

# 日治時期日文臺灣文獻數位典藏計畫概述

## The Current State of Digital Archives of Japanese Literature Materials for Taiwan Research in National Taiwan Library

蘇倫伸

Lung-Sheng Su

國立中央圖書館臺灣分館編輯  
輔仁大學圖書資訊學系兼任助理教授  
Editor, National Taiwan Library

Assistant Professor, Department of Library and Information Science, Fu Jen Catholic University

### 【摘要 Abstract】

我國推動數位典藏計畫，將豐富的文化藝術典藏及文獻資料數位化，不但充實網路資訊內容，同時可以提昇資訊的傳播與利用，促進社會文化的發展。國立中央圖書館臺灣分館現藏孤本（絕版）日文舊籍圖書資料約有十六萬餘冊，在臺灣學研究上是相當重要的文獻資料，如能將此資料數位化，並建立索引資料庫，對於臺灣資料的應用與推廣，將更具效益。

本文主要說明近二年來國立中央圖書館臺灣分館「館藏日文臺灣資料數位典藏計畫」執行情況，以促進資訊交流與經驗分享。

Digital archives will be abundant the contents of the internet and will speed to the information distribution. The National Taiwan Library owns a lot of Japanese literature materials for Taiwan research (about 160,000 titles). These are very important materials for Taiwan research. Digitalized and indexed these materials will be helpful for readers, especially for Taiwan study researchers.

This article describes the project entitled "Digital archives of Japanese literature materials for Taiwan research in NTL" in recent years. It also aims to share the experiences and results to whom concerned.

### 【關鍵詞 Keyword】

數位典藏；臺灣文獻；數位學習；數位圖書館

Digital Archive; Literature Materials for Taiwan Research; E-learning; Digital Library

## 壹、前言

近十年來，我國推動數位典藏計畫，將豐富的文化藝術典藏及文獻資料數位化，不但充實網路資訊內容，同時可以提昇資訊的傳播與利用，促進社會

文化的發展。國立中央圖書館臺灣分館（以下簡稱本館）現藏孤本（絕版）日文舊籍圖書資料約有十六萬餘冊，原系日治時期台灣總督府圖書館及南方資料館所藏，另有明清刊本線裝書約六萬餘冊館藏，前者由於二次大戰期間，迭遭

戰患，一再遷藏，致遭受嚴重破損；後者則因藏用年久，致蟲蝕水漬受損頗為嚴重。民國 74 年間，當時教育部李部長錫公蒞館巡視，審察其館藏圖書資料之珍貴及破損之嚴重，曾特別指示，務需妥善維護及保藏，尤須從速加以修補裝訂，並攝製縮影捲片以重蒐藏並供流通。為妥善辦理本項圖書典藏維修工作，本館乃擬訂「日文舊籍圖書修護裝訂十年計畫」暨「線裝圖書修補裝訂五年計畫」進行修補，又為便於永久保存及借閱流通起見，同時將上述珍藏資料攝製微捲（註 1）。目前，已完成圖書及期刊近 2823 捲微捲（16mm 微捲 2268 捲及 35mm 微捲 555 捲），報紙（包括：日據時期報紙、政府遷臺前後時期報紙、香港地區報紙等）約 2897 捲微捲（35mm 微捲）。

雖然，本館已將大部分珍貴日文舊籍圖書及報紙進行微捲拍攝並提供讀者利，惟於數 10 年前先進的微縮技術基於保存資料有其便利與效益，但在目前資訊化社會時代，讀者服務與利用上較為不便，必需要到館且利用特定的微縮閱讀機才能檢閱資料相當不便，同時，資料並未建立索引，對於取得日據時期台灣之資料，常常需要地毯式蒐尋，頗耗費研究人員時間與心力。如能將已製成微捲之資料數位化，並建立索引資料庫，對於臺灣資料的應用與推廣，將更具效益。因此，本館於 93 年 12 月遷館工作告一段落後，積極投入數位典藏工作，於民國 94 年 8 月規劃館藏日文臺

