

# 建構整合異質性數位典藏之開放原始碼資源探索系統先導研究

## Developing an Open-Source Resource Discovery System to Integrate Heterogeneous Digital Achieves: A Pilot Study

陳勇汀 (Chen, Yung Ting)

國立政治大學圖書資訊與檔案學研究所博士生

pudding@nccu.edu.tw

### 摘要

數位典藏是保存歷史文化資產的重要技術，但是珍貴的數位典藏資源往往卻因為被分散在各個在不同計劃中各自建置的數位典藏資料庫中，而使得數位典藏的價值難以彰顯。為了促進讀者有效發掘與使用豐富的數位典藏資源，利用數位資源交換技術發展一站式整合查詢與瀏覽的資源探索系統，是相當具有發展潛力的研究議題。本研究採用開放原始碼的模式建置資源探索系統 VuFind，整合三種不同類型的資料來源，包括圖書館自動化系統 Koha、基於 DSpace 建置的臺灣百年圖書館史數位圖書館與國立政治大學機構典藏、以及使用關聯式資料庫 MySQL 建置的孫中山紀念圖書館館藏目錄等異質來源的數位典藏後設資料，以達到一站式的數位資源瀏覽與檢索服務。文中除了介紹資源探索系統的建置過程之外，也歸納整合數位典藏時所遭遇的問題，並提出未來的發展方向，最後也將所建置的資源探索系統 VuFind 以虛擬應用範本的形式公開，供人直接架設、使用。資源探索系統 VuFind 的建置除了可讓圖書館帶來更全面的一站式資源整合查詢功能，更重要的是，開放原始碼軟體的自由更讓 VuFind 有著開拓未來無限可能性的潛力。

### Abstract

Digital repository is an important technology to preserve historical cultural heritages. However, valuable digital archive resources are often separate in different digital archive databases developed by different projects, thus made it hard to promote the value of digital repository. To promote effective discovery and use of the various digital resources, utilizing interoperability technology to develop a one-stop resource

discovery system that enables integrated searching and browsing is a research issue with high potentials. VuFind, an open source resource discovery system is used to integrate Koha, Taiwan Libraries' History Digital Library and NCCU Institutional Repository which are based on DSpace, along with Sun Yat-sen Memorial Library Catalog, which is based on MySQL, to implement an one-stop browsing and searching service for digital resource. Aside from introducing the establishing process, this essay also indicates problems encountered during the develop processes. Suggestions on future development directions are also provided.

關鍵字：資源探索系統、數位典藏、VuFind、Apache Solr、Koha、DSpace

Keywords:Resource Discovery System, Digital Achieve, VuFind, Apache Solr, Koha, DSpace

## 壹、研究背景

數位典藏是保存歷史文化資產的重要技術，而發揮數位資源交換技術來整合異質性數位典藏資料庫的研究，一直是產學界不斷努力達成的目標。數位典藏與數位學習國家型科技計劃的成功為我國帶來了數位典藏的熱潮，現今各大圖書館皆不約而同地著重於名人私文書的檔案典藏。國家圖書館的「當代名人手稿典藏系統」以當今之名人作家為徵求對象，蒐集各種詩歌、散文、小說、書信、日記、劇本、評論等等珍貴手稿；臺灣學術龍頭的臺灣大學圖書館也有王文興、林文月、葉維廉與白先勇等多位臺灣大學出身的文學創作相關者的手稿、文評、書信與授課講義等珍貴檔案，並以此發展數位人文的創新研究。此外，政治大學、交通大學、中山大學等多所大學也都各自擁有獨一無二的私人文書。

私文書的典藏固然對於臺灣歷史的保存具有重要貢獻，但是分別建置的數位化館藏資源卻深藏在一座又一座數位典藏網站中，往往讓讀者如入五里霧、不知如何找起。以政治大學的情況為例，政治大學至今建置了 32 個重要的學術資料庫與數位典藏網站，其中政治大學的機構典藏更擁有亞洲排名 13 名的傲人成績。但是目前這些資源卻僅能從學術資料庫入口網站所提供的超連結進入，然後得在各別的數位典藏系統中才能進行搜尋與瀏覽。為了達到 Webster (2004)強調的一站式服務(one-stop shop)終極目標，如此豐富的數位資源究竟該要如何整合，並發展成

具備整合檢索與瀏覽等功能的整合查詢(Federated Search)系統，即是目前亟待發展的重要議題。

數位資源的整合查詢方案可以借鏡檔案領域蓬勃發展的英國。英國在建置國家檔案網路(National Archives Network，簡稱 NAN)時考量了兩種方案 (Stocking & Craven, 2004)，第一種是將檢索指令分散到各個數位典藏系統中，亦即所謂的廣播檢索(broadcast search)。當時 NAN 是採用 Z39.50 通訊協定來作跨平臺的搜尋。而我國國家發展委員會檔案管理局也用類似的做法，發展出檔案資源整合查詢平臺(Archives Cross boundaries，簡稱 ACROSS)。同樣地，圖書館界的 Metalib 與 MUSE 亦是採行此道。但是使用廣播檢索的資源整合查詢系統卻有系統回應時間過長、難以在查詢中去除重複結果(de-duplication)、排序無法依照相關性呈現(relevance ranking)以及不易視使用者需求調整檢索條件等缺點(柯皓仁, 2011; 鄭伊廷, 2011)。跟理想的整合查詢系統相比，此方案仍有相當大的改進空間。

NAN 的第二個方案則是以 ISAD(G)規範相同的後設資料欄位，將各個數位典藏的資料彙整到單一檢索系統中。由於此方案需要保存大量資料，對與伺服器設備有較高的要求、也需要較多經費才能維持運作；而且每個數位典藏都需要以客製化設定資料彙整的方式，在版本控制與定期更新等技術上也是一大門檻。種種理由之下，當時 NAN 未採行此方案。但是時至今日，科技呈現出摩爾定律(Moore's law)的急速成長，儲存成本、網路速度、運算能力等伺服器設備也日趨合理。此外，數位典藏的資料交換技術也從早期的 Z39.50、SOAP 逐漸進展到 SRU (Search/Retrieve via URL)與 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)等支援程度更完整的進階技術，可直接依循網際網路主流的 TCP/IP 架構，以 XML 標準格式進行資料交換。搜尋引擎巨人 Google 的一站式檢索服務架構，往往被圖書館視為遙不可及，但現今圖書館自行建構跨資源整合查詢系統的技術已經成熟，那就是資源探索系統(Resource Discovery System, RDS)。

因此，本研究的目的係將圖書館自動化系統之館藏書目、各別獨立的數位典藏與機構典藏、甚至是不支援圖書館資料交換標準(interoperability standards)的自行建置數位典藏內容等各種異質性數位資源進行整合，以提供使用者一站式服務的檢索與瀏覽功能。為了達到上述目的，本研究係採用開放原始碼的資源探索系統 VuFind 進行實作，此先導研究將整合了圖書館自動化系統、支援 OAI-PMH 標準的數位典藏系統，以及自行開發的數位典藏系統這三種異質性的資料來源。以下

