化資料。後來平台內加入群眾共筆功能，並提供獨家設計之高效率中國古代文獻文字圖像之光學辨識技術 (Optical Character Recognition)，大幅降低將圖片文字內容數位化的門檻。也因此吸引到大量讀者參與文獻數位化作業，目前資料庫內容收錄超過 51 億個中文字的資料，是規模十分巨大的漢籍資料庫。而該資料庫同時也提供許多數位人文的分析工具，讓讀者得以利用該網站的內容，輕鬆的進行 N-gram 切割、正規表示法 (Regex) 搜尋、文字相似性、相異性比對、以及網絡圖、文字雲、統計圖等視覺化功能。而在開放內容服務方面，CTEXT 提供兩種方式，以便同時滿足一般用戶與程式開發者的需求。

(一) 原始資料下載

對於一般用戶，CTEXT 提供名為「全文輸出」的插件 (plug-in)³³。使用者在其帳戶內啟動該插件後，在閱覽界面內便會出現「全文輸出」的選項，以使用者下載閱讀中的文件的原始資料。使用者可以選擇下載整部叢書，或單獨的篇章內容。其下載格式為純文字模式，不含任何文字註解與文件格式紀錄。但可能是因為資料庫內容過於龐大，因此系統沒有提供完整原始資料下載的功能，不過使用者仍可以透過各篇章獨立下載的方式，逐一取得所需的內容。在資料授權方面，CTEXT 網站標示其資料是完全開放給個人或學術使用³⁴。

(二) Open API

CTEXT 資料庫也提供名為「JSON API」³⁵ 的 WEB API 資料開放服務。其 API 所提供的功能相當豐富，包含有：取得收錄文獻清單、取得文獻後設資料、取得文獻內章

33. 「中國哲學書電子化計劃 工具應用程式接口」，網址為：<https://ctext.org/tools/api/zh>，(2020/10/05 上網)

34. 「中國哲學書電子化計劃 版權說明」，網址為：<https://ctext.org/faq/zh>，(2020/10/05 上網)

35. 「Chinese Text Project API」，網址為：<https://ctext.org/plugins/apilist/zh/>，(2020/10/05 上網)

節清單、取得章節文字內容、取得文字搜尋結果，等與文獻內容相關之 API 函式。另外還有像是取得字典詞彙列表、字詞的詳細資料等取得參考資料的 API 函式。整體來說，API 數量共計有 12 個，是數倍於前述的名錄型資料庫所建立的 API 函式。這個現象是來自於全文資料通常具有較多不同用途的利用方式，連帶也會產生較多不同的資料取用需求。此外，CTEXT 也更進一步將其提供的 JSON API，包裝成為現今相當流行的 Python 程式語言的相容函式庫，方便開發者利用 Python 連接 CTEXT 資料庫，並取用相關內容。

二、Kanripo (漢リポ, Kanseki Repository)

Kanripo 也是一個相當知名的學術數位全文資料庫，由京都大學 Christian Wittern 教授所製作。該資料庫於 2015 年上線，內容主要包含四部叢刊、四庫全書、正統道藏、道藏輯要、CBETA 電子佛典集成等，共 9357 部叢書。相對於其他前述數位資源都採用以網頁界面搭配後端不公開的資料庫的傳統作法，Kanripo 則採用了另一種相當創新的作法 -- 將所有資料內容，直接放置 GitHub 的 Kanripo 公開倉庫中³⁶。而其前端網頁系統所提供的功能，都是透過直接讀取 GitHub 的倉庫內容來進行。這樣的作法導致 Kanripo 的網頁操作界面並沒有辦法提供太多新穎的使用者功能。但另一方面，因為原始資料本身即存放於 GitHub 公開倉庫中，因此 Kanripo 可以說是完全奉行以資料開放優先的數位資源。而 Kanripo 專案對於開放資料的設計，不僅在於資料存放地點，並且於資料內容也有相當的考量。Kanripo 的資料格式是以純文字紀錄為主，內容也只有少量的額外標籤，並且將每個檔案內容的相關背景資料 (metadata)，直接紀錄於資料檔頭之中，上述兩個要素，使得 Kanripo 的資料內容十分容易解析與再利用。也是因為 Kanripo 資料的高度開放性，因此許多數位人文的工具平台，如 DocuSky、Markus 等工具，紛紛主動撰寫相關程式，以便讓使用者能於該工具平台中，直接抓取 Kanripo 資料庫的內容。至於資料的授權部份，根據 Kanripo 的網站說明，所有 Kanripo 資料庫的內容是利用「非商業性使用 - 相同方式共享 4.0 國際許可協議 (CC BY-NC-SA 4.0)」的授權方式來釋出。並根據其網站紀錄，資料約每年更新一次。

三、CBETA 數位人文研究平台

由中華電子佛典協會所製作的 CBETA 電子佛典集成，可以說是目前世界上漢傳佛教經典數位化資料收錄最完整的資料集合。該協會自 1998 年創立後，便致力於進行漢傳佛教典

36. 詳情請參閱：「3 Using the Kanseki Repository」(<http://blog.kanripo.org/en3.html>)。而 Kanripo 所使用的資料倉庫位置為：<https://github.com/kanripo>，(2020/10/23 上網)。

籍的數位化工作，並約每 2-3 年便會發表新一期的內容更新。而自 2016 年開始，中華電子佛典協會注意到網路使用趨勢的變化，於是便與法鼓文理學院合作，製作完成「CBETA 線上閱讀 (CBETA Online)」網站，提供方便利用的佛典閱讀網站系統。而製作過程中，同時整合多種佛典相關資料，建立 CBETA 資料中心 (CBETA Data Center)³⁷，提供多種開放資料服務。

(一) 全文資料下載

自 2003 年由中華佛典協會所公開正式發行的第一張 CBETA 電子佛典集成光碟開始，在每張光碟內，都便包含有 CBETA 電子佛典集成的完整原始資料。該原始資料的內容，是以符合國際文獻數位化標準 TEI 的 XML 格式檔案來紀錄資料。自 2016 年開始，為了提供更方便的使用，CBETA 效仿 Kanripo 的作法，於 GitHub 服務上建立了一個公開的資料倉庫³⁸，並將原有的資料上傳公開。同時原來約 2-3 年進行資料更新的步調，縮短為每季釋出新資料內容，並依據發行的季度，給予資料版本編號。自 2016 到現在，已經發行過 14 次的更新，每次變動的內容皆經過詳加的比對與紀錄，並於網路上公開版本間的資料差異³⁹。

(二) Open API

除於 GitHub 倉庫提供完整原始資料外，CBETA 也於上面提到的「CBETA 資料中心」，提供各式的開放資料服務，包含提供完整的資料 API 服務⁴⁰。其所提供的 API 函式內容包含讓使用者取得目錄資料與全文內容、進行資料搜尋、資料匯出等，共約 30 餘種函式。該網站建構如此完善的 API 的原因，是因為其前端網頁 CBETA Online 便是透過此套 API 來取得所需資料，同時間他們也將這套 API 對外開放，讓更多需要取用電子佛典資料的同好，可以利用這些 API 製作其他形式的佛典網站。除自定義的 API 之外，CBETA Data Center 也提供了一組相容於 SHINE API 資料存取服務⁴¹。CBETA

37. 「CBETA Data Center」，網址為：<http://cbdata.dila.edu.tw/>，(2020/10/05 上網)。

38. CBETA 於 GitHub 所使用的資料倉庫位置為：<https://github.com/cbeta-org/xml-p5>，(2020/11/05 上網)。

39. CBETA 的資料更新差異紀錄於：http://cbetaonline.dila.edu.tw/doc/zh/07-14_2020-q3.php，(2020/11/05 上網)。

40. 「CBETA API」，網址，<http://cbdata.dila.edu.tw/v1.2>，(2020/11/05 上網)。

41. SHINE API 是個非常新穎的資料交換標準，由德國馬斯普朗克研究院科學史研究所 (Max Planck Institute for the History of Science) 所提出。該架構的目的，就是希望能夠日漸增加的數位人文資源與數位人工具間的資料交換需求，定義出一個共同的交換標準，減低各家系統資源間的溝通困難。在 SHINE API 的規範中，將資料分為 Collection(集合)、Resource(資源)、Section(區塊)、Content_unit(內容單位) 等四個類別。並且定義這四種類型資料的上下關係，其目的就是用來表達書籍內外部的階層結構。而在實做上，CBETA 將佛典文獻中的藏經集合、經典、各分卷、分卷文字內容映射到上述的 SHINE API 所要求的四個類別資料中，並實做了取得各類別清單與文字內容資料等五個 API。CBETA SHINE API，網址為：http://cbdata.dila.edu.tw/v1.2/static_pages/rise-shine，(2020/11/05 上網)。

目前已經通過 SHINE API 認證，並且已經註冊為其合法的資料來源。因此所有具有存取 SHINE API 資料源的能力的工具程式，皆可以輕易存取到 CBETA 的資料內容。目前 CBETA 資料已經藉由此途徑完成與 Docusky、MARKUS 等知名數位人文工具的介接。

四、小結

茲將上述提及之全文型的漢籍數位資源之資料開放服務，整理如下表 2。

表 2、全文型的漢籍數位資源之資料開放服務比較

	中國哲學書電子化計劃 (Chinese Text Project)	Kanripō (Kanseki Repository)	CBETA 數位人文研究平台
提供完整原始資料載	Yes	Yes	Yes
開放資料格式	純文字	純文字	TEI/XML
開放資料取得方式	透過網站功能下載	放置於 GitHub 中	放置於 GitHub 中
更新頻率	無特定說明，隨時可能更新	批次性更新，約每年更新一次。	批次性更新，約三個月會進行一次更新。
提供鍊結數據下載	No	No	No
開放式資料存取 API 數目	約 12 個	無	約 30 個
資料開放授權	自訂授權，完全開放給個人或學術使用	CC BY-SA 4.0	CC BY-NC-SA 3.0

