

# 紙質文獻類的 雜誌書籍之數位化

洪淑芬 國立臺灣大學圖書館特藏組編審、國立臺灣大學圖書資訊學系博士生

【摘要】本文主要就雜誌書籍類資料之數位化，指出宜因應此類資料之特性，進行有別於古老珍貴資料數位化之規格與規範之訂定，並基於費用的效能面之考量，提供如何有效地進行影像數位化、全文數位化、Metadata 建置及檔案命名等之建議。

【關鍵詞】紙質文獻數位化；雜誌數位化；書籍數位化；數位化規格；數位化規劃

## 一、前言

數位化的工作，因為資訊科技的持續發展而日益重要。然而因為數位化需要具備必要的知識，以利適切的規畫，又需要相當的經費，才可能執行；因此，不論哪一個國家，由公家出資進行數位化，在接受申請數位化補助的過程中，既需考量專業能力，也必須考量資料的珍貴稀有性，而以貴重資料為優先數位化的對象。臺灣發展數位化的情形也是如此。

但是，一方面是數位化執行有年之後，有些機構已完成古老珍貴資料之數位化；另一方面是因為網路科技的發達，使典藏機構開始思考透過網路推廣館藏利用。尤其是一些較早期的圖書雜誌，或因紙質脆化的問題，或因時代的關係，而

使留存的狀況未見理想。這類書籍雜誌之數位化，已漸為典藏機構與研究者所重視。因此，數位化的對象，比過去更為多元；雜誌書籍類資料之數位化，逐漸被人所重視。

過去國家執行數位典藏，基於規格的標準化及教育訓練與經驗分享之需，在由行政院國家科學委員會（簡稱「國科會」）主導的「數位典藏國家型科技計畫」方面，已出版一系列的教育訓練推廣教材（[http://d1m.ntu.edu.tw/01\\_1\\_2.htm](http://d1m.ntu.edu.tw/01_1_2.htm)）、工作流程指南（[http://content.ndap.org.tw/main/doc\\_c.php?class\\_format=18&format\\_type=數位化工作流程指南](http://content.ndap.org.tw/main/doc_c.php?class_format=18&format_type=數位化工作流程指南)）。在由行政院文化建設委員會（簡稱「文建會」）主導的「國家文化資料庫計畫」方面，於「國家文化資料庫」網站也提供了一系



列的技術規範標準 ([http://km.cca.gov.tw/download/rule/93\\_do\\_scale.doc](http://km.cca.gov.tw/download/rule/93_do_scale.doc))。但是，該等技術相關資源，對於雜誌書籍類的一文多頁情況之數位化處理，尚少相關介紹。筆者於數位化計畫的規劃執行經驗，早期主要是針對貴重的《淡新檔案》、《岸裡大社文書》、《古碑拓本》及老照片等，近期則曾針對雜誌類的資料進行規劃與執行。雜誌類資料之實體，與前面數者差異極大，因此，數位化各相關事項的規劃也與過去執行古文書類資料之數位化不同。筆者過去曾將數位化的實務經驗，撰著《文獻典藏數位化的實務與技術》一書，本文重點主要是在介紹該書所未涵蓋的雜誌類資料數位化之實務規劃。因為書籍類資料在編排上與雜誌有相似之點——雜誌之內含單篇文章，書籍之內可分章節，可能分別作為 Metadata 所將對應的文字內容之單位。因此，雜誌數位化之規劃，實亦可供規劃書籍數位化之參考。

## 二、從年代久遠的珍籍資料到近年的成冊文獻

年代久遠的珍籍資料之數位化，或因原件實體結構特殊（例如，長軸型、經摺裝、黏連成串）、或因資料尺寸大小不一、或因資料狀況脆化而有裂損之虞而不得不考量因應之道，因此使得製作上所需的費用單價偏高。但是，當所將數位化的資料對象是較近期的雜誌或書籍類時（例如，日治時期的書籍刊物、臺灣光復前後至民國六、七十年代的書刊等），則必須思考以下規劃上需考量的問題：