從資源探索系統的介紹開始，接著敘述本研究建構資源探索系統之系統配置與步驟，最後歸納建置時遭遇的問題以及未來發展的方向。本研究也秉持開放原始碼的分享精神，將先導系統建置成果之資源探索系統 VuFind 的虛擬應用範本下載連結附於文中，可供其他有相同需求的圖書館直接取用、建置。

## 貳、資源探索系統 (Resource Discovery System)

資源探索系統是一種網路規模(web scale)的檢索系統，具備直覺的瀏覽與檢索介面，可供使用者找到資源後直接取用全文。Vaughan (2011)認為資源探索系統的特色在於可為各種本地端或是遠端的數位資源建置索引，這些資源包括了圖書館自動化系統的書目記錄、數位典藏、機構典藏以及自行開發與架設的資料庫。黃明居 (2011) 指出資源探索系統為次世代的圖書館整合查詢系統之趨勢，兼顧簡單、效率、彈性與強大搜尋能力等特性，深度整合實體館藏以及電子數位資源。使用者在操作容易上手的「一站式查詢與探索各類資源」並「直接取用」全文的同時，也能夠將檢索結果依據不同主題與多個面向分類(faceted classification)進行過濾，找出更為精確的檢索結果。而現今資源探索系統所提供的社群互動機制，讓使用者能在為數位資源添增社會標籤(social tag)、推薦評分(rating)以及共同協作，讓資源探索系統更呈現 Library 2.0 的功能與特性。

我國圖書館界引入資源探索系統已有數年，亦有不少資源探索系統的評估與選擇研究 (麥綺雯, 2012; 孫宇、張磊、劉煒, 2013; 殷沈琴等, 2013; 黃明居, 2011; 楊志文、郭蕙貞, 2011)。國內主要的資源探索系統有四種產品，Ex Libris 公司推出的 Primo 是最早整合圖書館自動化系統館藏資料與電子資料庫產品；而資料庫廠商也有各自的資源探索系統，如 ProQuest 公司的 Summon 與 EBSCO 公司的 Discovery Service；圖書館自動化廠商也不落人後，擁有臺灣許多圖書館採用的圖書館自動化系統 Millennium 的 Innovative Interfaces 公司，也推出了資源探索系統產品 Encore。各種資源探索系統大約在 2009 年先後推出之後，隔年各家廠商便積極引入我國。而資源探索系統的選擇上，大多數圖書館較重於提供資源探索系統的廠商與電子資料庫廠商之間的合縱連橫。資源探索系統擁有越多電子資料庫之後設資料的索引，越能獲得圖書館的青睞 (柯皓仁, 2011)。

然而，相較於關注館藏資源與電子期刊資料庫的整合，資源探索系統在數位典藏與機構典藏等特色館藏的數位資源整合能力上卻較少人著墨。其原因可能有

二：一為資源探索系統在整合其他數位資源時需要較高的技術門檻。為了與不同數位典藏系統進行資料交換，大量的伺服器連結設定與排程操作是不可避免的作業；其次，各家廠商推出的資源探索系統多未提供充足的設置文件，甚至將數位資源交換功能列入收費的額外功能，使得圖書館不得不付費聘請廠商的技術人員才能進行設置，同時也矮化了圖書館自主操作的權力。

為了促使圖書館能夠主動進行圖書館自動化系統與異質性數位資源的整合，Balnaves (2013) 提出了一套以開放原始碼軟體(Open Source Software, OSS)構成的方案。其方案結合了圖書館領域中開放原始碼的先驅——Koha 圖書館自動化系統，也整合 MIT 發展的 DSpace 機構典藏系統，並以開放原始碼軟體的資源探索系統 VuFind 來整合其他異質性的數位資源。這一套結合圖書館自動化系統與機構典藏、數位典藏之資源探索系統方案，成功地將機構中供人開放取用的資料與圖書館所擁有之實體與數位館藏等各種異質性的數位資源，帶來單一入口即可檢索與瀏覽的一站式服務。然而當時 Balnaves 也僅針對圖書館自動化系統與機構典藏進行整合，本研究欲擴大 VuFind 的整合對象，將支援 OAI-PMH 的數位典藏系統與使用關聯式資料庫 MySQL 自行建置的資料庫也納入整合的對象，發揮 VuFind 整合異質性數位典藏資料庫的能力，以擴大資源探索系統的利用價值。

### 參、開放原始碼資源探索系統 VuFind

相較於 Koha 與 DSpace 等國內圖書館界較為知名的開放原始碼軟體，同樣是開放原始碼的資源探索系統 VuFind 目前在國內仍是乏人問津。VuFind (2015)是由維拉諾瓦大學法爾維紀念圖書館(Villanova University's Falvey Memorial Library)為圖書館所發展的圖書館資源入口網站，展示介面如圖 1 所示，其目標是取代傳統的 OPAC，讓使用者可以直接對圖書館所擁有的所有資源進行檢索與瀏覽。這些資源包括了館藏書目記錄、館藏期刊、數位圖書館數位物件、機構典藏、機構書目、其他圖書館的館藏與資源。VuFind 以模組化設計，可以任意設定所需要的基本功能或是開放所有功能。最重要的是，VuFind 係以開放原始碼軟體的方式發佈，允許系統管理員依據圖書館自身的需求修改模組，或是直接新增模組並開拓圖書館的創新服務。

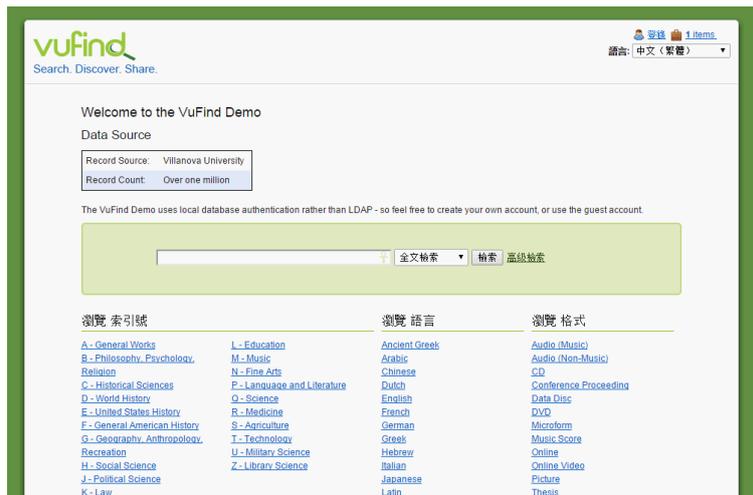


圖 1 VuFind 展示系統介面

在圖書館自動化系統的整合上，VuFind 支援來自 ExLibris 公司的 Voyager 與 Aleph、Innovative 公司的 Millennium、NewGenLib、SirsiDynix 公司的 Horizon、Symphony 以及 Koha 的 MARC 資料匯入與整合，也可以透過 SRU 資料交換協定整合 OCLC 公司所建置的 WorldCat 中的書目資料。而數位典藏與機構典藏的後設資料彙整主要則是使用 OAI-PMH 資料交換協定，能夠相容於 OCLC 的 ContentDM 與 MIT 的 DSpace 等數位典藏與機構典藏系統。然而 VuFind 不僅支援圖書館標準的資料交換協定，VuFind 使用的全文檢索引擎 Solr 更能夠整合各種關聯式資料庫的資料內容，即使是未能提供資料交換標準的自行建置之數位典藏系統，也能夠在 VuFind 中進行整合索引。

