



建起橋樑：

從歐洲數位圖書館到歐洲數位報紙

蘇珊·瑞利

Susan K. Reilly

荷蘭 LIBER

E-mail: susan.reilly[at]kb.nl

Session:

119 – Users and portals: digital newspapers, usability, and genealogy – Newspapers Section with Genealogy and Local History Section

摘要：

研究顯示，研究人員喜歡以方便獲取的方式，特別是一次就能滿足所有需求的方式來獲取數位內容，因為這種方式便捷有效。除了方便獲取外，有質量保障的內容也極為重要。

為此，歐洲數位圖書館（*Europeana Libraries*）專案直接從歐洲地區的 19 個具領導地位的研究圖書館中，選取 510 萬個圖像、書籍、視訊及論文和文章。這些資料的來源保障了中繼資料的可靠性以及圖像的質量，同時，透過歐洲圖書館和歐洲數位圖書館網站的一個搜尋工具，可以使得所有資料變得易於獲取。

這種合作關係為未來的合作與創新奠定了基礎。持續性的整合設施以及全文搜尋的能力已經透過歐洲數位圖書館工程開發，並應用於歐洲數位報紙工程中的一個新的內容主體。這項專案將使整個歐洲地區 2900 萬頁的報紙內容，得以透過歐洲數位圖書館和歐洲數位平台在線搜尋。

簡介：

研究顯示，研究人員喜歡以方便，特別是一次就能滿足所有需求的方式來獲取數位內容，因為這種方式便捷有效。除了方便獲取外，有質量保障的內容也極為重要。

本文描述了歐洲數位圖書館（*Europeana Libraries*）專案的部分工作成果。這項工程將使得國家和研究圖書館的全文內容，得以透過一個單一入口網站進行搜尋。文章闡述了為了實現一個共同目標，而達到跨組織合作的動機及利益。並

介紹了歐洲數位報紙（**Europeana Newspapers**）專案背後的動機及可預期的成果。歐洲數位報紙工程亦受益於歐洲數位圖書館（**Europeana Libraries**）專案開發的入口網站及全文搜尋功能。

奠基之石

歐洲數位圖書館專案始於 2011 年 1 月，為期兩年。這項專案的想法源於一項對當時歐洲的研究圖書館（包括國家和大學圖書館）進行單一整合的需求。這些圖書館在過去曾經合作，透過歐洲數位旅遊（**Europeana Travel**）專案為歐洲數位圖書館——即歐洲文化遺產入口網站——提供主題內容。在歐洲數位旅遊工程中，有關旅遊和觀光主題的內容分別透過兩個聚合器進行整合，其中一個位於國家圖書館，另一個則專門用於其他專案中的圖書館之內容整合。這項專案一方面為歐洲數位化提供了高質量的數位內容，另一方面也促成了國家圖書館和其他研究圖書館之間的合作，因而極為成功。不過，專案本身也提出了是否可持續使用兩個聚合器的問題。

歐洲數位圖書館專案藉著開放國家圖書館整合服務：歐洲圖書館（**The European Library**）至研究圖書館中，提出了與持續性有關的議題。它使用這項服務器來整合歐洲研究圖書館內數量龐大的珍貴內容。到2012年12月專案結束之時，可以預期將有來自各個研究圖書館的510萬筆項目：包括1200部影片和視訊、85萬張圖像及430萬個文本（書籍、新聞文章、論文、信件）。大部分的內容為全文文檔，對於研究人員有特別的價值。為最大程度地發揮這些內容的潛力，該專案還設立開發全文搜尋功能以及一個搜尋入口網站，提供專門的搜尋工具。

合作關係

將研究圖書館和國家圖書館的內容整合在一起，還加速了歐洲一些主要圖書館網路的生成，如 **CENL**、**CERL** 和 **LIBER**。

歐洲國家圖書館館長聯席會（**CENL**）包括 48 個國家圖書館，目前擁有歐洲圖書館（**The European Library**）的聚合器，這是目前唯一的歐洲地區圖書館域聚合器。到目前為止，它只在國家圖書館作內容整合。

歐洲研究圖書館聯盟（**The Consortium of European Research Libraries**）的重心，是開發歐洲的印刷遺產並提高它的開放性，目前已經擁有包括研究圖書館和國家圖書館在內的 33 個正式成員機構。**CERL** 傾向於編目和中繼資料的研究。