### （一）價格

雜誌書籍類的資料狀況，明顯的較年代久遠的珍籍資料單純，而且需要數位化的數量也遠為龐大。資料狀況單純，可以提升影像數位化的速率；數量龐大的情形，在商務交易上，一般合理的情形是可以獲得較低單價的優惠。因此，雜誌書籍類資料之數位化，單價上應該可以相當程度的降低。

### （二）機器的選擇

對於極珍貴之資料，為了避免數位化過程中造成對資料原件的損害，而不利於長久保存且減損資料本身的價值，可以針對不同的資料類型及資料之實體狀況，選擇適用的機器。只要數位化對象物是與國家歷史文化相關者，執行機構透過招標專案遴選廠商時，可被允許就機器、人員之專業等，進行規範。不過，對廠商的機器配備及人員之專業程度要求越高，相對地，製作單價也將隨之提高。

雜誌書籍類資料之數位化，因為資料型式較為一致，尺寸又大約在一定範圍之內，因此，選擇數位化機器之重點考量，主要在於速率、解析度及對雜誌書籍本身的保護等。

過去臺灣大學圖書館進行珍貴典藏資料之數位化專案，大力借重於掃描機器。尺寸偏大者（資料普遍高達一、二米以上），則採用先拍攝底片再掃描底片的方式。但是，近幾年來，數位相機本身的規格有大幅的提升之外，可搭配使用的軟體亦有長足的進步；兩者搭配之下，其功能已可比擬高規格的掃描機器；而且如果搭配適當的輔具（擺置雜誌書籍之裝置），則亦可達到相當於平床式掃描機（使資料文字面朝上，由上方擷取影像而進行數位化的機器）之資料保護效果。然



而因為數位相機之日產能甚高（因為擷取檔案與存檔速度極快的緣故），其價格又遠低於進口的平床機器，因此，可大幅降低製作單價，而成為雜誌與書籍類數位化的一個不錯的選擇。

### （三）規格

雜誌與書籍類資料之數位化，在影像檔的規格方面，如果比照珍貴資料文物，則數位化成果的檔案總容量將十分龐大，在儲存上不甚有利。現在電子資料庫中，此類資料之全文影像，大多是以 PDF 檔案格式呈現。PDF 檔案格式，已經是一文多頁的全文影像格式之趨勢。臺灣大學圖書館於進行大批雜誌數位化時，即以 PDF 檔案作為最終置於資料庫系統中的檔案；詳細的說明請參閱下文三之（一）。質言之，當數位化的對象資料不同時，應在規格方面作適當的調整，較為務實可行。

## 三、雜誌書籍類數位化的相關工作之實務

以下茲對照說明 1. 文件型或單件文物及 2. 雜誌書籍等二類資料之影像數位化、全文數位化、Metadata 建置、檔案命名、浮水印嵌入等實務工作可能的異同：

### （一）影像

#### 1. 文件型或單件文物

一般珍貴的文件型或單件文物的典藏數位化計畫，在數位影像的規格方面，大多分為以下幾種不同目的之格式：

- (1) 典藏級檔案（一般是 Tiff 檔，300dpi+）。
- (2) 商務應用級檔案（一般是典藏級檔案轉為 Jpeg 格式，或再將之作壓縮，壓縮為原檔案的 75%

較理想）。

- (3) 網路傳輸用檔案（一般是商務應用級檔案再轉檔為 150dpi 或 72dpi 之檔案，可加上再縮小圖檔尺寸，使檔案減小而利於網路傳輸）。

#### 2. 雜誌書籍

以臺灣大學雜誌數位化為例，雖然最終目標是產出 PDF 格式之檔案，將之載入資料庫系統中，以提供使用，但是因為是以高規格數位相機（數位相機的 CMOS 或 CCD 在 1100 萬像素以上的等級）進行數位化，所以在產出 PDF 檔的過程中，其實亦保留或轉檔出以下格式之檔案，並分述理由如下：

##### (1) DNG Archive File (RAW)

該種格式的檔案本身，以少量的空間保留了最完整的影像內容資訊，只要經由相關軟體，即可依設定而轉檔出 Tiff 格式、Jpeg 格式等不同類型、不同解析度之檔案，對原始影像的完整資訊之保存、儲存空間的節約等，都較為有利；唯需要對轉檔用的軟體有所了解，才能順利轉檔。