此外，VuFind 的特色還包括：

1. 搜尋以及層面(Faceted)檢索：搜尋系統可以讓讀者從基本檢索去找尋資料，再點選結果的不同層面來過濾需要的資料。
2. 使用 AJAX 查詢即時書目資訊以及位置：透過 AJAX 查詢，搜尋結果頁面可以呈現書目的即時資訊，以此可以避免頁面讀取緩慢的問題。
3. 「類似館藏」的資源建議：讀者將會看到類似現在所檢視資源的其他類似館藏。
4. 書目記錄備忘錄與匯出功能：讀者可以從搜尋結果或是單一書目的檢視頁面匯出書目清單。這份清單會保存在系統中，以便在需要時匯出。這功能可讓讀者省去學習使用複雜的書目管理軟體的困擾。
5. 瀏覽資源：讀者可以瀏覽所有圖書館的資源，掌握圖書館資源的整體脈絡，而

非僅能觀看搜尋結果。

6. 作者書目：讀者可以了解更多關於作者的脈絡資訊，並找到圖書館中所有作者發表的書籍。
7. 固定網址：每個書目頁面都有固定網址，因此讀者可以用瀏覽器的書籤保存他們的查詢結果或是書目資訊。
8. 相容於 Zotero：讀者可以使用書目管理軟體 Zotero 或任何基於網頁內嵌引用資訊標準 CoinS (OpenURL ContextObject in SPAN)的應用程式直接從網頁中儲存書目記錄，以便集中管理引用文獻。
9. 國際化翻譯：VuFind 已經被翻譯為正體中文、簡體中文、巴西葡萄牙語、荷蘭語、法語、英語、德語、日語、西班牙語等 25 種語言。而且可以輕易地新增其他翻譯與用詞。
10. 支援開放搜尋(OpenSearch)：使用者可以將 VuFind 的搜尋功能加入瀏覽器的搜尋引擎清單中，可直接在瀏覽器上檢索 VuFind 的數位資源。

#### 肆、資源探索系統之先導系統建置

由於 VuFind 具備了資源探索系統應有的完整功能，本研究係決定基於 Balnaves (2013)所提出之開放原始碼整合方案建置小規模的資源探索系統先導研究。為了含括圖書館可能面臨之各種使用情境，本研究規劃將圖書館自動化系統 Koha 與數位典藏兼機構典藏系統 DSpace 整合至 VuFind 系統中，除此之外，更將自行建置之數位典藏系統也納入 VuFind 的索引，藉以有效解決不具備圖書館資料交換標準之數位典藏系統的整合問題。

##### 一、系統架構

本研究所建置的資源探索系統架構如圖 2 所示。先導系統之建置係以資源探索系統 VuFind 為中心，整合其他各自獨立的數位資料來源。為了模擬實際可能的運作情境，本研究採用 Proxmox Virtual Environment 進行伺服器虛擬化來建置各個系統，操作介面如圖 3 所示。一部分的數位資源則是直接整合上線運作中的數位典藏與機構典藏系統。Proxmox Virtual Environment 同樣也是開放原始碼軟體，可從以下網址取用：<https://pve.proxmox.com/>。以下針對各個系統的建置與介紹進行更詳細的說明：

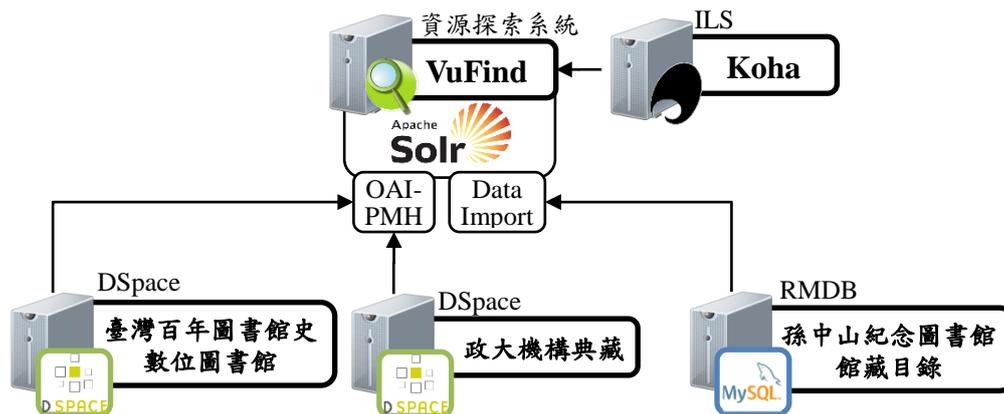


圖 2 資源探索系統之先導系統建置

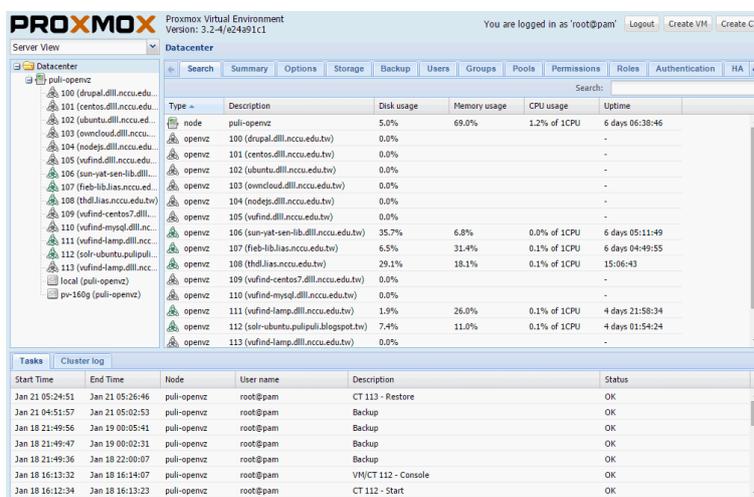


圖 3 Proxmox Virtual Environment 操作畫面

## 二、資源探索系統 VuFind

本研究所建置之 VuFind 伺服器係一 OpenVZ 容器虛擬化技術建置之虛擬機器，VuFind 則是安裝目前最新版本 VuFind 2。軟體與安裝則依照其網站上的指示進行。VuFind 網站的網址為：<http://vufind-org.github.io/vufind/>。為了配合 VuFind 的安裝需求，本研究採用了 TurnKey Linux 的 LAMP Stack 虛擬應用範本(Virtual Application Template)為基礎來進行安裝與配置。Turnkey Linux 以開放原始碼軟體的方式發佈了許多內建可直接使用的虛擬應用範本，供人直接在虛擬化環境中直接建置一臺可立即使用的虛擬機器，網址為：<http://www.turnkeylinux.org>。而其中 LAMP Stack 虛擬應用範本的作業系統為 Debian 7，提供 PHP 運作環境與 MySQL 資料庫。硬體配置為單核心 CPU，記憶體與記憶體置換空間(SWAP)各為 1GB，硬碟空間則是設為 80GB，其硬體要求並不高。