全文型資料庫，由於其全文資料結構複雜、內容較多、使用需求集中於文字內容，且文字之間較不易定義出明確關聯等特性，導致此類性的資源項目較少以提供鏈結資料的格式的開放資料服務。取而代之的是採用完整原始資料開放下載，或進一步提供開放 API 供其他系統介接。且相較於名錄型資料庫，全文性資料庫若有提供 Open API 的開放資料服務時，通常會有數量較為龐大的 API 函式。另外，透過 GitHub 服務公開原始資料，也是許多專案使用的方式。

伍、數位工具應用開放資料存取服務現狀

上述名錄型與全文型在漢籍數位人文資源，雖然資料類型不同，但還是可以將他們歸為以提供資料為主的「資料提供者」的類別。而相對的，近年來也出現純粹定位為輔助人文研究的數位工具。在這類的數位工具中，系統並不預先提供預先處理好的資料，而是讓使用者上載數位文件或抓取由系統中表列的開放資料，然後進行後續分析運用。因此這些數位分析工具，通常扮演的是「資料消費者」的角色。透過了解這些工具如何應用上述「資料提供者」的開放資料服務，我們也能觀察出數位工具偏好的開放資料應用方式。以下說明並比較二個經常被提及的漢籍數位人文工具，與其利用這些開放資料的方式。

一、MARKUS 半自動文本標記系統

MARKUS⁴² 是由何浩洋與魏希德 (Hilde de Weerdt) 的研究團隊所開發數位人文工具。其設計之功能相當單純，主要是讓使用者上傳自己的文本，然後透過該工具所建立自動比對功能，將文章與內建於 MARKUS 的實體清單進行比對，並產生參考連結。經過多年的發展，該工具已提供了許多便利的功能，包括手動、批次標記修改、自訂標籤等功能，讓研究者進行文字標注的工作，變得十分容易。這種讓使用者進行文本內容標記的工具，實際上可以讓研究者運用於許多不同的研究之中，因此頗受現代人文研究者歡迎，有逐漸變成數位人文界主流研究方式之一的趨勢。

在 MARKUS 系統中，有兩個運用到其他數位資源所提供的開放資料服務。首先，第一部份是其內建的地名、人名、時間、官職等資料庫的清單內容。這些清單有部份來自上述的佛學規範資料庫與 CBDB 資料庫等內容。而其取用的方式，便是透過取得上述資料庫所提供的完整原始資料檔案，經過內部解析後，成為 MARKUS 比對用清單的一部分。而另一個用到這些開放資料的部份，就是與其他全文系統所提供的資料 API 相介接，以便使用者的可以主動抓取由其他開放資料來源的全文資料內容，進行資料標記。實際上，MARKUS 也以資料取用者的角色，實做了 SHINE API 的內容。因若資料來源提供了符合 SHINE API 要求的資料提供者函式，則可以輕易與 MARKUS 平台相互連接。透過 SHINE API 的框架，使用者也能於 MARKUS 平台中，抓取 CBETA 資料與 Kanripo 的資料，CBETA 數位平台傳送資料給 MARKUS 平台。同時 MARKUS 平台也利用了 CTEXT 的 API，讓使用者可以在 MARKUS 平台中抓取所需的 CTEXT 資料。

42. 詳情請參考：Ho and De Weerdt(2014)。

二、Docusky 數位人文學術研究平台

Docusky 數位人文學術研究平台是由臺灣大學數位人文中心所開發的學術研究平台，其設計目的為提供一個符合人文學者研究需要，可以進行個人化文獻資料整理與分析的網路平台。以其運行邏輯來說，Docusky 的主要設計可以說是用來儲存並分析利用 MARKUS 平台標記完成的檔案。也就是當使用者利用 MARKUS 進行文獻內容的標記後，可以將檔案直接匯入到 Docusky 平台中，以便建立個人的文獻資料資料庫。建庫完成後，使用者除了能夠進行基本的瀏覽、搜尋之外，也可以利用 MARKUS 工具所完成的標記紀錄，進行以標記實體與文獻後設資料為基礎的過濾、比對、統計分析與視覺化等功能。

在 Docusky 工具操作中，資料來源除了由 MARKUS 檔案匯入之外，Docusky 也積極嘗試連接多種開放資料來源連接，包括透過 CBETA 所提供的 CBETA API，完成與 CBETA 數位研究平台的連接。以及透過 CTEXT API 取得 CTEXT 資料，並且也針對 Kanripo 的資料格式爬梳，提供 Kanripo 的資料匯入功能。而另一個 Docusky 使用開放資料的方式，則是發生在系統所提供的閱讀界面中。當使用者的標記資料內，有紀錄 DILA 規範資料庫或 CBDB 資料庫內的對應實體時，Docusky 的閱讀系統將會提供透過 DILA 規範資料庫與 CBDB 資料庫的 API，動態取得參考資訊的功能。

另外，Docusky 也提供一套稱為 Docuwidget 的自訂 API，該 API 以 javascript 撰寫，其設計目的為讓開發者端引入其網頁中，並按照其規範來與 Docusky 後端資料庫進行溝通，提供外部開發者與 Docusky 系統互動的機會。該函式功能包含系統登入，資料管理與資料擷取等功能。而 CBETA 數位人文平台，也利用此一 API 模型時做了資料推送的功能，讓使用者於閱讀資料時，便可以便利的將閱讀中的文獻，傳送至 Docusky 來建立個人資料庫。

三、小結

在上述兩系統中，我們看到這些數位研究平台，其實相當積極的尋求與外部資料來源的合作，並且致力於運用各系統的開放資料或利用 Open API 來進行系統間的連接。這也讓我們更清楚的看見，在實際的數位人文系統應用中，開放資料確實於跨系統的整合協作的實現，扮演相當重要的角色。不過我們會發現到，對於上述兩個數位平台而言，雖然也有部份功能來自於匯入各資源的開放原始資料，但似乎更趨向於利用 Open API 進行系統間的溝通。

陸、綜合討論與未來展望

在前述章節中，我們將目前常見的漢籍數位人文資源與其開放資料服務的設置與應用方式進行了整理與分析比較，也分別歸納出各類型的數位人文資源對於開放資料服務於設置上或使用上的偏好。總結來說，我們歸納出下面幾個趨勢與未來值得努力的方向。

一、開放資料服務已經是現在進行式

由本文所整理的漢籍數位人文資源所建置的開放資料服務的結果中，我們觀察到，開放資料服務對於漢籍數位人文資源而言，已不僅只是一個闡述理念的口號，而已經是許多數位人文資源實際進行的工作項目。而其中一個值得令人玩味的現象是，似乎知名度越高，越為人所使用的資源，其資料開放程度便越完整。而此現象究竟是因為採取各種開放策略，讓資源容易被使用，而推升了資源的知名度；還是因為資源知名度提昇、吸引了更多使用者後，也連帶產生更多不同需求，而導致需要製作更多開放資料的服務來滿足各種需求。其實原因尚待釐清，但無論如何，目前我們確實可以看見工具與資源之間，相互利用開放資料服務，串連成一個小型的服務網絡的成果。這樣的彼此合作所帶來的連動效果，一定有助於提昇各資源的使用率，並有利於建構更方便利用的數位人文環境。

二、需建立共同的全文資料交換 API 的標準

根據本研究結果，我們發現到數位平台十分熱衷於對於透過各資源的 Open API 取得所需的資料。在我們研究中討論的兩個數位人文平臺，實際上都竭盡所能的利用各式的 API，來提供使用者取用不同來源的資料。但實際上，目前各數位資源所提供的 API 仍具有相當高的異質性。這樣的情形對於平台與資源兩端，都不見得是好事，極可能導致後續整合上與維護上的負擔。因此為了促進工具與資源間能夠更方便的達到資料交換與流程整合的目的，定義一組全文資料交換 API 的標準，有其必要性。

在實務上，以在筆者所知的範圍內，尚未有機構定義出可以應用於所有情形之 API 設計的標準規範，目前僅見一些 API 設計上的指引建議⁴³。此現象的原因可能是因為各家 API 所

43. 例如，Google 提出了 API Design Guide (網址為：<https://cloud.google.com/apis/design/?hl=zh-TW>)，Microsoft 也提出了關於 API 設計的建議指引 (網址為：<https://docs.microsoft.com/zh-tw/azure/architecture/best-practices/api-design>)，但兩者都屬於較為大方向的指引，並非針對某特定工作目標的細節 API 定義。

提供的資料面向、與系統流程差異性極大，極難討論出所有資源能共同遵守的規範⁴⁴。不過但若將需求面縮小，例如針對漢籍全文資料庫，並以如何取得全文資料為目標，就有極有可能訂出一組大家可以接受且能符合所用的 API 界面。例如上述提及的 SHINE API 規範便是一個以取得漢籍全文資料為主的共通 API 設計，該規範目前仍積極群求各組織的認可與支持。無論最後是否由 SHINE API 成為大家認可的標準，但各漢籍數位資源能提供符合相同規範的 API，相信能大大的簡化資料取得的壁壘，讓數位人文的應用發展更加快速。