LIBER 是一個歐洲研究圖書館的團體，擁有分布於歐洲及周邊地區的 420 所成員機構（國家、大學及其它研究圖書館）。其中有 19 個 **LIBER** 成員機構為本項專案提供內容。不過，歐洲數位圖書館工程開發的服務器最終將擴大充到所有的 **LIBER** 成員機構中。

上述三個組織基於一個強大的共性首次藉歐洲數位圖書館工程進行合作，這個共性就是他們的成員機構都擁有非常有研究價值的資料，並且全部有意願將這些內容開放給研究團體。歐洲數位圖書館工程透過對這項共性的剖析，將其轉

化為實際的、有共識的價值建議，並鞏固了圖書館網路之間的合作關係——這些都保障了工程產出成果的可持續性。

定義合作的價值

這些工程的價值，不僅在於為圖書館開發了一個單一整合服務，更為顯著的一點，是它還體現了一個可能，即把圖書館內有價值的研究內容開放給全世界的研究人員。它潛在地擴大了國家和研究圖書館本身已經建立的研究領域和地域的邊界。它運用了圖書館的共同聲譽，即值得信賴的、擁有可靠質量的訊息和良好中繼資料的提供者。圖書館與書籍緊密連結的關係已經是一個不容置疑的事實。提供數位書籍的全文內容和其他數位化檔案，例如：圖像、視訊和影音檔案，更不用說學術內容，比如技術論文和論文等，意味著研究人員可以獲得更加豐富的研究結果。透過這種方式提高這類內容的能見度，可以提升圖書館大量投資數位化運作的影響。

對於研究人員來說，其價值在於能夠在單地點獲取並研究大量的文化遺產和相關研究內容。研究顯示，研究人員正提出更為複雜的研究項目，而研究活動的時間卻更短。因此，這種一次就能滿足所有需求的方式實在頗具吸引力。

設計入口網站

大部分歐洲數位圖書館工程整合的內容將可透過入口網站搜尋。但是考量到整合內容的性質，歐洲圖書館入口網站重新開發，將人文科學的研究人員視為終端用戶。這個特別與人文科學研究族群相關的命題，對於他們來說，研究資料的定義非常複雜：

「人文社群如果要從『精品計畫』的招牌轉移至擁有穩固的理論和方法基礎上，就需要非常大量的數位資源，以及共同的工具、服務器和資料庫。」

數場專題討論會、終端用戶測試以及案頭調研，都被納入新的入口網站的設計中，它具有如下的特點：

- 全文內容搜尋的功能。
- 檢查每一份物件中繼資料的機會，以便最終為研究目的獲得更大的數據資源庫。
- 更廣泛地使用的 CERIF 主題標題。這使得進行一個跨文本資料庫物件的研究（透過共同的主題或時間框架連接的）變為可能。
- 與歐洲各國家圖書館、研究圖書館和大學圖書館等合作開發泛歐典藏。這是當前視覺展覽（歐洲範圍內一系列資源展覽）的延伸。
- 年表顯示幾世紀來對某特定詞彙進行檢索的時間。
- API 將使資料庫可以在歐洲圖書館入口網站之外的情境下加以分析和演示。這意味著研究人員可以將他們的資料內容帶到自己的研究情境，並探索使用該內容的新方法。
- 將紀錄直接輸出於廣泛使用的文獻管理服務器，比如 Mendeley and Zotero。

入口網站上線後會持續重複開發，與人文科學和數位人文科學新興的研究活動保持同步。針對研究人員如何開發並使用這一獨特內容進行深入的研究，並同步確立使用策略的內容。

建立橋樑

和整合服務器一樣，歐洲數位元圖書館工程開發的入口網站和全文搜尋功能將被用來呈現一類新型內容：報紙。OCR 的最新發展使得 IMPACT 專案可以應用於來自歐洲 12 個國家和研究圖書館的報紙內容。1800 萬頁的報紙在精細化處理後便可在歐洲數位圖書館和歐洲圖書館的入口網站獲取。歐洲數位報紙專案的主要目標是要應對一個非常特別的挑戰，就是可以搜尋舊報紙的全文內容。它會應用精細化的方法來進行 OCR、OLR／文章切割、命名實體辨識（NER）以及網頁等級識別。