### 三、異質性資料來源系統介紹

本研究欲在 VuFind 中整合三種不同的資料來源，各別是圖書館自動化系統 Koha、支援 OAI-PMH 的 DSpace 系統、以及自行建置不支援常見資料交換標準的關聯式資料庫 MySQL。以下逐一介紹：

#### (一) 圖書館自動化系統 Koha

本研究中欲整合的圖書館自動化系統為開放原始碼軟體的 Koha。作為一套企業規模水準的圖書館自動化系統，Koha 具備採訪、流通、編目、期刊管理、權威控制、可自訂的報表、標籤列印、多格式的標示、離線流通功能等多種功能模組 (Official Website of Koha Library Software, 2015)。除了從 Koha 官方網站下載套件進行安裝之外，陳勇汀 (2014) 也提供 Koha 3.16 的 OpenVZ 虛擬應用範本供人直接取用。本研究採用後者的方式建置 Koha 系統，並匯入數筆書目資料，作為整合測試使用。最終建置的 Koha 書目記錄頁面如圖 4 所示。

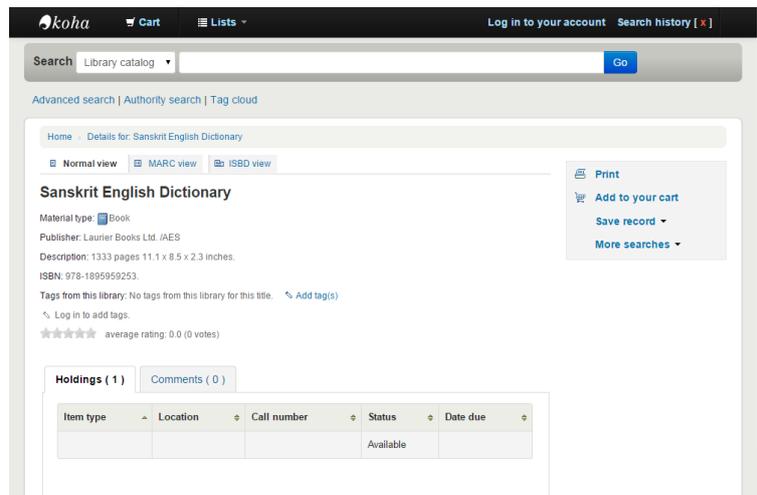


圖 4 Koha 的館藏書目記錄檢視畫面

#### (二) 數位典藏與機構典藏系統 DSpace

在數位典藏與機構典藏等開放近用數位資源的整合上，本研究選擇支援 OAI-PMH 資料交換標準的 DSpace。DSpace 係由麻省理工學院 (Massachusetts Institute of Technology, MIT) 所開發的開放原始碼機構典藏系統，不僅可以用來保存文字、影像、影片、聲音等各種形式的資料，並為其建置索引供使用者檢索之外，

更重要的是 DSpace 支援各種圖書館與檔案館採用的資料交換標準，包含 OAI-PMH，可有效提高數位資源的能見度 (DURASPACE, 2015)。為了使先導研究貼近實際架設時的可能情境，本研究規劃整合兩種以 DSpace 架設的數位典藏系統與機構典藏系統到 VuFind 之中，前者是 2008 年建置的臺灣百年圖書館史數位圖書館，後者則是擁有豐富全文資源的國立政治大學機構典藏。

臺灣百年圖書館史數位圖書館係採用修改後的 DSpace 數位圖書館系統，將數位典藏的資料建入系統平臺中，利用其分類與檢索的功能來提供服務。網站內容架構分為六大類，包括：(1) 大事記；(2) 各類圖書館；(3) 圖書資訊學教育；(4) 圖書館人物；(5) 圖書館建築；(6) 各類文件。系統畫面如圖 5 所示，公開取用的網址為：<http://tlh.lias.nccu.edu.tw/>（王梅玲等人，2008）。



圖 5 臺灣百年圖書館史數位圖書館首頁

政治大學機構典藏則是由政治大學圖書館所推動，秉持著維護學術傳播體系重要的典藏與提供者之精神，將政治大學機構內各項豐富學術產出進行永久性的典藏。並透過圖書館與校內各學院系所、行政單位緊密搭配的合作模式，讓政治大學機構典藏永續經營。除了增進機構內學術產出的取用更加便利外，機構典藏更能夠改善國內的學術傳播環境，並與世界各地學術研究者接軌共同彼此分享豐沛的知識資產。政治大學機構典藏的首頁如圖 6 所示，公開取用的網址為：<http://nccur.lib.nccu.edu.tw/>（國立政治大學圖書館，2015）。



圖 6 政大機構典藏首頁

### (三) 關聯式資料庫 MySQL

許多學術資料庫與數位典藏係以自行規劃的方式建置，其做法大多係以關聯式資料庫(Relational Database Management, RMDB)技術作為保存資料的核心。但是自行建置的資料庫大多並未考量資料交換標準的相關功能，因此一般的資源探索系統往往無法整合這些自行建置的資料庫。然而，VuFind 因為採用了 Solr 全文檢索引擎為核心，使它具備了能夠與關聯式資料庫進行整合的能力，也使得本研究可將自行建置的孫中山紀念圖書館館藏目錄有效納入整合資源探索平臺之中。

孫中山紀念圖書館的歷史可追溯到民國 16 年，為紀念國父孫中山先生而在南京成立之紀念圖書館。在歷經長久的遷臺、復館與搬遷之後，2012 年移至國立政治大學社會科學資料中心進行管理與活化。孫中山紀念圖書館館藏資料包括中國國民黨中央委員會藏書、孫科先生「補不足齋」藏書、故臺大哲學系主任方東美先生贈書、中山文化教育館藏書、國防研究院藏書、孫中山先生言行思想、國民黨黨史資料及部分明清善本、新善本書籍等珍貴史料，其館藏資源可說是近代歷史研究的重要瑰寶（社會科學資料中心，2013）。然而，孫中山紀念圖書館的館藏目錄目前僅有以 MySQL 關聯式資料庫保存的資料，本研究欲藉此機會將其豐富的館藏資源整合進資源探索系統 VuFind，供讀者方便查檢。

### 伍、系統建置與整合過程

VuFind 係以開放原始碼軟體的方式發佈，其最大的優勢就是可讓系統管理者

掌握系統的一切設定。除了 VuFind 提供的說明文件之外，VuFind 在 SourceForge 的電子郵件討論群組中也有許多開發者社群，透過互相協助的方式解決眾人在安裝與使用上的問題。以下本研究將概述資源探索系統先導研究的建置與整合過程，並在文中註明需要注意的重點。