三、鏈結資料的應用仍然十分缺乏

根據本研究結果，名錄型資料庫似乎都相當熱衷於提供以鏈結資料為基礎製作之開放資料服務。但相對於資料供給端如此熱烈的投入於鏈結資料的製作中，作為資料消費者的數位人文平台卻似乎興趣缺缺。在本研究提及的兩個數位人文資料，皆同時應用到了 CBDB 與法鼓規範資料庫的兩個名錄型資料庫的開放資料服務，但都不是透過該兩資料庫所建置的鏈結資料服務，而是採用下載兩專案原始資料或透過 API 進行整合的方式。這可能是因為兩平台在進行與 CBDB 和規範資料庫之資料整合時，兩者的鏈結資料服務都尚未建立。但無論如何，目前尚未出現真正運用到上述資料庫所建置的鏈結開放資料服務。但實際上鏈結資料本身之高度資料鏈結性與異質資料的整合性，確實是可以被期待的特性，筆者感覺此部份具有一定潛力，值得深入研究。

四、數位人文的資源整合仍待大家努力

本文所提及的具有資料開放服務的數位人文資源，實際上還是只佔人文相關數位資料庫建設成果的一小部份，許多國內重點單位過往所建立的數位資源，多半欠缺在資料開放方面的考量。若能有更多相關單位，願意投入為資料庫製作開放資料的服務，我們應該便可以期待我國的數位人文發展能快速的提昇到另一個水平。

44. 一個較為接近此方向的實際成果，是由 Swagger Open Source 組織，所提出 OpenAPI 3.0 的標準（網址為：<https://swagger.io/specification/>）。但該標準也僅要求各個工具資源在設計 WEB API 時，必須製作可機讀的且具有共同格式的 API 描述文件。該標準的目的是希望規範系統說明文件的格式，未來就有能發展出能讓人類與機器都能夠更了解遠端 API 的使用方式，藉此達到以最小邏輯便能應用該 API 的目的地。

參考文獻

Berners-Lee, T. & Hendler, J. & Lassila, O. (2001). The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. ScientificAmerican.com.

Berners-Lee, T. (2006). Linked Data. Design Issues. W3C. Retrieved from <https://www.w3.org/DesignIssues/LinkedData.html>, 2010-12-18.

DARIAH (2020), DARIAH Strategic Action Plan II 2019-2022, pp.1-5, Retrieved from <https://www.dariah.eu/wp-content/uploads/2020/05/DARIAH-Strategic-Action-Plan-II-2019-2022.pdf>

Google Maps API Blogs(2006), Google Maps API Version , <http://googlemapsapi.blogspot.com/2006/04/google-maps-api-version-2.html>, Last Access: 2020/11/12)

Ho, Hou leong Brent, and Hilde De Weerd. (2014) MARKUS. Text Analysis and Reading Platform. <http://dh.chinese-empires.eu/beta/> Funded by the European Research Council and the Digging into Data Challenge.

Merton, R.K. (1973). The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press.

Open Science and Research Initiative (2014), open science and research handbook, p.3, Retrieved from <https://www.fosteropenscience.eu/index.php/content/open-science-and-research-handbook>

Programmableweb (2013), 9,000 APIs: Mobile Gets Serious , <https://www.programmableweb.com/news/9000-apis-mobile-gets-serious/2013/04/30> , Last Access: 2020/10/03

Sören Auer, Christian Bizer, Georgi Kobilarov, et al. (2007), DBpedia: A Nucleus for a Web of Open Data The Semantic Web, In The Semantic Web, Vol. 4825, pp. 722-735 , doi:10.1007/978-3-540-76298-0_52

謝順宏 (2020)。臺灣歷史人物傳記數位人文系統設計與建置之研究（未出版之博士論文）。國立臺灣師範大學，臺北市。

劉昭麟 (2018)。從《數位人文研究》展望資料開放、工具協同與議題網絡。數位典藏與數位人文 1 期 (2018 年 4 月)，頁 11-50。DOI: 10.6853/DADH.201804_1.0002

【子題四】 學術前瞻議題

**Session IV
Academic Foresight Issues**



陳淑君
Shu-Jiun Chen

陳淑君，現職中央研究院歷史語言研究所助研究員，獲得國立臺灣大學圖書資訊學博士、英國雪菲爾大學 (University of Sheffield) 資訊學所碩士。目前同時兼任中央研究院數位文化中心執行秘書，及國立臺灣大學圖書資訊學系兼任助理教授。自 1998 年開始參與臺灣數位典藏國家型科技計畫，2013 年獲得美國蓋堤研究所 (Getty Research Institute, GRI) 獎助前往 GRI 擔任訪問學者，進行數位藝術史的學術研究，並自 2014 年起獲美國 GRI 聘任為 Getty Vocabularies as Linked Open Data 研究計畫的外聘顧問。專長於數位圖書館研究、知識組織系統、後設資料 (Metadata) 研究與設計、鏈結開放資料 (Linked Open Data)、數位人文研究。目前同時負責規劃與美國蓋堤研究所「藝術與建築索引典」(Art & Architecture Thesaurus) 中文化研究計畫，及主持中央研究院數位文化中心鏈結開放資料實驗室。

Shu-Jiun Chen is an Assistant Research Fellow at the Institute of History and Philology, Academia Sinica, the Executive Secretary of the Academia Sinica Center for Digital Cultures (ASCDC), and also Adjunct Assistant Professor of Department and Graduate Institute of Library and Information Science, National Taiwan University. She received her PhD degree in library and information science from National Taiwan University in 2012. Her research interests include digital libraries, metadata, knowledge organization, linked data and digital humanities. Dr. Chen initiated the Chinese-language Art & Architecture Thesaurus (AAT) research project with the Getty Research Institute in 2008, and she is also the principal investigator of the Linked Open Data Lab in ASCDC.

數位人文環境下 BIBFRAME 書目框架在古漢籍善本鏈結資料的 再用與擴展

BIBFRAME Reuse and Extension for Linked Data of Chinese Rare Books in the Context of Digital Humanities

陳淑君

Shu-Jiun Chen

中央研究院歷史語言研究所助研究員兼數位文化中心執行秘書

Assistant Research Fellow, Institute of History and Philology, Academia Sinica;

Executive Secretary, Academia Sinica Center for Digital Cultures

摘要

本研究目的在於以書目框架 (BIBFRAME) 為基礎，建立古漢籍善本的鏈結資料集，並探索其在數位人文學的運用。以中央研究院傅斯年圖書館國際合作善本資料庫為例，依據 BIBFRAME 2.0 之書目本體為基礎進行對應，並探討 BIBFRAME 2.0 的再用與擴展；最後，建立鏈結開放資料集，並徵求人文學者進行回饋與調整，進而提出數位人文研究的使用案例。本研究結果，提出古漢籍善本知識本體，以及五顆星的鏈結開放資料集，並展示如何以此為基礎支援數位人文研究。

關鍵字：古漢籍善本、鏈結資料、數位人文學、知識本體、書目框架

Abstract

The study aims to construct a linked dataset for the Chinese Rare Books collection based on the BIBFRAME, and examine its applications for further digital humanities research. Taking as a case study the Chinese Rare Books database built with international cooperation by Academia Sinica's Fu Ssu-nien Library, the structure of the bibliographic database were mapped to BIBFRAME 2.0, and the reuse and extension of BIBFRAME 2.0 were explored. Finally, a linked open dataset was constructed and humanities scholars were solicited for feedback and revisions, ultimately producing use cases for digital humanities research. The results of this research put forward a Chinese Rare Books ontology as well as a 5-star linked open dataset, and illustrate how digital humanities research is supported on these bases.

Keywords: Chinese Rare Books, Linked Data, Digital Humanities, Ontologies, BIBFRAME

壹、研究背景

BIBFRAME 是由美國國會圖書館發展的書目知識本體，主要目的是用來取代圖書館社群的機讀編目格式標準。從 2012 年第一版 BIBFRAME 1.0 首次公布後，至 2016 年提出第二版 BIBFRAME 2.0，在圖書館專業與學術社群已有諸多的研究實驗，例如：倫敦大學學院資訊學系發展基於 BIBFRAME 的鏈結開放資料書目資料集（鏈結開放書目資料）、Linked Data for Production（簡稱 LD4P）研究計畫聯合了包括美國國會圖書館及史丹佛大學在內等六個機構共同合作，將圖書館的技術服務生產工作流程，從傳統轉換為以鏈結開放資料為基礎的方式（Schreur, 2008）。雖然目前已經有很多 BIBFRAME 相關研究，大多仍著重於探討此模型的發展史、或與不同格式、模型之間的對應（Kroeger, 2013; Steele, 2019; Xu, Hess, & Akerman, 2018; Sfakakis, Zapounidou, & Papatheodorou, 2020; Malssen, Hunter & Leigh, 2014）。本文旨在探索以語意網為基礎的知識本體及鏈結資料，如何運用於傳統的書目資料，並進而支援數位人文研究。具體而言，本研究以古漢籍善本的書目資料為對象，採用 BIBFRAME 作為知識本體，並轉置為鏈結資料，以此為基礎，探究如何運用語意檢索（SPARQL）及分析工具進行人文學者在不同研究階段的問題探究。本文提出兩項研究問題，包括：

- 一、BIBFRAME 作為書目知識本體，滿足古漢籍善本書目的程度為何？對於無法滿足的部分，如何擴充以維持 BIBFRAME 架構的一致性？
- 二、古漢籍善本書目的鏈結開放資料集如何支援數位人文研究？

貳、研究設計與實施

本文的研究對象是來自民國 96 年至 98 年間，中研院傅斯年圖書館受蔣經國國際學術交流基金會委託，與美國國會圖書館亞洲部（Library of Congress, Asian Division）、普林斯頓大學東亞圖書館（East Asian Library of the Princeton University）、哈佛大學燕京圖書館（Harvard-Yenching Library of the Harvard University）合作建置的「古漢籍善本數位化資料庫」（<http://rarebookdl.ihp.sinica.edu.tw/rarebook/Search/index.jsp>）書目藏品，包含 13 部集叢（series）、1,071 部單行本（monographies）、6,355 部子目（volumes）等 7,439 筆階層式的書目資料，內容涵蓋明朝初年至二十世紀中葉（1368-1953）近六百年時間之書目資訊（包含印刷刊本、手抄本資訊等資訊）。