##### (2) 24 bit 彩色 Jpeg (optional)

有的雜誌本身含有一些彩色頁，轉出彩色的 Jpeg 檔，可以在必要提供彩色頁時使用。

##### (3) 8 bit 灰階 Tiff (optional)

如欲節省儲存空間，不需彩色頁，但需要以照片的灰階效果呈現雜誌中的照片，則可以只轉出灰階之檔案。

##### (4) 黑白 Tiff

為轉置黑白的 PDF 檔之前置歷程，因檔案並不大，所以可以考慮保留。

##### (5) PDF

為資料庫系統中所將使用的檔案格式。經轉



為 PDF 格式之檔案後，再進行同一文各頁之串檔，串成一文多頁。而臺灣大學圖書館近期進行數位化的雜誌，大多數只有封面是彩色，其餘為黑白之刊物。因考量保留封面的豐富彩色資訊，以及黑白格式可以使檔案降至最小，乃決定只當含彩色印刷時（主要是封面與封底的部分）才取用彩色頁，其餘都取用黑白格式，進行串檔。串檔之容易與否，與拍攝之時的檔案命名極為相關，詳見下文三之（四）。

如果產出 PDF 檔的方式是以一般的掃描機直接設定為 PDF 檔，而且是一文多頁的模式進行掃描，則大多在一次設定後，即產出所設定的一種格式之檔案，能夠彈性運用的可能性較小，也無法保留影像內容資訊最完整之原始檔案。

串連一文多頁的途徑，尚可利用 J-Tiff 格式或 Jpeg 格式之檔案。但是，以此二種格式之檔案進行串連一文多頁，會使最終檔案極大，而不利傳輸，因此，需考慮配合使用壓縮圖檔軟體。唯此種情形，使用者端亦必須下載開啟該種壓縮圖檔之軟體，才能開啟影像圖檔。相較之下，仍以網路上較為通用的 Adobe Acrobat Reader 即可開啟的 PDF 格式之檔案，通用性與便利性較高。

至於近幾年來新產出的電

子文件，如果希望將之以 PDF 檔案格式在網路上提供使用，則將已排版好的電子檔直接轉存為 PDF 檔，可以獲致不錯的效果；尤其如搭配 Adobe 的 CS2 軟體進行轉檔壓縮，將可使檔案壓縮至極小，卻可使字體影像達到最佳效果，例如，一篇 21 頁之 PDF 格式之檔案，檔案大小僅 165KB，但放大倍率可以極高，仍無損解析度，如〔圖一〕。



圖一、以文字檔直接轉檔之 PDF 檔：頁面文字放大示意圖



## (二) 全文

### 1. 文件型或單件文物

年代較久遠的珍貴文件型資料，其原文大多為書寫文字或者版刻印刷之文字。此二類型之文字，均無法透過 OCR 的方式，進行全文之數位化。因此，如需進行全文建檔，一般均必須透過人工逐字輸入。

### 2. 雜誌書籍

雜誌書籍全文之數位化，其主要功能之一是在資料庫的後端支援全文檢索，否則提供影像檔，即足以滿足使用者閱覽的需求。要將雜誌書籍的內文進行文字輸入之數位化，經測試結果，以光筆逐行閱讀，或整頁一次進行 OCR 辨識轉換，都只有當印刷字體極為清晰，頁面為純白、完全無底色，而且版面單純無特殊排版或頁首、頁尾、頁框邊緣文字的情形，才可獲致理想的結果。而雜誌書籍實際的情形是，往往含有各種版面的設計、頁首頁尾文字、大小或粗體文字混雜、套印底色的情形，因此使 OCR 的效果大打折扣；所需校對的功夫人力，難以為計。因此，只有近年來以電子檔產出的文件，應用原來的文字電子檔的部分，比較易行。而因為印刷類的雜誌書籍，文字在閱讀上不會有難辨識之問題，因此除非有特殊目的，否則全文數位化並非必要。