## 一、資源探索系統 VuFind 的建置

本研究所建置之 VuFind 係基於 Turnkey Linux 的虛擬應用範本 LAMP Stack，並參考 VuFind 的 Ubuntu 安裝教學進行建置，教學文件網址為 [https://vufind.org/wiki/vufind2:installation\\_ubuntu](https://vufind.org/wiki/vufind2:installation_ubuntu)。LAMP Stack 的環境除了 VuFind 所需要的 PHP 與 MySQL 之外，仍需要安裝 Java 運作環境以供全文檢索引擎 Solr 正常運作。本研究的 VuFind 選擇使用 Debian 作業系統下進行安裝，大部分的安裝動作都可以使用套件軟體管理工具 APT (Advanced Packaging Tools) 直接從網際網路下載、並直接安裝相關套件。唯需注意的是，VuFind 的 Ubuntu 安裝教學雖與 Debian 類似，但若以系統管理者 root 登入，就不需要額外的 sudo 指令，直接操作 apt-get 等安裝指令即可。

在經過簡單的安裝步驟之後，VuFind 預設將安裝在伺服器的 /usr/local/vufind2/ 底下。為了方便接下來的說明，本文中將以 [VuFind] 表示 VuFind 的所在目錄，而以 [VuFind-server] 表示 VuFind 伺服器的網址。舉例來說，要啟動 VuFind 需在伺服器指令列中執行「[VuFind]/vufind.sh start」；重新啟動的指令則是「[VuFind]/vufind.sh restart」。而最後建置完成之後則可直接從 [http://\[VuFind-server\]/vufind/](http://[VuFind-server]/vufind/) 連線使用。

## 二、VuFind 與異質性資料來源的整合

VuFind 在整合圖書館自動化系統 Koha、支援 OPI-PMH 標準的 DSpace、以及自行建置的關聯式資料庫 MySQL 等不同類型的資料來源有各自的做法，以下逐一介紹不同類型資料來源的整合設定概要：

### (一) VuFind 與圖書館自動化系統 Koha 的整合

在 VuFind 安裝的最後步驟為自動設定功能(Auto-Configuration)。在此步驟中需要設定圖書館自動化系統的相關參數，此處即可選用 Koha，並輸入可登入 Koha 之 MySQL 資料庫的連線設定。

然而 MySQL 的帳號一般並不會開放本機之外的主機範圍使用，因此需要額外  
在 Koha 中新增特別供 VuFind 使用的帳號。本處設定建議增設 root 之外的其他帳  
號名稱，例如 vufind。帳號的連線「主機」也必須指明 VuFind 伺服器的網址，而  
非預設的 localhost 本機端。最後還要將 Koha 的資料庫權限開放給新增的帳號，如  
此 VuFind 才能如預期一樣取用 Koha 內建置的館藏書目資料。若自動設定功能中  
圖書館自動化系統成功設置之後，VuFind 即會自動將 Koha 的資料加入索引。重新  
啟動 VuFind 之後即可看到來自 Koha 圖書館自動化系統的館藏記錄。

除了 Koha 之外，VuFind 也可以整合來自 Voyager、Aleph、Millennium、  
NewGenLib、Horizon 與 Sysmphony 等各種圖書館自動化系統的館藏資料。操作方  
式皆於 VuFind 網頁中說明：[https://vufind.org/wiki/marc\\_export\\_notes](https://vufind.org/wiki/marc_export_notes)。本研究未  
來將積極爭取與其他系統合作的機會，進一步拓展 VuFind 資源整合的對象。

## (二) VuFind 與數位典藏、機構典藏系統 DSpace 的整合

VuFind 係使用 OAI-PMH 資料交換標準來獲取以 DSpace 建置的數位典藏與機  
構典藏中每一筆典藏的後設資料，因此必須事先確認 DSpace 系統提供 OAI 資訊的  
網址來源。臺灣百年圖書館史數位圖書館的 OAI 連線網址為  
<http://tlh.lias.nccu.edu.tw/oai/request>；而政治大學機構典藏係基於機構典藏計劃所  
建置，根據計劃中的安裝常見問題說明，即可找到機構典藏 OAI 網址 (NturAdmin，  
2013)，其 OAI 連線網址則是 <http://nccur.lib.nccu.edu.tw/ir-oai/request>。

VuFind 提供了一份 DSpace 整合的說明，全文可從此網址取得  
[https://vufind.org/wiki/importing\\_records:how\\_to\\_index\\_dspace\\_with\\_vufind](https://vufind.org/wiki/importing_records:how_to_index_dspace_with_vufind)。其設定  
流程大致上可分成五個步驟：

1. 設定 OAI 提供的網址，設定檔案的路徑為「[VuFind]/harvest/oai.ini」；
2. 設定後設資料的對應，設定檔案有兩個，各別是  
「[VuFind]/import/dspace.properties」與「[VuFind]/import/xsl/dspace.xml」；
3. 執行獲取指令，以下載 DSpace 所典藏數位資源之 XML 檔案。獲取指令為「php  
[VuFind]/harvest/harvest\_oai.php」；
4. 執行匯入指令，將 XML 檔案的後設資料匯入 VuFind。匯入指令必須先切換到  
「[VuFind]/harvest」目錄底下，再執行「sh

batch-import-xsl.sh ./DSpace ../import/dspace.properties」，其中「./DSpace」與「../import/dspace.properties」皆與後設資料對應檔案相關，若匯入時發生錯誤，則需檢查第二步驟中的兩個檔案是否正確設定。

5. 最後需要重新啟動 VuFind，即可看到由 DSpace 匯入的後設資料。

使用 OAI-PMH 獲取 DSpace 後設資料的做法得採用非同步排程的方式進行。從上述第三步驟之後的指令可設定為排程檔案，設定間隔固定時間進行獲取、匯入、重新啟動 VuFind 即可。

### (三) 關聯式資料庫孫中山紀念圖書館館藏目錄的整合

本研究最後要整合的對象係以關聯式資料庫 MySQL 所建置的孫中山紀念圖書館館藏目錄。VuFind 的系統本身預設不具備關聯式資料庫的整合功能，但是它所使用的核心卻是以能夠索引各種資料而聞名的全文檢索引擎 Solr。因此本研究的關聯式資料庫整合即是仰賴 Solr 的 DataImport 功能來實作。

為了使 VuFind 能夠索引關聯式資料庫，首先必須先在關聯式資料庫中為 VuFind 新增帳號與設定可連線的主機，此做法與上述將 VuFind 與 Koha 的 MySQL 進行整合的方式相同。以下用[MySQL-server]代表該關聯式資料庫的網路位址，其登入的帳號設為 vufind，密碼為 password。

接著需要設定 VuFind 系統內多個檔案，其步驟如下：

1. 為 Solr 加入必備的函式庫(library)，這包括 DataImport 功能需要的 solr-dataimporthandler-x.jar (x 為版本號，例如 solr-dataimporthandler-4.2.1.jar)、solr-dataimporthandler-extras-x.jar，還有連線到 MySQL 資料庫所需要的 mariadb-java-client-x.jar，必須先將這些 jar 函式庫放置到 [VuFind]/solr/lib 路徑底下。
2. 則是設定 VuFind 儲存書目資料的 Solr 核心(core) biblio，將其後設資料設定檔 [VuFind]/solr/biblio/conf/solrconfig.xml 加入程式 1 設定：

```
<requestHandler name="/dataimport"  
class="org.apache.solr.handler.dataimport.DataImportHandler">  
  <lst name="defaults">  
    <str name="config">dataimport-config-mysql.xml</str>  
  </lst>
```

```
</requestHandler>
```