本研究與傅斯年圖書館合作，以美國國會圖書館 (Library of Congress) 發展之「BIBFRAME 書目框架模式」(Bibliographic Framework, BIBFRAME) 作為鏈結開放資料轉換的模型基礎，並採用 2016 年發布之 BIBFRAME 2.0 版。以「資料開放授權」、「資料剖析」、「資料建模」、「資料清理」、「資料調和」、「資料加值」、「資料轉置」、「資料 SPARQL 語意查詢」、「資料發佈」、「資料應用」等方法與步驟執行本研究。

參、研究結果與討論

一、古漢籍善本知識本體：BIBFRAME 的再用與擴展

本研究以 2016 年由美國國會圖書館 (Library of Congress) 發布的「書目框架」(Bibliographic Framework, BIBFRAME) 2.0 版本為基礎，建構「古漢籍善本知識本體」(Chinese Rare Book Ontology, 以下簡稱 CRBO)，分別由 26 個「類別」(Classes) 及 89 個「屬性」(Properties) 組成 (表 1)，包括作品 (Work)、實例 (Instance)、單件 (Item) 等三大核心類別 (Core classes) (圖 1)。語彙集類型，除了直接採用 BIBFRAME 外，本研究也再用其它語彙集 (諸如：DBpedia、Skos、Time、dcterms、rdf、schema、vra 等)，以及由本研究自行設計的語彙集 (ASCDC)，藉此擴展以完整描述古漢籍善本書目資訊。

表 1 古漢籍善本知識本體的類別與屬性之語彙集類型

類別與屬性	語彙集類型			總計 (個)
	BIBFRAME	其它語彙集	ASCDC (本研究自行設計)	
類別	20	4	2	26
屬性	34	24	31	89

在所有 26 個類別當中，本文直接再用 BIBFRAME 的 20 個類別、4 個外部語彙集，以及 2 個本研究自行設計的類別。採用外部語彙集的目的是為了補充 BIBFRAME 在描述本研究個案的古漢籍善本之不足，例如：以「DBpedia Ontology」描述古漢籍善本中與指涉藏品相關圖像資源概念、以「Time Ontology」描述書籍藏品相關出版時間、貢獻者所屬朝代或生卒時間或區間之類別概念。總計 89 個屬性方面，本文再利用 34 個 BIBFRAME 已有的屬性，不足之處以採用其它語彙集補充 (24 個屬性)，或自行設計 (31 個屬性)。例如：以「skos」描述古漢籍善本作品概念及實體所屬附註補充資訊、以「ascdc」的不同屬性描述善

本古籍作品概念及實體相關版本及版型設計、印記、避諱、貢獻者所屬朝代時間、典藏分類及典藏取得方式等資訊。

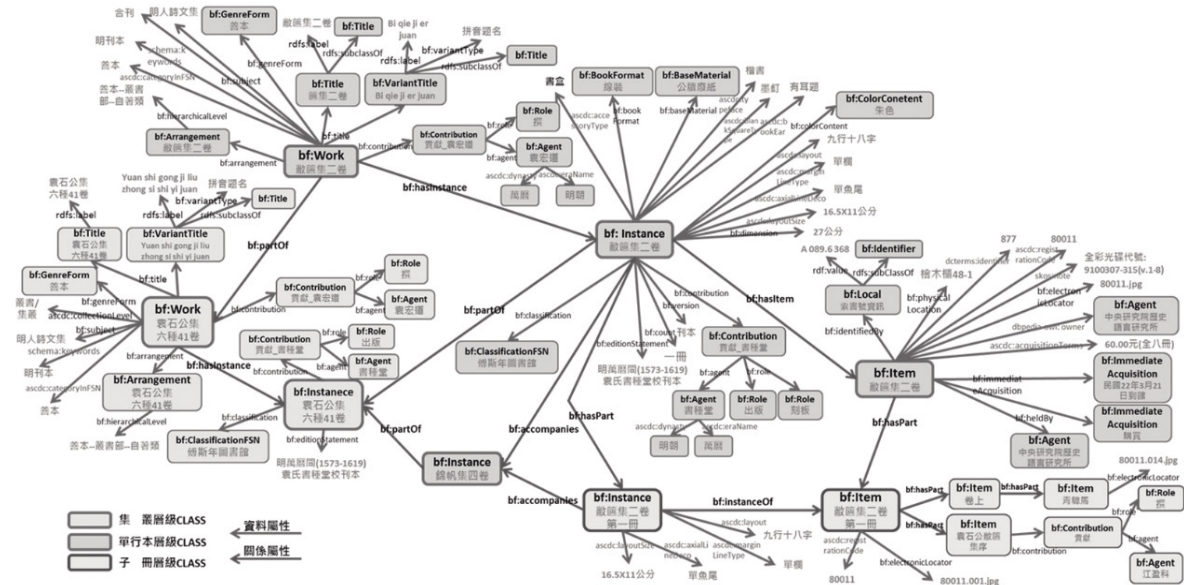


圖 1 古漢籍善本知識本體的結構示意圖

以集叢《袁石公集六種四十一卷》中的單行本《敝篋集二卷》架構為範例示意。灰字為欄位資料值或欄位實例

(一) 書目結構的對應與融合

基於本研究個案「古漢籍善本數位化資料庫」的善本古籍群原始書目結構，主要由集叢、單行本、子目（冊）等三個層級組成，藉此顯示藏品間從屬、平行等關係；而 BIBFRAME 書目框架則是以規範單一書籍作品、實例、單件等核心類別架構，藉以描述書籍由抽象著作概念至單一實體館藏間的資訊內容及關係。本研究提出 CRBO，結合原有集叢、單行本、子目等架構，並整合至以 BIBFRAME 為基礎的作品、實例、單件等核心類別（圖 2）。

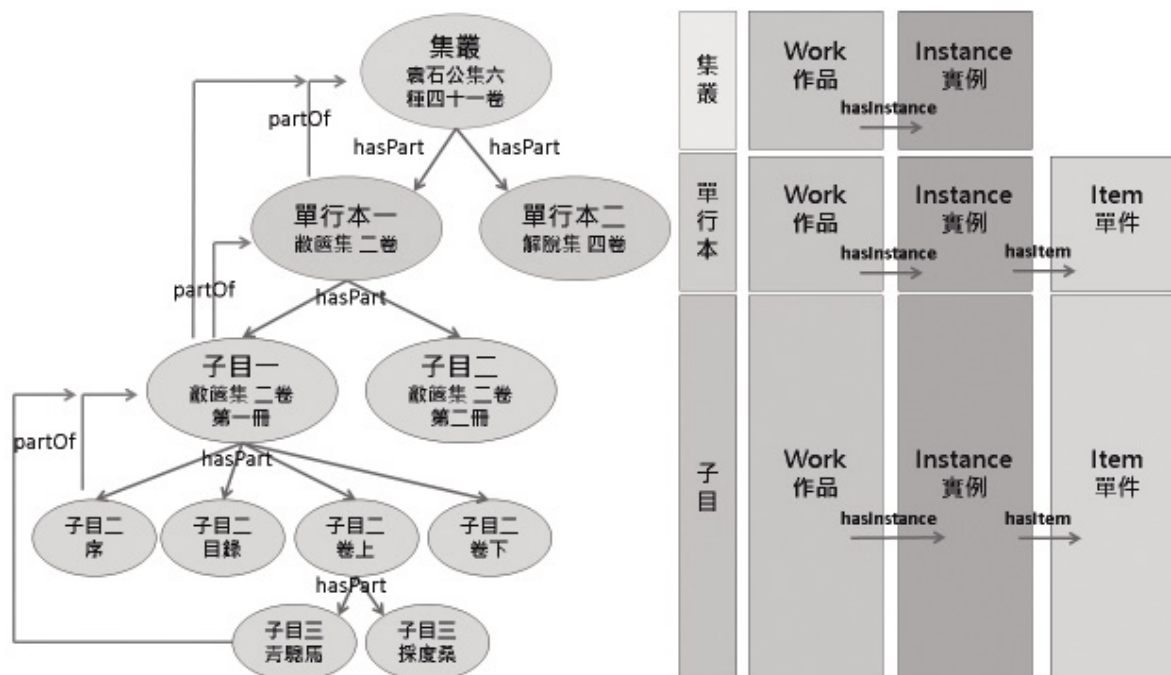


圖 2 古漢籍善本藏品架構對應 BIBFRAME

以《袁石公集六種四十一卷》為例，書目的「集叢」階層可對應於BIBFRAME的「作品」與「實例」層級資訊。集叢項下所屬「單行本」，如：《敝篋集二卷》則對應於「作品」、「實例」及「單件」等層級資訊；而單行本下所屬「子目」，如：《敝篋集二卷第一冊》則亦對應於「作品」、「實例」及「單件」等層級資訊。

(二) 古漢籍善本知識本體模型—Work 核心類別

現有古漢籍善本模型對應、整合 BIBFRAME「作品」(Work) 類別模型中，主要包含書目所屬著作各類型題名、主題內容、分類、著作者、摘要、語言等相關欄位資訊內容。而古漢籍善本各「集叢」、「單行本」、「子冊」層級中所包含、對應於 BIBFRAME 之「作品」與「實例」間資訊則以「具實例」屬性 (bf:hasInstance) 相連，表達兩者間向下之整部從屬關係 (圖 3)。