## (三) Metadata

### 1. 文件型或單件文物

數位化的對象資料如果是文件型或是單件的文物，則其 Metadata 大多是以「件」為單位；大多數情形是一筆 Metadata 對應一個數位物件。古文書與手稿資料則亦有一筆 Metadata 對應多個影

幅之情形，但是，基本狀況是針對「件」為單位建置 Metadata。然而各筆 Metadata 的內容，同一欄位的資訊相異度高，不易共用欄位內的資訊。

### 2. 雜誌書籍

數位化的資料對象如果是雜誌書籍，則建置 Metadata 可能是以雜誌中的「篇」、書冊中的「章節」為單位。而同一本雜誌及同一本書籍中的各筆 Metadata，其可以共用欄位資訊的欄位較多，例如，書刊名、出版社、出版年、書籍的作者或編者、雜誌的卷期、雜誌的專欄名稱等；換言之，針對同一本雜誌或書籍建置 Metadata 時，各筆 Metadata 資料中，真正不同而需個別鍵入資料的，只有篇（章節）名、作者、頁碼等少數欄位而已。因此，在規劃設計建檔軟體時，可以考量如何充分利用到共用的欄位，以提高建檔效率。

## (四) 檔案命名

任何型式的資料之數位化，在進行相關規劃之初，即必須考量如何使數位影像、數位全文、詮釋資料 (Metadata) 等各類資料能夠在最少人工的情況下，順利互相關聯。臺灣大學圖書館進行各種類型資料之數位化計畫，主要都是透過檔名的規劃，使未來資料庫系統中的 Metadata 能夠順利連結到相對應的數位影像檔與全文檔。檔名規劃的基本原則是使影像數位檔、全文數位檔、Metadata 編號等各類數位檔案之主檔名一致，只有副檔名不同 (Metadata 編號只有基本檔號，無副檔名)：

### 1. 文件型或單件文物

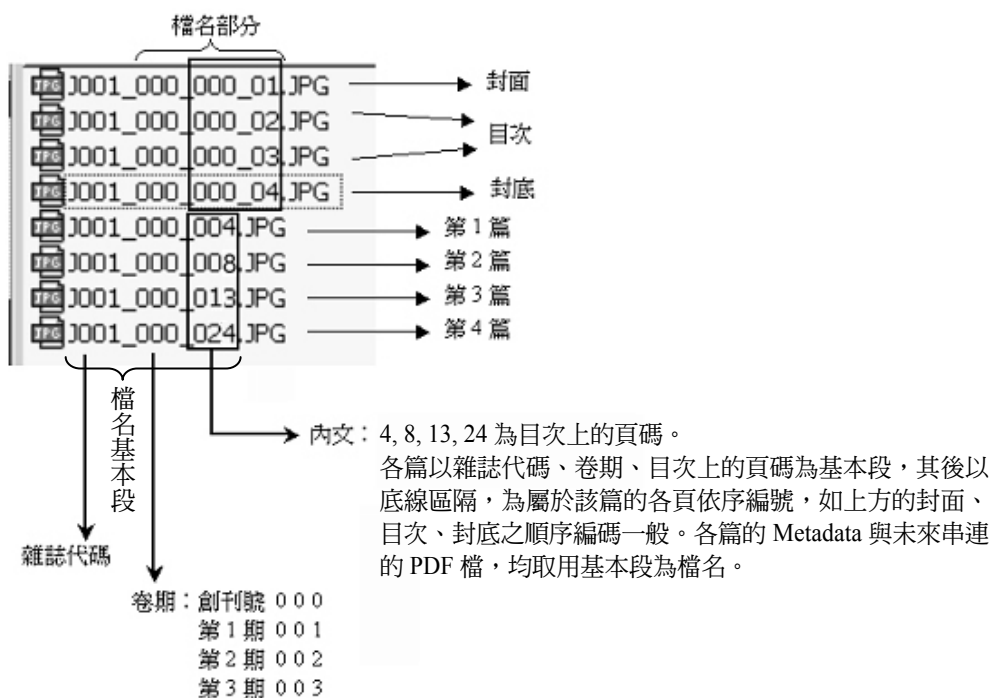
以資料原有的編號作為基本段，同一文件如因包含附件、浮貼、文件過大而分區數位化，因