歐洲數位圖書館所開發的成果目前也應用於歐洲數位報紙工程：

1.整合：

目前已可以識別四種類型的數位報紙：

- a) 只有結構性中繼資料的圖像
- b) 用於搜尋的帶有結構性中繼資料和全文的圖像（OCR）
- c) 帶有結構性中繼資料、文章識別（OLR／文章追蹤）及 OCR 的圖像
- d) 帶有結構性中繼資料、OLR、OCR 和充實語義的圖像

所有可搜尋的資料都將為歐洲圖書館獲得。資料將轉換成 EDM（歐洲數位圖書館的資料模型）並傳佈。

2.中繼資料標準化：

目前所使用的中繼資料格式有許多類型。為提高數位內容的開放性，必須採用共同的標準。將對現有的所有中繼資料格式加以識別，並提交社群最佳實踐方案。

3.更好的顯示功能：

報紙目前所面臨的挑戰是搜尋的便利性及線上顯示效果。線上報紙專案將著眼於歐洲數位圖書館專案的成果，在此基礎上詳細說明合適的搜尋和顯示要求，這也可以用於歐洲數位圖書館。

目前，最遠可追溯到 18 世紀的報紙內容，都將可以完全在線上搜尋。而且，用戶將可以瀏覽這些報紙，以及藝術圖片、照片、相關論文、書籍和文章。透過此一新型的搜尋方式展示報紙文本，將產生新的連結，並打開新的合作大門。

這項專案也會帶來其他利益，包括 OCR、OLR／文章追蹤、NER 以及網頁等級識別等，將成為報紙內容更新的方式。針對每一種技術任務都將確認並發佈最佳實踐建議。這對於 CERL、CENL 和 LIBER 網路的拓寬將會產生巨大的影

響。它在現在以及將來都會降低報紙數位化的成本，並提高數位報紙資源的開放性。

結論

歐洲數位圖書館是一項絕佳地運用網路。其源自歐洲圖書館對於單一整合這項特別實際的需求。這項工程使得主要的圖書館網路連結在一起。透過致力於一項共同的願景，網路已經為研究機構創造了一種擁有巨大潛值的資源。基於連結館藏內容和研究人員的願景，它還為國家及研究圖書館創造了合作無間的條件。歐洲的圖書館擁有豐富且多樣的館藏，尤其是報紙產業。在精細化技術開發到透過有意義的方式，進行全文展示地步的同時，將這些產業線上化，為研究人員創造了一個巨大的機會，使得他們可以與歐洲豐富的文化遺產資料互動，並建立新的連結。這也是組織、網路及機構合作產生創新成果、提高效率及得以相互傳遞的真切體現。這項工作的影響所即將不僅涉及圖書館，或增強歐洲地區文化遺產的傳遞，還將對全世界每一個擁有大量印刷文化資料的國家產生深遠的影響。

參考文獻：

Bulger, Monica, et al., *Reinventing Research? Information Practices in the Humanities*. London: Research Information Network, 2011.

Christine L. Borgman. "The Digital Future is Now: A Call to Action for the Humanities" *Digital Humanities Quarterly* 4.1 (2010). Available at: <http://works.bepress.com/borgman/233>

De Rosa, Cathy et al. 2005. *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. Dublin, Ohio: OCLC. <http://www.oclc.org/reports/2005perceptions.htm>.

European Commission, Maurice Lévy, Elisabeth Niggemann, and Jacques de Decker. *The New Renaissance*. Brussels: European Commission, 2011. http://ec.europa.eu/information_society/activities/digital_libraries/doc/reflection_group/final_report_20cnds.pdf

Europeana (2012) Retrieved from <http://www.europeana.eu> on 19th April 2012.

Europeana Libraries (2012), Retrieved from <http://www.europeana-libraries.eu>, on 9th March 2012.

Europeana Travel (2012), Retrieved from <http://www.europeanatravel.eu/> on 19th April 2012.

IMPACT, Retrieved from <http://www.impact-project.eu/> on 19th April 2012.

The European Library(2012), Retrieved from <http://www.theeuropeanlibrary.org> on 9th March 2012.