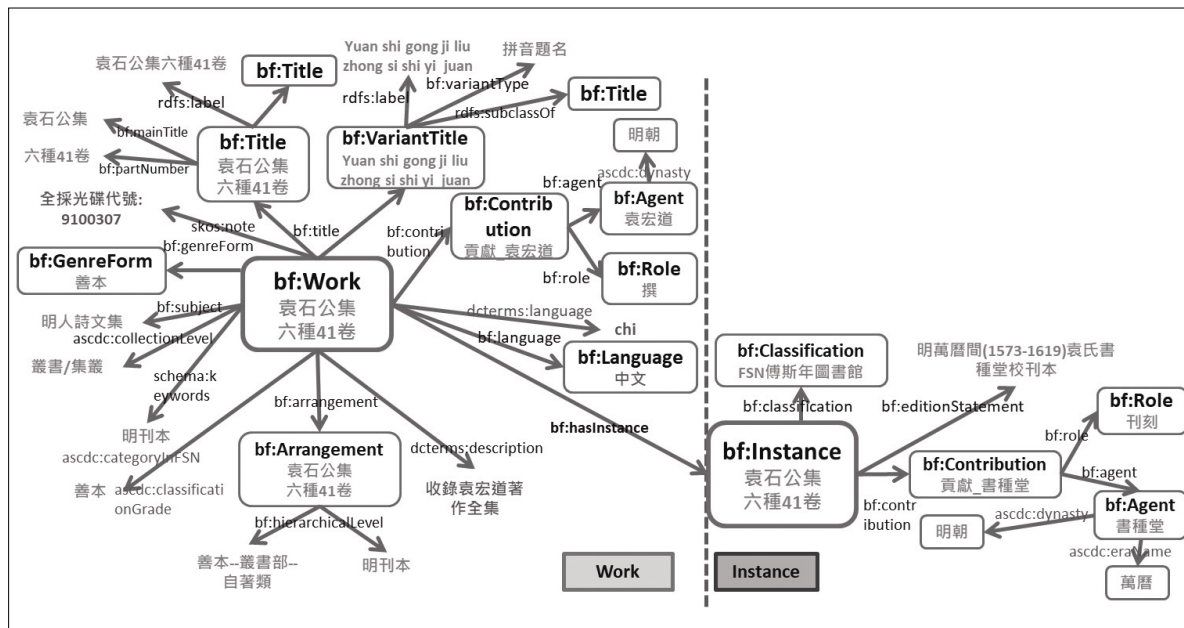


圖 3 古漢籍善本資料架構 Bibframe 模式 - Work 核心模式架構

以古漢籍善本集叢《袁石公集六種 41 卷》示意。灰字為欄位資料值或欄位實例，綠字為該欄位可對應 Work 的欄位資訊，但本範例中並無該欄位資訊，以虛擬方式顯示

(三) 古漢籍善本知識本體模型—Instance 核心類別

現有古漢籍善本模型對應、整合 BIBFRAME 「實例」 (Instance) 類別模型中，主要包含書目所屬版本分類、版本裝幀、型制、設計、材質、版本出版地、出版參與者、避諱等相關欄位資訊內容。而古漢籍善本各「集叢」、「單行本」、「子冊」層級中所包含、對應於 BIBFRAME 之「作品」與「實例」間資訊則以屬性 `bf:hasInstance` 相連，表達兩者間向下之整部從屬關係；而「實例」與「單件」間則以屬性 `bf:hasItem` 相連，表達兩者間之從屬關係 (圖 4)。

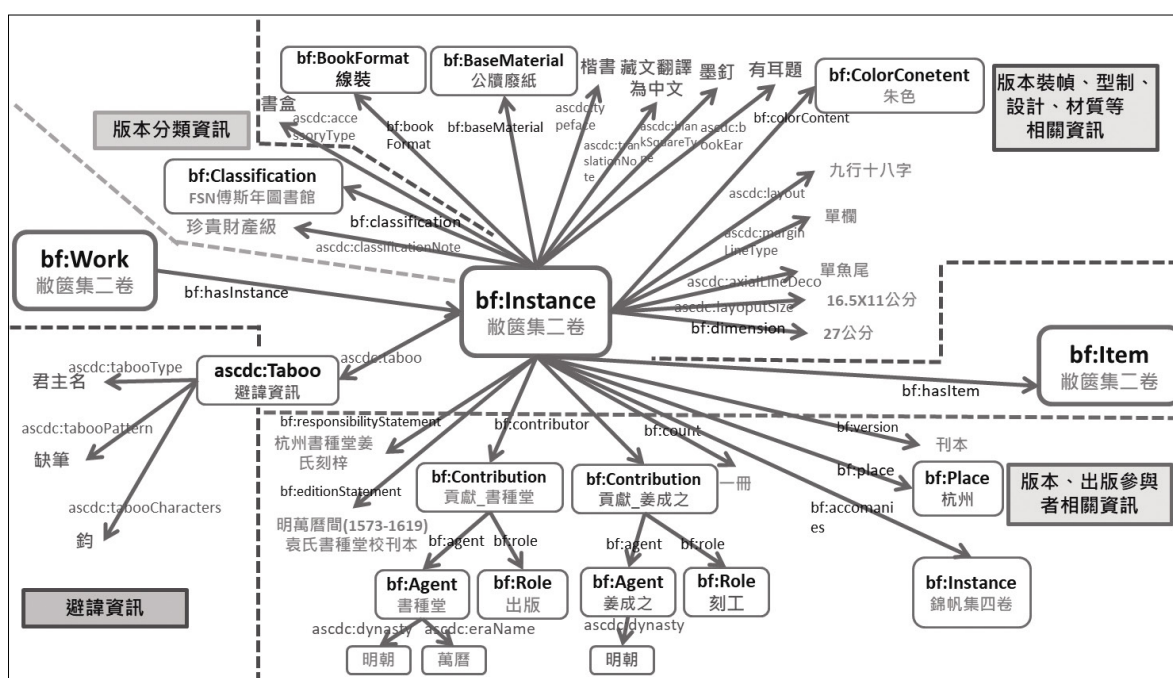


圖 4 古漢籍善本資料架構 Bibframe 模式 - Instance 核心模式架構

以古漢籍善本單行本《敝篋集二卷》示意。灰字為欄位資料值或欄位實例，綠字為該欄為可對應 Instance 的欄位資訊，但本範例中並無該欄位資訊，僅以虛擬方式顯示

以單行本「敝篋集二卷」為例，在 IBFRAME 模型「實例」層級，如欲描述其所屬書目裝幀資訊，可以利用「裝幀類型」屬性 (bf:bookFormat) 描述「敝篋集二卷」作品 (bf:Instance) 所屬「裝幀類別」(bf:BookFormat) 之實例值為「線裝」。藉由單階段 RDF 三元組描述方式，達成「敝篋集二卷」書目作品裝幀類型為「線裝」的語意表述。

(四) 古漢籍善本知識本體模型—Item 核心類別

在 BIBFRAME 模型「單件」層級架構，如欲描述現有古漢籍善本藏品所屬索書號資訊，首先需利用「識別碼資訊」屬性 (bf:identifiedBy) 描述「敝篋集二卷」作品 (bf:Work) 所屬「索書號資訊」實例 (bf:Local)；其次再運用「資訊值」屬性 (rdf:value)、「超類別資訊」屬性 (rdfs:subClassOf)，分別描述該「索書號資訊」實例所屬之「索書號值」為「A 089.6 368」及「超類別類型」(bf:Identifier) 為識別號。藉由這種兩階段 RDF 三元組描述方式，達成「敝篋集二卷」書目作品索書號為「A 089.6 368」的語意表述 (圖 5)。

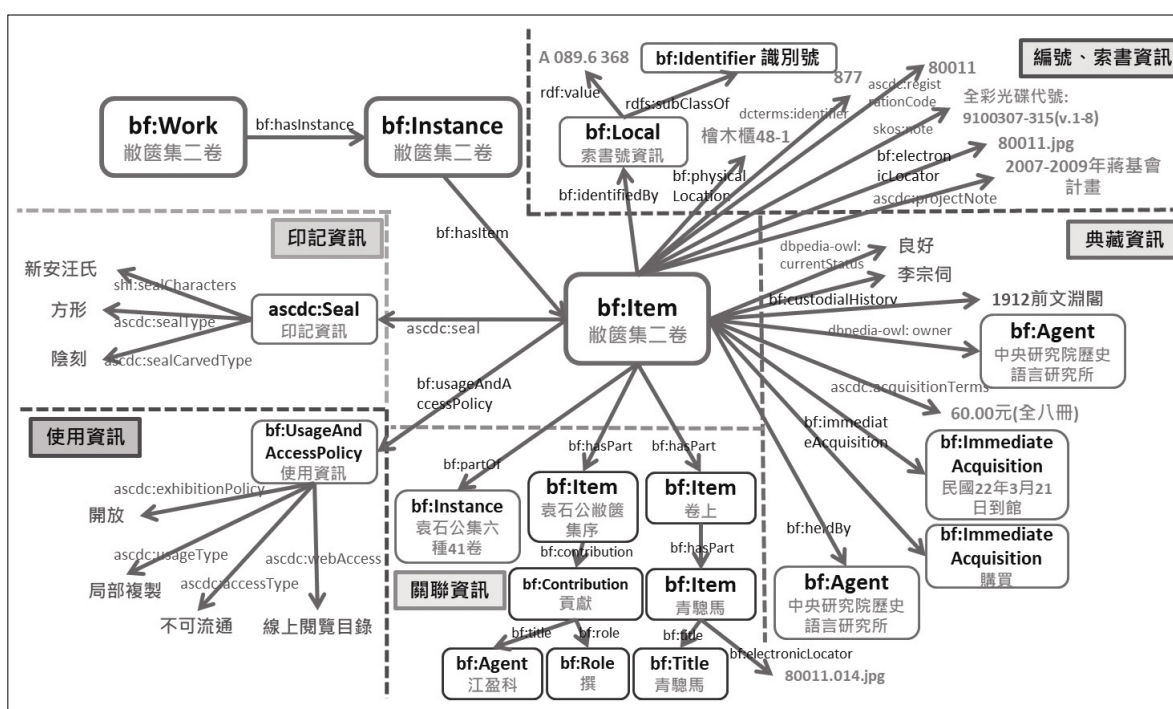


圖 5 古漢籍善本資料架構 Bibframe 模式 - Item 核心模式架構

以古漢籍善本單行本《敞篋集二卷》示意。灰字為欄位資料值或欄位實例，綠字為該欄為可對應 Item 的欄位資訊，但本範例中並無該欄位資訊，僅以虛擬方式顯示

二、古漢籍善本知識本體：BIBFRAME 的擴充

由於 BIBFRAME 模型架構及內容元素的設計，主要是由現代圖書館書目資訊及西方現代出版品資訊管理觀點為出發，故模型中包含之「類別」、「屬性」等元素並無法完整包含、指涉東方長期歷史文化脈絡及出版環境下所孕育之特有書目資訊分類觀點。針對 BIBFRAME 模型對描述中文善本古籍特殊書目欄位之不足，在進行「古漢籍善本知識本體」規劃設計過程中，本文分別就「作品」、「實例」、「單件」等三種核心類別進行延伸、擴充設計，滿足語意描述中文善本古籍特有欄位資訊之需求。