### 程式 1. solrconfig.xml 中的 DataImport 設定

3. 建立[VuFind]/solr/biblio/conf/dataimport-config-mysql.xml，並依據程式 2 的內容進行設置。其中[MySQL-server]、user 與 password 的設定值需依據 MySQL 伺服器各別調整之外，還需要在<entity>的 query 屬性中設定進行檢索的 SQL 語法，例如「select \* from books」；下面還要搭配檢索結果與 VuFind 的欄位對應，例如「<field column="id" name="id" />」。VuFind 提供的預設欄位在 [VuFind]/solr/biblio/conf/schema.xml 中有詳細設定，常用的欄位可參考表 1 的摘要。

```
<dataConfig>
  <dataSource driver="org.mariadb.jdbc.Driver"
url="jdbc:mysql://[MySQL-server]:3306/lib" user="vufind"
password="password" batchSize="5" />
  <document name="actions">
    <entity name="actions_item" query="select * from books">
      <field column="id" name="id" />
      <field column="bookname" name="title" />
      <field column="author" name="author" />
      <field column="publisher" name="publisher" />
      <field column="description" name="description" />
    </entity>
  </document>
</dataConfig>
```

### 程式 2 dataimport-config-mysql.xml 設定檔案

表 1 VuFind 常用欄位名稱與說明

欄位名稱	資料格式	說明	欄位名稱	資料格式	說明
id	string	辨識碼	callnumber	string	索書號
title	text	標題	description	text	描述
author	textProper	作者	contents	text	目錄
publishDate	string	出版日	language	string	語言
url	string	網址	format	string	格式
allfields	text	後設資料所有欄位	physical	string	實體描述
fulltext	text	全文內容	publisher	textProper	出版者
instuion	string	典藏機構	edition	string	版本

collection	string	典藏系列	series	text	系列
building	string	實體所在位置	topic	text	主題
dewey-full	string	杜威十進分類法	genre	text	類別
lccn	string	美國國會圖書館分類法	geographic	text	地理位置
isbn	isn	ISBN 號碼	era	text	時代
issn	isn	ISSN 號碼	thumbnail	string	縮圖網址

4. 到網址 [http://\[VuFind\]:8080/solr/#/biblio/dataimport](http://[VuFind]:8080/solr/#/biblio/dataimport) 中開啟 DataImport 頁面，執行匯入的動作。

若 dataimport-config-mysql.xml 的設定正確，而且 MySQL 的帳號也能夠順利連線，Solr 就能很迅速地索引 MySQL 資料庫中的內容。本研究索引 MySQL 資料庫中的 1 萬 3 千多筆資料，花費時間大約僅需 10 分鐘。最後重新啟動 VuFind 之後，即可看到來自 MySQL 資料庫的後設資料內容。若設定運作能夠正常運作，則可以將 DataImport 的匯入與重新啟動 VuFind 指令設為排程檔案，即可定期與數位典藏系統進行整合。

### 三、整合成果

在整合了圖書館自動化系統 Koha、數位典藏與機構典藏的 DSpace、以及自行建置的 MySQL 關聯式資料庫之後，最後 VuFind 即可直接針對匯入的後設資料直接進行檢索，檢索畫面如圖 7 所示。在 Stocking 與 Craven (2004) 所遇到資料整合之間的版本控制問題上，VuFind 的全文檢索引擎 Solr 係採物件導向資料庫概念，資料的匯入會依據 id 欄位來判斷是否為新增或更新，因此不會出現重複匯入之後資料重複的問題。而本研究所建置完成的開放原始碼資源探索 VuFind 也依照 GPL2 的授權規範，將建置成果在 GitHub 上發佈，網址為：<https://github.com/pulipulichen/vufind/>。

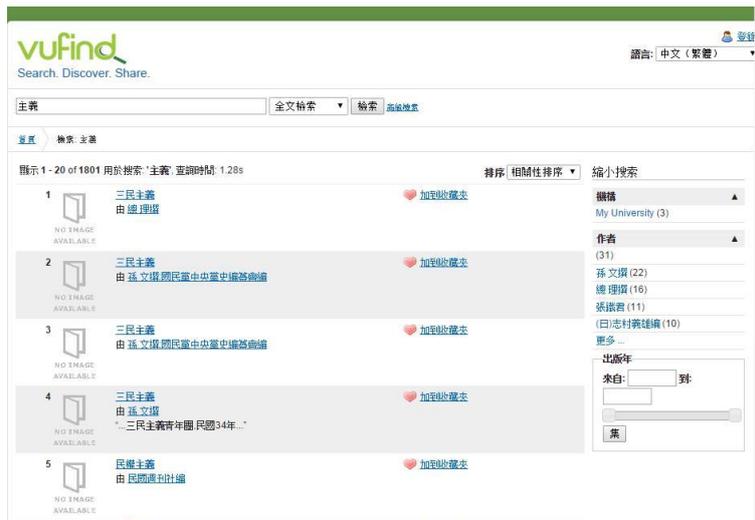


圖 7 VuFind 的檢索畫面

## 陸、討論

在先導系統的建置過程中，除了安裝系統時需要注意的設定與運作環境配置之外，目前仍有部分問題尚未克服，導致資源探索系統索引的後設資料未能完整呈現。以下說明本研究所遭遇的問題以及可行的解決方向。

### 一、DSpace 的 OAI 提供資訊問題

本研究將臺灣百年圖書館史數位圖書館與政大機構典藏作為資源探索系統的資料來源，它們的基礎系統 DSpace 皆具備 OAI 提供的功能，使得 VuFind 得以輕易獲取它們的後設資料並加入索引。但是這兩個基於 DSpace 所提供的後設資料卻都有各自的問題，必須額外設定或將後設資料進行清理之後才能使用。

臺灣百年圖書館史數位圖書館為其數位典藏資料制訂了團體名稱、人物、建築物、書目、照片與檔案六種不同的後設資料綱要(metadata schema)。其做法是在邏輯上拓展了都柏林核心集(Dublin Core, DC)並建置自訂的後設資料欄位，但是由於並未使用預設的 DC 欄位，OAI 在預設狀態下無法匯出這些自訂欄位，仍需要手動設定額外的後設資料欄位對應。

而基於機構典藏計劃的政大機構典藏則是遵循預設的 DC 後設資料綱要，但是在後設資料的內容中卻加入了自訂的額外資料，使得資料類型並非完全遵照 OAI 的規範。以程式 3 為例，<dc:identifier>應為指示資源位址的網址，但是政大機構典藏匯出的 OAI 中卻多了[[uri]]的額外資訊，導致資料不符合網址格式，使得 VuFind 無法直接匯入<dc:identifier>的資料內容。不過只要改善後設資料對應的轉