灣資料數位典藏計畫報陳教育部，列入教育部國立社教機構服務升級第 2 期計畫中，95 年 11 月修訂本館國立社教機構服務升級第 2 期計畫——「館藏日文臺灣資料數位典藏計畫」報部審查，96 年 5 月教育部核定同意辦理。本文主要說明近 2 年來「館藏日文臺灣資料數位典藏計畫」執行情況，以促進資訊交流與經驗分享。

## 貳、「館藏日文臺灣資料數位典藏計畫」內容與成果

### 一、計畫內容

本項計畫主要目標為將本館已拍攝成微捲之現行藏孤本（絕版）日文舊籍圖書資料及報紙（包括：日據時期報紙、政府遷臺前後時期報紙、香港地區報紙等）進行數位化典藏與利用，共計 5720 捲微捲。同時建立線上索引資料庫（預計 8 萬 6 千筆），透過網際網路提供讀者查詢臺灣資料並可即時線上取得所需資料之數位內容，提昇服務品質與研究效益。

在執行策略方面，本項計畫

#### （一）主要工作項目包括：

1. 本館已拍攝成微捲之現行藏孤本（絕版）日文舊籍圖書資料及報紙微捲清點及複製。
2. 數位典藏系統軟、硬體設備建立及測試。
3. 委託廠商進行微捲資料轉換為數位影像資料，並載入數位典藏系統。

4. 於數位典藏系統進行資料索引編製工作及與數位影像資料連結。
5. 辦理臺灣學研究相關主題之國內或國際會議。

## (二)執行方法與步驟：

1. 本館已拍攝成微捲之現行藏孤本（絕版）日文舊籍圖書資料及報紙微捲清點。
2. 於進行微捲資料數位化前，將先行複製一份，以此備份提供廠商進行資料數位化，以避免因數位化過程損及微捲母片或遺失。
3. 進行數位典藏系統軟、硬體設備需求規格制定、招商作業及系統安置、測試與營運管理制度建立。
4. 聘請具有精通日文及圖書館學域之專業知識人員進行檢核委外整理之文獻及索引品質。
5. 聘請具有精通日文及圖書館學域之

專業知識人員進行索引編製，及資訊專業背景人員進行資訊系統維護管理。

6. 委託廠商進行微捲資料轉換為數位影像資料，並載入數位典藏系統。
7. 進行資料索引編製及校對工作及並與數位影像資料連結。
8. 辦理臺灣學研究相關主題之國內或國際會議，推廣利用本項計畫成果。

## 二、執行成果

### (一)臺灣文獻數位化資料內容與數量

自 96 年 5 月起至 97 年 8 月止，「館藏日文臺灣資料數位典藏計畫」在數位影像掃描共計數位化微捲影幅數為 597,116 影幅（頁），詮釋資料撰寫筆數為 29,863 筆，建置情形如表一。預估至 97 年 12 月底，數位化頁數 1,265,000 頁，圖書 3842 種，雜誌 156 種（125,867 篇）。

表一：數位影像檔與詮釋資料統計表

資料名稱	數位影像檔 (影幅)	詮釋資料 (筆)	備註
臺灣教育會雜誌暨臺灣建築會誌資料庫	32,000	13,683	
日治時期臺灣地圖資料庫	764	764	
日治時期電子書資料庫	16,159	200	
《臺灣警察協會會誌》	54,149	12,599	含摘要 6,673 筆，無摘要 5,926 筆
《臺灣通信協會雜誌》	1,079	575	摘要 328 筆，無摘要 247 筆
其他雜誌（圖書）（註 1）	219,426	0	詮釋資料尚未建置，預定於 97 年底建置完成
日治時期圖書全文影像系統	273,539	2,042	摘要 2,020 筆，不含摘要 22 筆
合計	597,116	29,863	