(一) 延伸擴充說明

在以 BIBFRAME 模型架構為基礎的「古漢籍善本知識本體」設計上，可由類別屬、性兩方面進行擴充：

1. 類別

「古漢籍善本知識本體」類別設計主要採用 BIBFRAME 2.0 版規範的「作品」、「實例」、「單件」等三種層級式「核心類別」進行建構，在現有知識本體的

26 個類別，其中 20 個類別是再利用 BIBFRAME 現有設計。雖然 BIBFRAME 既有類別架構設計大致可滿足東方善本古籍資訊描述需求，但在描述藏品所屬「避諱資訊」、「印記資訊」時，尚無法滿足這兩種東方善本書籍特色之類別需求，因此本研究在「古漢籍善本知識本體」設計過程中以延伸 BIBFRAME 方式新增類別、補充知識本體對這些東方書目在地特色進行完整描述。

以描述書目「單件」層級中的「印記資訊」類別為例，「古漢籍善本知識本體」以自訂的類別「印記」(ascdc:Seal) 表示某書目「單件」層級資源所屬的印記資訊實例 URI，兩者間同時利用自訂的屬性「印記」(ascdc:seal) 進行資源實例間的鏈結。以哈佛大學燕京圖書館藏為例，明朝崇禎年間刊印、由浙江文人張毓睿撰寫《三國志纂八卷》單件藏本上蓋印「楊晉印信」四字章記(圖 6)。首先，在該印記資訊的描述過程，依據知識本體架構設計，利用自訂延伸屬性「印記」(ascdc:seal) 描述《三國志纂八卷》所屬的印記資訊實例資源；其次，利用延伸設計之屬性 ascdc:sealForm、ascdc:sealCarvedType 分別表示該印記資訊實例資源所屬的「形狀」(如方形)、「刻法」(如陰刻)等資訊。其中印記「刻法」資訊部分，「陰刻」等資料值則可直接再利用外部的語彙資源，如：AAT「藝術與建築索引典」等，達成資料加值、豐富原有書目資料的目的。現有知識本體設計中，印記的刻法屬性並延伸、再利用 SKOS 知識本體中規範之 skos:Concept「概念」類別，整合、鏈結 AAT 收錄之「陰刻」詞彙(Intaglio, AAT300186231)，提升、擴展原有資料內容。

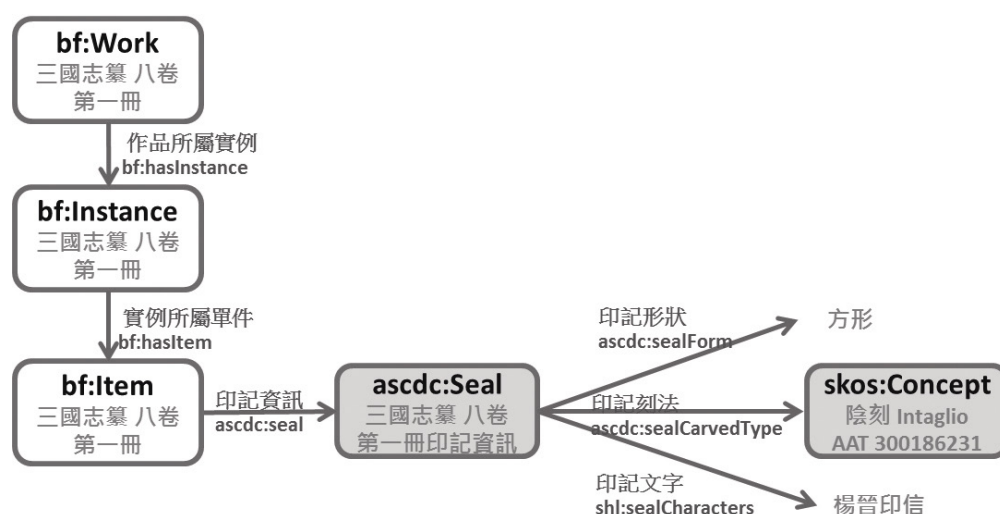


圖 6 「古漢籍善本知識本體」語意型模型設計中，「印記資訊」類別設計內容(灰底方塊為「古漢籍善本知識本體」中，於 BIBFRAME 2.0 延伸之部分)

除上述古漢籍善本書目資源所屬印記資訊 (ascdc:Seal)、避諱資訊 (ascdc:Taboo) 等資訊類別描述延伸設計外，古漢籍善本知識本體中有關書目相關刊印、抄寫等時間資料描述，本研究利用 W3C「全球資訊網協會」規範「時間知識本體」(Time ontology) 架構設計中的「時間實體」(time:TemporalEntity)、「時段實體」(timeProperInterval) 等類別，分別針對 BIBFRAME 結構中，書目「作品」(Work)、「實例」(Instance) 層級中的書目出版年份時間、朝代或年號等相關資訊類別進行延伸、描述。再以前述《三國志纂八卷》為例，該冊書目刊印時間為明朝，故於古漢籍善本實例層級中，利用自訂之屬性 ascdc:dynasty 鏈結至表示書目出版朝代資訊的「時段實體」實例資源中，藉此鏈結、再利用中心建構之時間知識本體詞彙中收錄之「明朝」一詞詞彙 (ascdc:Dynasty/030)。

此外，現有古漢籍善本類別的延伸設計，則是運用 SKOS 知識本體中 skos:Concept「概念」類別，達成概念導向資源內容加值目的。以清咸豐年間刊印之《論語古韻二十卷》為例，該「實例」層級所屬「字體」欄位資源即利用自訂之屬性 ascdc:scriptType，鏈結至表達該「字體」類型資源的 skos:Concept 類別實例中，以表述其欄位值為「楷書」。鑑於「楷書」一詞亦已收錄於 AAT 索引典中，故在「字體」欄位設計中可藉由 skos:Concept 的導入，整合「楷書」詞彙 (Standard script, AAT300343590)，達成補充原有欄位資料值之可能。

2. 屬性

現有「古漢籍善本知識本體」設計中在屬性元素延伸設計上，主要針對「單件」層級中的「裝飾」(page's decoration)、「版式設計」(layout design)、「避諱」(taboo) 及「實例」層級中的「印記」(seal) 等相關欄位資訊進行新增屬性設計。其中除「避諱」、「印記」等欄位屬性因其東方在地書目文化特殊性，無法在 BIBFRAME 模型中找到可對應之相似屬性元素設計外，其餘如「裝飾」、「版式設計」等屬性類型均屬由 BIBFRAME 模型「實例」層級屬性元素中延伸之類型。

目前 BIBFRAME 屬性，有關書籍裝飾、版式等多種資訊描述內容均只利用單一「版式設計」屬性 (bf:layout) 描述各類「書籍資源中的文字、圖像、觸覺資訊等安排」(Arrangement of text, images, tactile notation, etc., in a resource.)。此一屬性定義粒度設計甚為廣泛、可涵蓋上述自訂「裝飾」及「版

式設計」延伸屬性中有關「耳題」(ascdc:erTi)、「墨釘」(ascdc:moDing)、「書口」(ascdc:mouthOfBook)、「魚尾」(ascdc:yuWei)、「邊欄」(ascdc:frameLineType)等較為細緻化之版型設計相關欄位屬性定義(表2)。

表 2 古漢籍善本語意模型設計中擴充之 BIBFRAME 屬性架構設計類型概覽

擴充類型	屬性	上層屬性	書目層級
裝飾資訊	ascdc:erTi (耳題)	bf:layout (版式設計)	實例 (Instance)
	ascdc:moDing (墨釘)	bf:layout (版式設計)	
	ascdc:mouthOfBook (書口)	bf:layout (版式設計)	
	ascdc:yuWei (魚尾)	bf:layout (版式設計)	
版式設計	ascdc:frameLineType (邊欄)	bf:layout (版式設計)	實例 (Instance)
	ascdc:layoutSize (版框高廣)	bf:dimensions (度量資訊)	
	ascdc:linesPerPage (行格)	bf:dimensions (度量資訊)	
	ascdc:scriptType (書體)	N/B	
避諱資訊	ascdc:taboo (避諱資訊)	bf:Note (註記資訊)	實例 (Instance)
	ascdc:tabooCharacter (避諱字)	bf:Note (註記資訊)	
	ascdc:tabooPattern (避諱對象)	bf:Note (註記資訊)	
	ascdc:tabooType (避諱方式／類型)	bf:Note (註記資訊)	
印記資訊	ascdc:seal (印記資訊)	bf:Note (註記資訊)	單件 (Item)
	ascdc:sealCurveType (印記刻法)	bf:Note (註記資訊)	
	ascdc:sealType (印記形狀)	bf:Note (註記資訊)	