譯檔案[VuFind]/import/xsl/dspace.xsl，讓 VuFind 剖析 XML 檔案時過濾多餘的[[uri]]字串，即可克服這個問題。

```
<dc:identifier>[[uri]]http://nccur.lib.nccu.edu.tw/handle/140.119/4579</dc:identifier>
```

### 程式 3 OAI 提供的 XML 之 identifier 欄位舉例

## 二、 關聯式資料庫連線的安全性問題

本研究將圖書館自動化系統 Koha 以及自行建置的關聯式資料庫 MySQL 納入資源探索系統中，以資料庫直接連線的方式獲取後設資料內容。其中提供資料的關聯式資料庫中需額外建置帳號、設定可連線的伺服器（亦即資源探索系統的網路位址）、以及資料庫操作權限。由於此處開放了關聯式資料庫供本機之外的伺服器存取，因此必須格外注意資訊安全的控管。除了資源探索系統存取的帳號切忌千萬不要開放系統管理者 root 等級的權限之外，更要確實規範可連線使用的伺服器主機。

另一方面，由於關聯式資料庫一般較少開放給外部伺服器使用，因此網路防火牆或是入侵偵測防禦系統(intrusion detection system, IDS)很有可能會阻擋資料庫的外部連線，例如 MySQL 預設使用的 3306 連結埠。若 VuFind 整合的過程中發現一直無法順利連線到 MySQL 關聯式資料庫，不妨先檢查防火牆與入侵偵測防禦系統的設定，確保 VuFind 被列在可通過的白名單(whitelist)之中。

## 三、 資源探索系統 VuFind 仍有待發掘

本研究的先導系統建置中將異質來源的數位資源整合到資源探索系統 VuFind 當中，但是到目前為止僅是建置起 VuFind 作為資源探索系統的基本功能，確認使其檢索功能可以正常運作，未來仍有進一步探索 VuFind 的進階設定。

除了檢索功能之外，VuFind 也具備瀏覽目錄的功能。如圖 8 所示，使用者可以根據標籤、索引號、作者、主題、流派、地區、時代等七種不同層面的後設資料進行瀏覽。但是 VuFind 尚未考量到英文之外的館藏，故無法根據中文順序進行瀏覽。此外，後設資料在匯入 VuFind 時，也需要考量多值、縮寫、「著」與「譯」等作者類型的區分等情況。為了使不同縮寫的作者、主題能夠進行有效的彙整，VuFind 與 Koha 整合之權威控制將會是改善瀏覽目錄的重要功能。除了書目資料有

待整理之外，將使用介面進行客製化調整，使其更符合中文情境的使用邏輯，也是要將 VuFind 用於正式線上服務之前必須要克服的問題。



圖 8 VuFind 的瀏覽目錄功能

## 柒、結論

本研究採用開放原始碼的模式建置資源探索系統 VuFind，並成功地整合了來自圖書館自動化系統 Koha、基於 DSpace 建置的臺灣百年圖書館史數位圖書館與國立政治大學機構典藏、以及使用關聯式資料庫 MySQL 建置的孫中山紀念圖書館館藏目錄等異質來源的後設資料，達到一站式的數位資源瀏覽與檢索服務。VuFind 的建置並不僅是讓圖書館入口網站有了傳統 OPAC 之外的另一種選擇，更重要的是，開放原始碼軟體的自由讓它具備開拓未來無限可能性的潛力。

多重脈絡的呈現是資源探索系統 VuFind 未來值得發展的方向之一。項潔、翁稷安（2012）強調檔案系統應呈現的多重脈絡，在彙整館藏書目資料、機構典藏的計劃與論文、以及數位典藏的手稿等私文書檔案的豐富巨量資料(Big Data)之後，VuFind 已經具備了發展跨資料庫數位資源脈絡呈現的先備條件，可以進一步朝向數位人文研究發展。以檔案私文書來說，私文書的編排描述著重於創作者的個人脈絡，應以創作者的生平經歷作為彙整相關作品的來源原則。透過事件、時間與空間等多重面向，串聯起不同獨立數位資源之間的時空脈絡。這種資源探索系統才能呈現的多重脈絡整合成果，是傳統 OPAC 與各自運作的數位典藏之中難以達成的境界。

作為一個先導研究，本研究主要僅針對 Koha 圖書館自動化系統、使用

OAI-PMH 標準的 DSpace 機構典藏與數位典藏系統、以及以 MySQL 資料庫自行建置的資料庫這三種不同類型的資料來源進行整合。未來有待與更多其他資料來源合作，以進行國內較為知名的圖書館自動化系統與 WorldCat 的整合。

除了數位資源的編排與描述之外，在注重學術影響力呈現的國際化衝擊之下，網路計量新指標 Altmetrics 也將是提升資源探索系統 VuFind 應用價值的發展重點。有別於傳統評鑑學術成果的引用計量分析，Altmetrics 係基於文件層級計量 (Article-Level Metrics, ALMs) 的概念、在網際網路環境下針對新型學術傳播模式進行的新型計量方法。其計量的分類大致上包括了使用率(usage)、擷取(capture)、提及與討論(mention)、社群媒體(social media)、引用(citation)等五種計算方式 (Cave, 2013)。根據蔡明月、曾苓莉 (2014) 所彙整的 Altmetrics 相關服務，只要資料庫系統提供必備資料或整合服務所需要的 API，就能夠在直接在系統上整合 Altmetrics 的計量資訊。舉例來說，題名與作者的資訊可使用 Publish or Perish (PoP) 計算來自 Google Scholar 的引用數據；數位物件辨識碼 (Digital Object Identifier, DOI) 與 PubMed ID 有助於更精確地連結論文與引用資訊，前者能夠支援 Altmetric.com 服務，後者則可結合 CiteIn，而 ScienceCard 服務則能夠同時支援 DOI 與 PubMed ID；若系統整合 Mendeley API 之後，也可以直接使用來自 PaperCritic 與 ReaderMeter 所提供的計量資訊。

然而作為開放原始碼資源探索系統的方案之一，VuFind 也應該與其他資源探索系統一樣進行多方面的多重評估。評估方法可以依循 Luther 與 Kelly(2011)所提出的架構，從內容、檢索、需求符合程度與價格等四個面向規劃的評估準則；或是參考香港教育學院圖書館的做法，從學生、教職員、圖書館員等不同族群評估其館藏目錄整合程度、文獻索引以及介面設計等面向 (麥綺雯, 2012)；或是根據殷沈琴等人 (2013) 歸納三種不同資源探索系統評估的研究，從檢索功能和使用介面功能評估、讀者評估、後設資料評估、資源整合產品功能評估、後續系統可擴展性與可維護性等面向來規劃資源探索系統的評估架構。

優使性(usability)也是從使用者的角度評估資源探索系統的重要議題。建議未來應該融合 Nielsen (1993)所提出的優使性五大構面及十項評估準則，安排使用者針對記憶性、學習性、容錯性、效率性與滿意度等五個面向進行使用評估。姜義臺 (2010)、林佳穎與吳明德 (2011) 以及鄭伊庭 (2011) 等人的研究都是針對數位資源整合查詢目錄進行優使性評估，相當值得參考。