註 1：其他雜誌包括：《台灣通信協會》、《臺中州教育》、《敬慎》、《新臺灣》、《臺灣產業雜誌》、《臺灣の水利》、《臺灣醫事雜誌》、《實業之台灣》、《臺灣の山林》、《內外情報》、《臺灣總督府圖書館新著圖書目錄（圖書）》、《臺灣年表（圖書）》、《外國為替管理關係法令（圖書）》、《臺灣に於ける母國人農業植民（圖書）》、《婦人界》、《台灣公論》、《台灣鑛業會報》、《語苑》、《臺灣鐵道》、《高山國》、《ニヒタ力》、《臺灣文藝》、《木瓜》、《蕃情研究會誌》、《臺灣經濟雜誌》、《蕃界》、《綠珊瑚》、《台灣水產》、《社會事業之友》、《臺灣農友會會報》等。

## (二)品質管制與追蹤工作會議

本館為有效掌握廠商數位化工作進度與品質，邀請學者專家召開會議，提供專業意見與諮詢。本館於97年2月27日召開「館藏日文臺灣資料數位化諮詢委員會」第一次會議，邀請林富士委員、項潔委員、薛理桂委員、蔡錦堂委員、鍾淑敏委員出席指導，並對如何兼顧數位影像品質與原件保存、數位典藏電子影像的開放原則與管理機制、詮釋資料審查原則等多項議題提供諮詢建

議。另外，本館分別成立「臺灣教育會雜誌」文獻數位化工作會議、「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」工作會議及97年度「館藏日文臺灣資料數位化」工作會議，邀請得標承作廠商及本館數位典藏計畫工作同仁定期召開工作會議。截至97年9月為止共召開16次會議，主要針對各類型資料數位化、檔案命名及詮釋資料撰寫所遭遇問題進行討論與處理，會議日期摘要如表二。

表二：品質管制與追蹤工作會議日期表

編號	會議名稱	會議時間
1	「臺灣教育會雜誌」文獻數位化第一次工作會議	96年8月7日
2	「臺灣教育會雜誌」文獻數位化第二次工作會議	96年8月21日
3	「臺灣教育會雜誌」文獻數位化第三次工作會議	96年9月11日
4	「臺灣教育會雜誌」文獻數位化第四次工作會議	96年10月17日
5	「臺灣教育會雜誌」文獻數位化第五次工作會議	96年11月14日
6	「館藏日文臺灣資料（雜誌類）數位化」第一次工作會議	96年12月25日
7	「館藏日文臺灣資料（圖書類）數位化」第一次工作會議	96年12月27日
8	「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」第二次工作會議	97年1月23日
9	「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」第三次工作會議	97年3月6日
10	「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」第四次工作會議	97年4月2日
11	「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」第五次工作會議	97年5月2日
12	「館藏日文臺灣資料（雜誌類）暨（圖書類）數位化」第六次工作會議	97年6月5日
13	97年度「館藏日文臺灣資料數位化」第一次工作會議	97年6月4日
14	97年度「館藏日文臺灣資料數位化」第二次工作會議	97年7月3日
15	97年度「館藏日文臺灣資料數位化」第三次工作會議	97年8月12日
16	97年度「館藏日文臺灣資料數位化」第四次工作會議	97年9月4日

## (三)數位典藏系統硬體建置與檢索系統

為促進數位典藏系統功能發揮，本館「館藏日文臺灣資料數位典藏系統」硬體建置考量設備擴充性、資料查詢之

便利性、數位資訊安全性、與研究人員互動性及未來發展前瞻性等因素進行規劃。在主機系統部份，採用多主機服務，以避免單一主機負荷過重，影響檢

索效能，規劃以圖書資料庫及其刊資料庫分別由不同主機服務。在影像檔案儲存採用 SAN 磁碟陣列儲存系統，支援 RAID-0、RAID-1、RAID-0+1、RAID-5 等執行方式，以避免因單一硬碟毀損造成資料遺失。在網路傳輸方面，採用光纖網路，增進資料傳輸效能，同時搭配入侵偵測防禦（IDS）系統及防毒閘道設備，以過濾病毒或降低駭客入侵之機會。