(1) 裝飾資訊屬性延伸設計

在「古漢籍善本知識本體」模型設計中，裝飾資訊意指善本書籍翻開書頁中，在書頁邊框版式線條上或線條中出現或包夾之細部特殊裝飾標記，主要包含「耳題」、「墨釘」、「書口」、「魚尾」等部分。這些邊框中的細部裝飾或用以承載資訊(如「耳題」、「書口」)，或做為書頁對折的標準(如「魚尾」)。以「魚尾」為例，這類裝飾出現在善本書籍版心中央部位，以類似魚尾開岔的外型標註出每張書頁對折之中對折之中線(嚴佐之，1989)。

鑑於這種「魚尾」裝飾資訊為中國傳統善本書目資訊中甚為特殊之欄位類型，在現有 BIBFRAME 語意模型中無法對應近似、可再利用之屬性，因此本文「古漢籍善本知識本體」以自行設計屬性 `ascdc:yuWei` 之方式，在善本書目「實例」(Instance) 層級中描述某書目資源所屬之「魚尾」欄位裝飾資訊，以文字方式鏈結、帶出書目「魚尾」欄位類型(如「單魚尾」、「雙魚尾」)(表 3)。

表 3 「古漢籍善本知識本體」裝飾資訊型屬性延伸設計範例(以「魚尾」為例)

ascdc:yuWei/ 魚尾	
URI	<code>http://data.ascdc.tw/property/yuWei</code>
名稱 Label	Yu Wei (Axial Line Decoration)/ 有魚尾版式類型
類型 Type	Property
定義 Comment	描述古漢籍善本版式所屬魚尾類型資訊
定義域 Domain	<code>bf:Instance</code>
值 域 Range	Literal
量詞 Quantification	0-n
資料類型 Value type	Literal (@zh)
範例	七經偶記十二卷第一冊 有魚尾版式類型 (yuWei) 單魚尾

(2) 版式設計屬性延伸設計

版式，特指印版書頁中，每塊書版的版面格式，這類資訊主要包含「邊欄」、「版框高廣」、「行格」等，也包含版面中刊印使用的「書體」類型，可謂善本書籍版面中的主要設計架構(嚴佐之，1989：8)。古漢籍善本版式中，「行格」意指每個印版書頁所組成的垂直行數及每行中所刊印之字數，行格字少，版面即顯得醒目、行格字多，版面就顯侷促。在中國傳統古籍研究中，善本版式資訊是鑑定版本的重要依據之一，行字數目多寡反映出書籍出版經費充裕性、出版作坊之出版策略，或整體出版環境興盛、蕭條與否。

本研究延伸設計「版框高廣」(`ascdc:layoutSize`)、「行格」(`ascdc:linesPerPage`)等屬性，分別用以描述東方善本書目中特有，刊本書目每個印版頁面邊欄範圍的長寬資訊及印版書頁內的垂直行數及每行中刊印之字數等欄位資訊。這類度量相關之屬性資訊粒度性、專指性甚高，相較 BIBFRAME 模型，僅有「實例」層級中表達「書籍資源載體或容納物之的

度量」(Measurements of the carrier or carriers and/or the container of a resource.) 的粗粒度「度量資訊」屬性 (bf:dimensions) 可模糊對應。故古漢籍善本本體中 ascdc:layoutSize 及 ascdc:linesPerPage 等屬性可視為 bf:dimensions 之延伸。

(3) 避諱資訊屬性延伸設計

「避諱」是中國古代封建社會中，由周朝開始，盛行唐宋、乃至明清時期的特殊書寫、刊印文化風俗。礙於不得在文字創作中書寫與當代統治者相同之尊名，舉凡文中如需使用與統治者相同、甚至形音相近之文字時，須以缺筆、改字、代字、留白或其他方式進行替換（嚴佐之，1989：133-134）。因此在傳統善本書籍版本研究中，包含「避諱字」、「避諱對象」、「避諱方式」等資訊就成為辨別善本書籍內容真偽、創作時代的重要參考資訊。

「避諱」相關類型資訊因涉及特有之中國傳統政治、文化發展背景，故同樣無法在現有 BIBFRAME 語意模型中獲得可滿足、再利用之屬性及類別元素對應。援此，以「避諱字」欄位資訊描述為例，在現有「古漢籍善本知識本體」模型「實例」層級中，亦利用自訂規範屬性 ascdc:tabooCharacter，以文字值方式描述某書目版本實例資源中所屬「避諱字」資訊（表 4）。

表 4 「古漢籍善本知識本體」避諱資訊屬性延伸設計範例（以「避諱字」為例）

ascdc:tabooCharacters/ 避諱字	
URI	http://data.ascdc.tw/property/tabooCharacters
名稱 Label	Taboo Characters/ 有避諱字
類型 Type	Property
定義 Comment	描述古漢籍善本所屬之避諱文字內容
定義域 Domain	ascdc:Taboo
值 域 Range	Literal
量詞 Quantification	0-n
資料類型 Value type	Literal (@zh)
範例	文淵殿上卷第一冊避諱資訊 有避諱字 (tabooCharacters) 玄

三、古漢籍善本書目的鏈結開放資料集與數位人文研究

數位人文研究有賴於學者的問題意識、高品質的巨量資料，以及適當的數位分析工具。本文以 BIBFRAME 2.0 知識本體為基礎，將古漢籍善本書目轉換為鏈結資料集，並從數位人文視野，以研究者可能感興趣的問題意識為基礎，發展不同樣式的「Sparql 語意查詢樣版」（表 5）。讓此資料庫成為知識庫，並提供語意豐富的探新功能，提供研究者利用此資料庫查詢與取得資訊的過程，提供使用介面與工具，以協助其回答複雜的研究提問。本文依研究者常用檢索需求（如各種或特定屬性資訊）、問題意識進行設計，並與領域專家確認問題設計正確性，以及對善本書目研究者在執行數位人文研究的意義。語意檢索範例的命題，涵蓋古漢籍善本基本資訊、版本研究、書目著作群體社會關係、跨資料集查詢、版式裝幀等等研究命題。議題設計除考量能涵蓋善本書目研究範疇、角度的多元性、廣泛度外，並得以反映不同研究階段間問題意識深淺度設計，包含不同研究命題深淺度。

表 5 「古漢籍善本 LOD 資料集」檢索範例設計、研究議題及研究階段反映概覽

編號	檢索範例	研究議題	研究階段
1	善本資料集中收錄的各單行本的出版年份，分別位於哪些皇帝統治期間？各有多少數量？	目錄學 - 出版史 - 思想文化	研究初步查考階段，輔助研究者從大方向掌握各時期書籍出版數量分布
2	善本資料集中收錄的各單行本的出版地點，分別位於今日的那些地區？各有多少數量？	目錄學 - 出版史 - 書籍史	研究初步查考階段，藉由統計結果呈現作為研究主題
3	清道光年間出版書籍有哪些？請列出主要題名與卷端題名。	目錄學 - 出版史 - 學術思想史	研究拓展的階段
4	列出古漢籍善本資料集的集叢、單行本與子目一層次中，所有亦出現在明清人名權威檔 LOD 資料集的著述名稱。	目錄學、善本書目架構研究、跨資料集查詢	研究基礎階段
5	在古漢籍善本 LOD 資料集中，各單行本分別有哪些貢獻者？又從曾共同參與同一單行本的次數來看（含撰述、序跋題記…），貢獻人物彼此間關聯性如何？	目錄學 - 出版史 - 合作關係、文人史 - 文人社群（文人網絡）- 學術思潮、善本書目著作群體社會關係研究（SNA）	完整研究問題，並具清晰問題意識

編號	檢索範例	研究議題	研究階段
6	古漢籍善本資料集中所收錄的“嘉興府志”版本為何，又該作品在美國國會圖書館典藏目錄 (Library of Congress) 中還有那些其他版本？	善本書目版本研究、 跨資料集查詢	基礎問題，可輔助文獻學學者進行資料查考，具協助研究的價值
7	古漢籍善本中那些單行本子目書籍同時具有雙魚尾、雙欄等版式裝飾？這類書籍所屬出版朝代資訊為何？	善本版式裝幀研究	深度研究問題拓展，輔助進行古籍出版景氣環境、版式設計演變

以編號 5「在古漢籍善本 LOD 資料集中，各單行本分別有那些貢獻者？又從曾共同參與同一單行本的次數來看（包含撰述、序跋、加工出版等），貢獻人物彼此之間的關聯性如何？」為例，本文從善本書目著作群體社會關係研究的角度，以原始資料集中所著錄的書誌資料，提取書籍序跋、校勘、編纂等出版貢獻者資訊，進行社會網絡分析。本研究將此命題以語意檢索 (SPARQL) 古漢籍善本鏈結資料集的各單行本資訊後，將搜尋到的資料再運用社會網路分析工具，以視覺圖像探究明中國過往歷朝文人間（明清時期為主）因參與書籍刊印、著述創作而形成之人際網脈關聯及流動性。

具體而言，本文嘗試呈現「古漢籍善本鏈結開放資料」中，單行本書目共同貢獻者間合作關係與外部社群樣貌，有關本項命題研究實作過程茲概述如下：

（一）資料篩選

研究執行過程中，首先以 SPARQL 語意檢索方式，查詢「古漢籍善本鏈結開放資料」中集單行本貢獻者資訊，總計獲得 1,469 位、其次將分析重點聚焦於較為活躍貢獻者間之互動連結。本處以單一書目中貢獻紀錄達兩筆作為條件，進行貢獻者的篩選，共得 122 位文人貢獻者（圖 7）。

The screenshot shows the BIBFRAME SPARQL Query interface. On the left, there are example queries and their explanations. The main area contains a SPARQL query input field with the following query:

```

1 PREFIX bf: <http://id.loc.gov/ontologies/bibframe/>
2 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
3 PREFIX dcterms: <http://purl.org/dc/terms/>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX ddpedia: <http://dbpedia.org/ontology/>
6 SELECT ?id ?contributor ?dynasty ?role
7 FROM <http://data.ascdc.tw/Book/crb/>
8 WHERE{
9   ?s dcterms:identifier ?id.