此產出多影幅，則這些隸屬於同一文件的所有影幅，都是同一基本段之下作檔名的延伸，進行檔案的命名存檔。如此則不論是程式或人工，都可辨識屬於同一文件的所有影幅。

## 2. 雜誌書籍

雜誌書籍之數位化，因為也需數位化其封面、封底、目次，又希望未來以自動排序整理數位檔案時，各雜誌之封面、目次、內文的各篇能夠依序排列。因此，臺灣大學圖書館將雜誌進行數位化時，單篇全文影像之存檔檔名及串連一文多頁的 PDF 檔之存檔檔名如(圖二)之圖示說明：串連一文多頁時，以及為各筆 Metadata 編號時，均根據目次各篇起始頁碼作為檔名的尾數，如此則兩方面的工作成果即可容易地互相關連。



圖二、單篇全文影像之存檔檔名與串連為一文多頁的 PDF 檔之檔名

## (五) 浮水印

### 1. 文件型或單件文物

以臺灣大學圖書館執行古文書類的數位化計畫為例，數位化的成果包括典藏級的 Tiff 格式檔案、商務應用級的 Jpeg 格式檔案以及網路傳輸用的 Jpeg 格式檔案，但只在網路傳輸用的檔案上嵌入浮水印。因為此類的數位影像檔並未進行一文多頁的串連，所以浮水印是在各幅影像檔一一嵌入浮水印。

### 2. 雜誌書籍

數位化紙本雜誌並轉檔為單頁的 PDF 格式檔案之後，單頁影幅之檔案在未嵌入浮水印之前的大小約為 90KB；如於單頁嵌入浮水印，則每頁的



檔案大小約增加至略多於 200KB (〔圖三〕之例為 216KB)。串連一文多頁之後再嵌入浮水印，可以使平均每頁檔案大小約為 150KB。串連越多頁的情形，平均每頁的大小有降低的趨勢。因此，先串連檔案再嵌入浮水印，較為理想。

ntul-jn247_0018_000_01_2頁	325 KB
ntul-jn247_0018_001_6頁	856 KB
ntul-jn247_0018_007_1頁	216 KB
ntul-jn247_0018_008_6頁	927 KB

圖三、串連一文多頁並加浮水印之後的檔案大小之例

#### 四、結論

典藏數位化的工作，在國內推行已約有十年的歷史，在此十年間，數位化的設備有長足的進步。影像數位化的速率、規格等，隨著設備的更新而提升；相對的，亦可減低操作數位化機器的人力成本支出。另一方面，因為需要執行數位化的資料，益趨多元化，有的機構基於推廣館藏利用的目的，需要將大量的雜誌書刊進行數位化。此種情形之數位化，需要從設備與數位化規格方面作調整，以及將工作流程改善至最佳化，使製作成本合理化。因此，計算合理的數位化單價時，已不只是從文件大小張進行考量，而應根據現行可用的機器之數位化效率，從日產能估算廠商所需花費的人力成本，據以評估合理的委外數位化單價。

本文以國立臺灣大學圖書館進行雜誌類資料數位化之經驗為例，說明如何在影像檔規格、數位化設備、檔案命名、Metadata 建檔、浮水印嵌入等規格與流程作調整異動，以產出最佳效果。不過，在要求製作價格合理化的同時，也必須為品質嚴格

把關。當放寬對機器型式的規定時，仍必須維持影像規格及符合各種規範的水準；以嚴格的招標規範徵募廠商，於招標文件中要求廠商於投標時提供各種檔案格式之數位檔及其列印樣品，或要求廠商於決標後短時間內提供此等樣品供審核，且於招標文件中規定審核通過之樣品列為未來驗收成果之品質標準依據。如此，則可能在最優惠的製作單價之條件下，獲致高水準之成果。

後記：本文為根據作者在 2006 年 11 月 15 日於國立臺中圖書館之「宗教史料數位典藏研習會」之「數位典藏技術與實務」經驗分享內容，擇要撰述而成。