在數位典藏檢索系統功能方面，本館日治時期圖書全文影像系統及日治時期期刊全文影像系統之檢索系統功能需求，係參考國科會數位博物館專案計畫 Metalogy 系統功能特色為基礎，委託廠商開發，主要功能包括：可載入 DTD 建立資料庫、定義資料庫之 Schema、Metadata 編目功能、索引典建立及權威控制功能、數位化資料的管理與描述、一般查詢及 Web 查詢功能、資料顯示功能、影像瀏覽功能、轉入 XML 記錄、載出 XML 記錄、使用者權限控制、訊息管理（註 2）、管理系統、統計、用戶管理、支援 OAI-PMH 分散檢索協定、全文授權管理等。本館日治時期期刊全文影像系統網址：<stfj.ntl.edu.tw>，日治時期圖書全文影像系統網址：<stfb.ntl.edu.tw>。

## 參、遭遇問題與因應策略

### 一、原件保存與數位化資料品質之權衡

本館「館藏日文臺灣資料數位典藏

計畫」採用微捲影幅資料為主轉製成數位檔案，而非以原件資料掃描或翻拍，其主要的目的是基於原件的保存與資料的利用。20 餘年前，本館針對日治時期總督府圖書館所遺留圖書資料進行微捲拍攝工作，部份圖書資料已拆開拍攝並重新裝回，為避免數位化工作進行需再重新拆裝圖書資料，造成原件損害，對於不影響資料內容辨識的微捲影幅內容，以微捲影幅轉製成數位檔案。然而，由於微捲無法顯示採色效果，所以針對彩色（2 色以上）印刷之圖書資料或者微捲拍攝不清的資料，在不拆圖書資料的原則下，進行以原件資料掃描或翻拍。在極特殊的情況，例如：圖書資料內容夾於圖書裝訂線內，非拆卸圖書無法辨識內容且該原件資料尚容許拆卸情況下，基於資料利用的完整性，才進行原件拆裝掃描。因此，有些數位檔案靠近書脊中的文字略有弧度（但不影響資料內容辨識），也是因為保護原件，在不拆圖書資料的原則下，進行以原件資料掃描或翻拍所產生的結果。

### 二、研訂本館數位化檔案命名原則

數位資源命名的一致性是在網路上是相當重要也是必要。WWW 是網際網路主要的服務，而 URL（Uniform Resource Locators）則是 WWW 進行數位資源超連結的基礎。URL 是一種容易實作及富彈性的機制，但也是一種不穩定，不可靠的連結機制，一旦主機移位、貯存路徑或檔案名稱變更時，URL 就無法正確定位，我們在瀏覽網頁時，

常會出現 http v1.0/1.1 Error404 訊息，這表示所點選欲連結的資源已經被移除，以致無法更進一步利用（註3）。另外，資源的命名在網路環境中為何會成為重要的問題，從下列幾個角度來看，亦可得知數位資源的命名確為要長久典藏資訊的數位資訊系統中必需處理的問題（註4）：

- (一)從數位資源儲存管理的角度而言，命名問題非常重要：在一機構內，進行資料數位化的單位可能不只一個，各單位的檔案，各種類型都有，如果各單位對於數位資源名稱不能事先協調，而有各自的命名方式，則可能產生重複、系統不一致等題，時間久了之後，將難以辨識，硬體資源也不易有效分配。
- (二)當一個區域或一個國家，數位資源日漸豐富後，民眾將不再滿足於分別連結到各個資料庫查詢的服務，而希望有一整合性的使用方法，如建立聯合目錄，或將數位資源集中。此時，可能需將相同領域的資源予以集中，而若各機構事先有一致的命名方式，將會減少重複、難以辨識、不一致、書目資料與數位資源連結不易等困擾。
- (三)從電子商務的角度視之，也需要賦予數位資源唯一的識別碼。以目前圖書、期刊或錄音錄影資料為例，若非有 ISBN、ISSN 及 ISRC 等號碼，則圖書、期刊或錄音錄影資料的國際銷售將困難重重。而今，當要銷售的資料單元是書中的一個章節、一個圖、