```

Below the query input, there are tabs for 'Triples' and '關係結構圖'. The 'RESULTS' tab is selected, showing a table with the following data:

id	contributor	dynasty	role
102487	林起龍	清	纂
102487	林起龍	清	鑑定
102138	鄭良	清	纂
102486	鄭良	清	纂
102487	鄭良	清	鑑定

圖 7 古漢籍善本單行本貢獻者語意檢索資料篩選（以範例命題「在古漢籍善本 LOD 資料集中，各單行本分別有哪些貢獻者？又，從曾共同參與同一單行本的次數來看（包含撰述、序跋題記……），貢獻人物彼此間的關聯性如何？」為例）

（二）料篩選結果匯出

將上述 SPARQL 語意檢索結果資料匯出、下載，以進行資料後續再處理工作，並做為再處理後資料匯入 Gephi 軟體、執行社會網路分析之基礎。後續與領域專家合作，確認檢索資料正確性並預先整理。

（三）資料再處理

資料匯出後，將資料內容依據任兩貢獻者間對應到相同單行本者進行分組，並將沒有與其他位貢獻者共同對應到任一單行本的文人資料移除，以進行網絡建構。將網絡分析結果，對照歷來文學史記述，從大方向印證文學史、文人史交游考察，透過書誌資料，發掘文人於出版事業上經由合著、協著書籍，所呈現之具體往來事實，並加以提示歷來文人史研究所未觸及之面向，尋繹文人於出版活動往來的實際線索。

「古漢籍善本鏈結開放資料」中，有關單行本書目共同貢獻者間合作關係社群樣貌經上步驟處理、過濾並套用 Gephi 軟體繪製後，總計於網路圖上產生 69 個節點，代表該資料集中善本書目出現共同貢獻參與之總人數。而各人物節點對應其所處朝代，分別賦予不同顏色標示。如橘點代表宋朝人物、紫點為明朝、藍點為明末清初、綠點為清朝人物等區分。另外，由於原始書目後設資料中並包含書目各貢獻者參與書目創作之角色資訊（如序、跋、撰等等），考量人物間合作貢獻類型的不同，經與領域專家請教後，初步區分為序跋、合作序

跋、合著、加工出版、編、合作加工出版等六種合作關係類型，並分別以不同顏色標記，作為標示人物間邊線所指涉之合作關係類型（圖 8）。

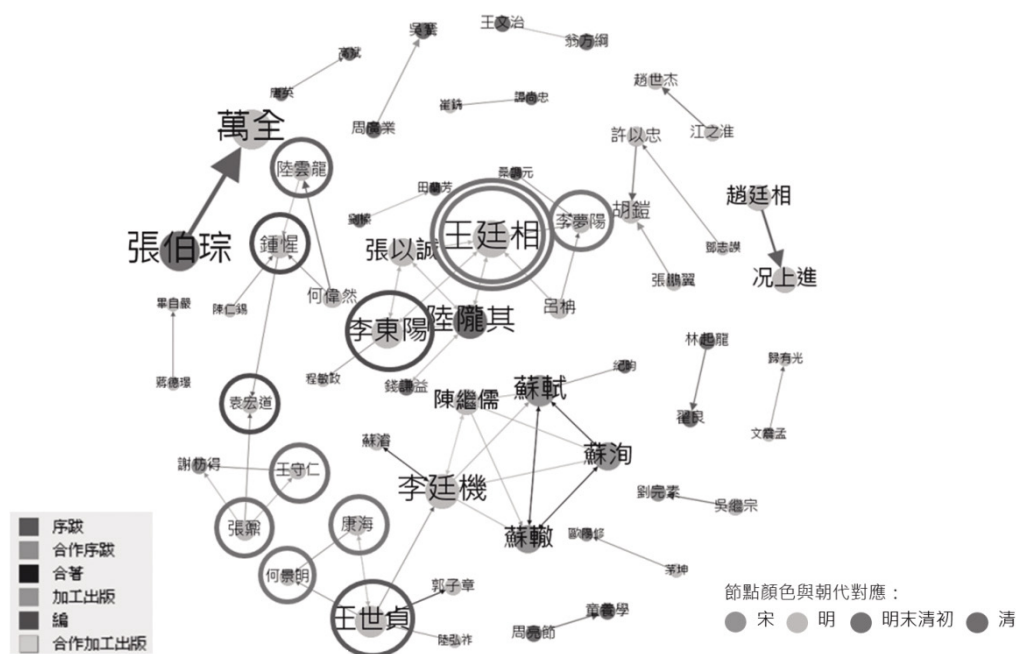


圖 8 古漢籍善本單行本貢獻者 SNA 社會網路分析圖（以範例命題「在古漢籍善本 LOD 資料集中，各單行本分別有哪些貢獻者？又，從曾共同參與同一單行本的次數來看（包含撰述、序跋題記……），貢獻人物彼此間的關聯性如何？」為例）

綜觀本研究產生之人際網脈圖，鑑因原始書目資料內容含括經、史、子、集等各種書目類型，並未特定以收錄如醫家類、方志類等主題書目為主；加以書目貢獻創作者所屬時代甚廣，從宋朝至清朝均有紀錄，復以書目樣本數較少之因，故貢獻者間合作網路社群較為分散，無法明顯揭示該書目資料集中，貢獻者間整體交織細密社群樣貌。此外，由脈絡圖中顯示由宋代三蘇連結出陳繼儒、李廷機，復連結至王世貞、何景明、康海，有別於歷來「文人集團」（如前七子／後七子）的劃分與理解。袁宏道（文學家 - 公安派）與鍾惺（文學家 - 竟陵派）之往來，可經書籍資訊，再連結到陸雲龍（小說家）、張鼐（思想家）、王守仁（思想家）。李東陽（文學家 - 茶陵派）則經由王廷相（思想家）與李夢陽（前七子）產生連結。由此結果提示出研究者留意中國文學史的刻版分類，跳脫既有的文人集團歸類印象，重新注意文人具體真實的往來關係，為文人史、文學社會學之研究，提供可能發展之新面向。

參考文獻

林聖智編 (2018)。羣碧樓藏書特展。臺北市：中央研究院歷史語言研究所。頁 198。

嚴佐之 (1989)。古籍版本學概論。上海：東華師範大學出版社。頁 10。

Kroeger, A. (2013). The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC Future. *Cataloging & classification quarterly*, 51(8), 873-890.

Malssen, K.V., Hunter, C., & Leigh, A. (2014). BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources. Available: <https://www.loc.gov/aba/pcc/bibframe/bibframe-avmodelingstudy-may15-2014.pdf>

Schreur, P. (2018) The Evolution of BIBFRAME: from MARC Surrogate to Web Conformant Data Model. Paper presented at: IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies in Session 141 - Cataloguing. Available: <http://library.ifla.org/2202/>

Sfakakis, M., Zapounidou, S., & Papatheodorou, C. (2020). Mapping Derivative Relationships from BIBFRAME 2.0 to RDA. *Cataloging & Classification Quarterly*, 58(7), 603-631.

Steele, T. D. (2019). What comes next: understanding BIBFRAME. *Library Hi Tech*. 37(3), 513-524. <https://doi.org/10.1108/LHT-06-2018-0085>

Taniguchi, S. (2018). Mapping and merging of IFLA Library Reference Model and BIBFRAME 2.0. *Cataloging & Classification Quarterly*, 56(5-6), 427-454.

Xu, A., Hess, K., & Akerman, L. (2018). From marc to bibframe 2.0: Crosswalks. *Cataloging & Classification Quarterly*, 56(2-3), 224-250.

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2014, November). Library data integration: towards BIBFRAME mapping to EDM. In *Research Conference on Metadata and Semantics Research* (pp. 262-273). Springer, Cham.

國家圖書館出版品預行編目 (CIP) 資料

第 13 次中文文獻資源共建共享合作會議：「跨域經緯・創新翻轉 - 圖書館與數位人文的對話」會議論文集：13th Conference on cooperative development and sharing of Chinese resources : conference proceedings of dialogue between libraries and digital humanities/ 國際合作組編輯. -- 臺北市：漢學研究中心，2021.07
面； 公分

部分內容為簡體字版

ISBN 978-957-678-699-0 (平裝)

1. 數位人文 2. 資源共享 3. 學術研討會 4. 文集

028.07

110010266

第 13 次中文文獻資源共建共享合作會議 跨域經緯・創新翻轉—圖書館與數位人文的對話 會議論文集

發行人 | 曾淑賢

出版者 | 漢學研究中心

100201 臺北市中山南路 20 號

02-23619132 (代表號)

編輯 | 國際合作組

國圖網址 | <https://www.ncl.edu.tw/>

設計印刷 | 邁視國際有限公司

印刷冊數 | 200 冊

出版日期 | 2021 年 7 月

ISBN 978-957-678-699-0 (平裝)

GPN 1011000891



漢學研究中心

Center for Chinese Studies

ISBN 978-957-678-699-0



9 789576 786990