



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. 「MAARKUS」，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方面的開放科學 (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一个環節，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 (Sematic 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。

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)

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/chinesecbdb/> 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/chinesecbdb/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的資料集內容，無是整個資料集，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 億個中文字的字樣，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，在瀏覽界面內便會出現「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，內容主要包含四庫全書、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，就是希望能夠為日漸增加的數位人文資源與數位人工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 Weerd) 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 Weerd(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