一個表而不是整本書，則其編碼方式也需細到章、節、圖、表才行。

本館為使數位檔案便於管理及日後與其他單位數位內容交換，本館參考「行政院文化建設委員會國家文化資料庫數位化檔案命名原則」（註5）及數位典藏國家型科技計畫之經驗，研訂本館數位化檔案命名原則，並從檔案命名中可以辨識是那一種資料（圖書或期刊）、第幾期、第幾篇（章或節）的第幾頁及那一年進行數位化的資料，以便於數位檔案與原件資料審核校對及確保檔案名稱的唯一性，檔案名稱越長表示資料型態變化程度越大。本館檔案名稱結構與文建會檔案命名原則相似，其結構為：

**「計畫代碼—物件類別—物件代碼—多部份序號—使用目的.附屬檔名」**

在此檔案命名結構下，物件代碼與多部份序號由於需因應資料型式的多變性，檔案命名原則中需加以規定。例如：日治時期期刊的物件代碼：以「總督府代碼（sxt）+索書號+卷期」表示，在「卷期」的表示中，由於日治時期期刊卷期有多種表示方法，目前在檔案命名原則中，「卷期」表示方式有四種：

1. 期刊只有期號(總號)—以期號(n+3碼總號)表示，若期號超過一千期以上，則直接以期數表示。例如：臺灣總督府鐵道部報（第183期和1944期） sxt\_0798\_27\_n183；sxt\_0798\_27\_1944

2. 期刊有卷期號—以卷(v+3 碼卷號)期(n+3 碼期號)表示。例如：sxt\_0773\_19\_v001n001，如果期刊出現“創刊號”、“第一期/一月號(新年號)”同時並存的情況。“創刊號”以 n000 命名，第一期/一月號(新年號)以 n001 命名。
3. 期刊無任何期號或編碼—以出版年月(y+4 碼西元年，m+2 碼月份)表示，例如：sxt\_0773\_19\_y1943m08
4. 期刊以春、夏、秋、冬或第一~四季編號—以實際出版年月表示。例如：1950 年夏季號(於 6 月出刊) sxt\_0773\_19\_y1950m06，以上 4 種情況，若期刊卷期同時有上述多種情況並存時，依狀況編號順序優先採用。

在多部份序號亦有因應資料型式的多變性，檔案命名原則中也需加以規定，除此之外，圖書類及期刊類不同類型的檔案命名原則，以及 PDF 檔案(以多頁為主)與 Tiff 檔案(每頁)的檔案命名原則，在多部份序號表示方法亦有差異，由於篇幅所限，可以參考本館「期刊資料數位檔案命名原則」、「圖書資料數位檔案命名原則」及「地圖資料數位檔案命名原則」不再贅述。

## 肆、結論

本館「館藏日文臺灣資料數位典藏計畫」預計進行三年，目前已將詮釋資料格式、數位檔案命名原則、硬體設備

及檢索系統等數位典藏的基礎工作已初步建立完成，未來一年將朝持續增加資料數量、建立整合查詢介面、推廣服務以及加值運用等方向努力，以增加日文舊籍圖書資料及日治時期報紙、政府遷臺前後時期報紙、香港地區報紙等珍貴資料得以保存與利用，增進本館處理數位資訊的能力，促進數位典藏的經驗成果分享。

## 附 註

- 註 1 國立中央圖書館臺灣分館編印，館藏資料微縮目錄（臺灣分館，民 84）序文。
- 註 2 陳昭珍、陳雪華、陳光華，「數位圖書館與博物館 Metadata 管理系統——Metalogy 之設計」TANET 2000 台灣區網際網路研討會論文集，頁 497-498。<<http://www.ncku.edu.tw/TANET2000/download/paper/C4-2tanetN05.doc>>（2008/09/09）
- 註 3 陳昭珍、何佳欣，「數位化圖書館互通檢索機制之探討」中國圖書館學會會報 70 期（2003 年 6 月），頁 55。
- 註 4 同前註，頁 55-56。
- 註 5 「行政院文化建設委員會國家文化資料庫數位化檔案命名原則」（2004 年 6 月）<[http://km.cca.gov.tw/download/數位檔案命名原則\\_v20\(20040608\).pdf](http://km.cca.gov.tw/download/數位檔案命名原則_v20(20040608).pdf)>（2008/09/07）