一站式服務是圖書館的最終目標，但是這目標並非一蹴可及。開放原始碼的資源探索系統 VuFind 儘管不具備廠商強調的電子資料庫預索引資源，但是卻有更多供人探索與開發的無限可能性。所幸的是，VuFind 的 PHP 與 MySQL 運作環境是國內大多工程師都具備的專業技能，因此可以降低工程師開發 VuFind 的門檻。最後，本研究也秉持開放原始碼的分享精神，不僅開放目前先導系統所架設的 VuFind 虛擬應用範本功能下載，也歡迎欲建置資源探索系統的同道一同討論、合作，以及進一步分享開發的成果。

### 誌謝

感謝國立政治大學圖書館與社會科學資料中心協助提供本研究所需的數位資源。

### 參考文獻

- NturAdmin (2013)。NTURInstallFAQ。機構典藏計劃網站。上網日期：2015 年 1 月 20 日，檢自：<http://ir.org.tw/Wiki.jsp?page=NTURInstallFAQ>
- 王梅玲、蔡明月、陳志銘、柯雲娥、蔡佳縈、陳勇汀、林怡甄 (2008)。台灣圖書館史數位圖書館建構之研究。《圖書館學與資訊科學》，34 (1)，15-38。檢自：<http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/view/504>
- 林佳穎、吳明德 (2011)。圖書館電子資源整合查詢系統之好用性評估：以國立臺灣大學圖書館 MUSE 電子資源整合查詢系統為例。《大學圖書館》，15 (2)，1-18。上網日期：2015 年 1 月 20 日，檢自：<http://www.lib.ntu.edu.tw/Publication/univj/uj15-2/01.pdf>
- 社會科學資料中心 (2013 年 5 月 18 日)。重現人文歷史風 孫中山紀念圖書館政大揭幕。《國立政治大學 - 校園新聞》。臺北市。
- 姜義臺 (2010)。圖書館電子資源整合檢索系統優使性之研究：以靜宜大學為例。《圖書與資訊學刊》，75，25-36。上網日期：2015 年 1 月 20 日，檢自：[http://www.lib.nccu.edu.tw/blis/fulltext/75/75\\_2.pdf](http://www.lib.nccu.edu.tw/blis/fulltext/75/75_2.pdf)
- 柯皓仁 (2011)。圖書館自動化與數位化。在《中華民國圖書館年鑑 100 年》。臺北市：國家圖書館。
- 殷沈琴、唐武京、邵誠敏、謝瑩、王乐、张计龙 (2013)。三家资源发现系统的调研，测试和评估。《图书馆杂志》，32 (12)，82-86。上網日期：2015 年 1 月 8 日，檢自：

- <http://www.libnet.sh.cn:8081/tsgzz/CN/article/downloadArticleFile.do?attachType=PDF&id=320>
- 國立政治大學圖書館 (2015)。關於政大典藏。政大機構典藏。上網日期：2015年1月21日，檢自：<http://nccur.lib.nccu.edu.tw/about/index.jsp>
- 陳勇汀 (2014)。OpenVZ 虛擬應用範本：Koha。布丁布丁吃什麼？。上網日期：2015年1月21日，檢自：  
<http://pulipuli.blogspot.tw/2014/10/openvzkoha-openvz-virtual-application.html>
- 麥綺雯 (2012)。如何挑選合適的探索工具 (Discovery Tool)——香港教育學院圖書館的經驗分享。2012年第十一屆海峽兩岸圖書資訊學學術研討會論文集。2012年第十一屆海峽兩岸圖書資訊學學術研討會。
- 項潔、翁稷安 (2012)。多重脈絡——數位檔案之問題與挑戰。在數位人文要義：尋找類型與軌跡。臺北市：國立臺灣大學出版中心。
- 黃明居 (2011)。次世代圖書館館藏整合查詢系統。在中華民國圖書館年鑑 100年。臺北市：國家圖書館。
- 楊志文、郭蕙貞 (2011)。E 世代的館藏查詢系統——資源探索系統 (DDS)。上網日期：2015年1月20日，檢自：  
<http://nchuir.lib.nchu.edu.tw/handle/309270000/147380>
- 蔡明月、曾苓莉 (2014)。網路計量學新指標 Altmetrics。教育資料與圖書館學，51 (5)，91-120。doi：10.6120/JoEMLS.2014.51S/0655.OR.AM
- 鄭伊廷 (2011)。檔案目錄整合檢索系統之優使性研究——以 ACROSS 為例 (碩士論文)。國立政治大學圖書資訊與檔案學研究所，國立政治大學。上網日期：2015年1月20日，檢自：  
<http://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=6EYFGH/record?r1=1&h1=0>
- 孙宇、张磊、刘炜 (2013)。图书馆资源发现系统选型研究。图书馆杂志，32 (12)，63-70。上網日期：2015年1月8日，檢自：  
<http://www.libnet.sh.cn:82/gate/big5/www.libnet.sh.cn:8081/tsgzz/CN/abstract/abstract317.shtml>
- Balnaves, E. (2013). From OPAC to Archive: integrated discovery and digital libraries with open source. Retrieved from <http://library.ifla.org/id/eprint/79>
- Cave, R. (2013). Overview of the Altmetrics Landscape. *Charleston Library*

- Conference*. Retrieved from  
<http://docs.lib.purdue.edu/charleston/2012/Communication/2>
- DURASPACE. (2015). About DSpace. *DSpace*. Retrieved January 21, 2015, from  
<http://www.dspace.org/introducing>
- Luther, J., & Kelly, M. C. (2011, March 15). The Next Generation of Discovery. *Library Journal*, p. 66. Retrieved from  
<http://www.lexisnexis.com/lxacui2api/api/version1/getDocCui?lni=52BF-J0N1-DY5K-R1N3&csi=242161&hl=t&hv=t&hnsd=f&hns=t&hgn=t&oc=00240&perma=true>
- Nielsen, J. (1993). What is usability. *Usability Engineering*, 23–48.
- Official Website of Koha Library Software. (2015). About. *Koha*. Retrieved January 21, 2015, from <http://koha-community.org/about/>
- Stocking, B., & Craven, L. (2004). Chapter 6. Metadata and the UK archives network. In G. E. Gorman & D. G. Dorner (Eds.), *Metadata applications and management* (pp. 109-138). Lanham, Md.: Scarecrow Press.
- Vaughan, J. (2011). Investigations into library web scale discovery services. *Information Technology and Libraries*. Retrieved from  
[http://digitalscholarship.unlv.edu/lib\\_articles/44](http://digitalscholarship.unlv.edu/lib_articles/44)
- Villanova University's Falvey Memorial Library. (2015). VuFind - Search. Discover. Share. *VuFind*. Retrieved January 20, 2015, from  
<http://vufind-org.github.io/vufind/>
- Webster, P. (2004). Metasearching in an academic environment. *Online Information Review*, 28(2), 20-23. Retrieved from  
<http://cat.inist.fr/?aModele=afficheN&cpsidt=15488717>