

國家圖書館 期刊報紙典藏數位化計畫 數位化工作流程圖文說明

最後修訂日期：2004/01

計畫簡介：

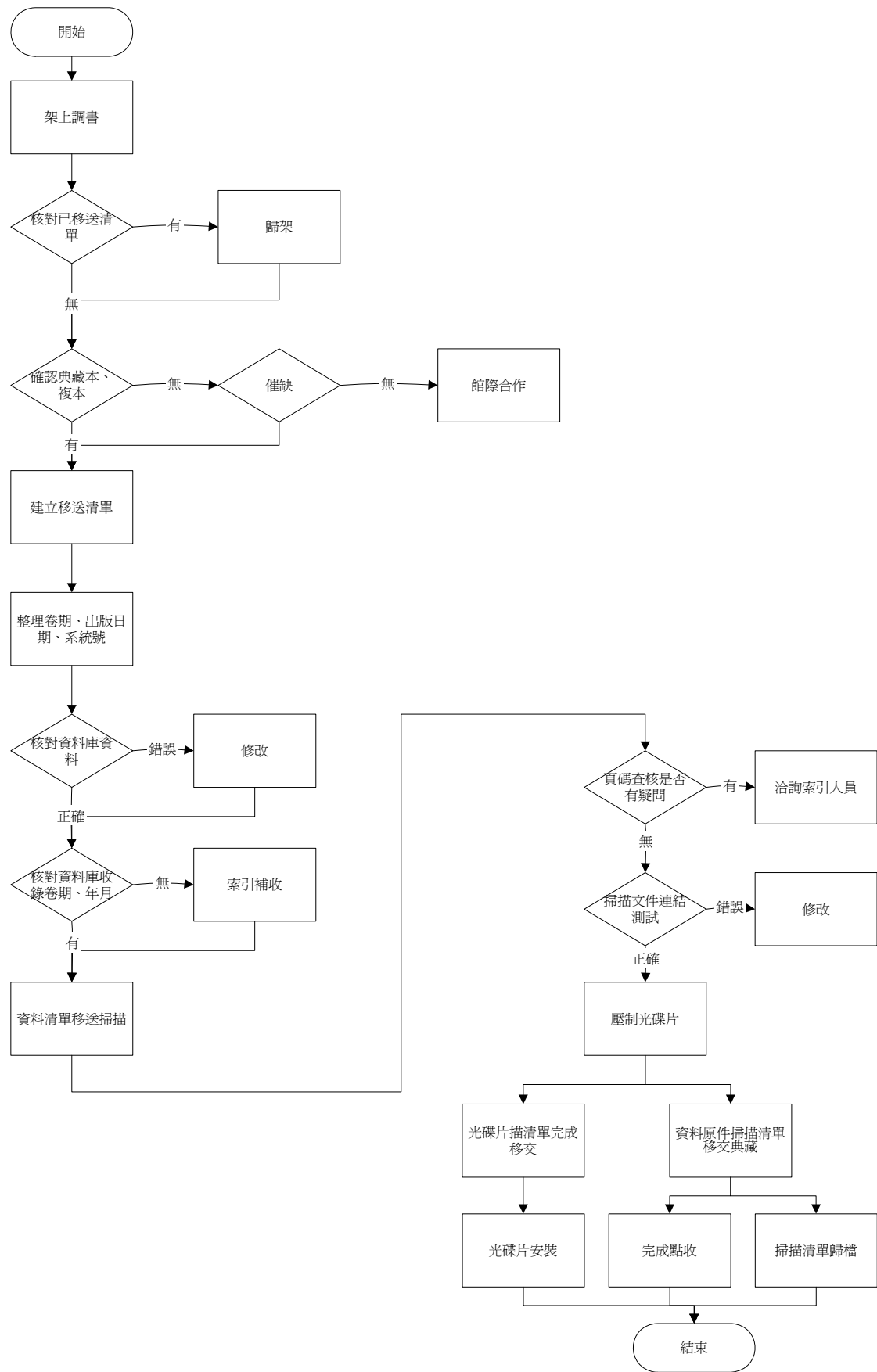
期刊報紙是一般訊息傳遞、研究發展之成果發表最快速的園地，也是記錄當代人文、社會、科技等活動最重要的媒體，讀者使用量極高，但保存不易，使用不便，因此極需數位化以改善。

此項計畫擬在五年內完成國家圖書館（以下簡稱國圖）館藏臺灣地區發行之期刊約 1,000 種、臺灣地區發行之報紙約 30 種之數位化。充實國家圖書館期刊報紙影像資料庫，便利民眾取閱，並得以數位方式永久保存報刊資料，傳之久遠，以備學術文化資產之永久典藏，並期能藉助資訊網路科技之進步，提供使用者方便快捷的文獻傳遞服務，達成無牆圖書館之理想。

九十二年度此項計畫將加強報紙掃描工作數量，預計進行國圖館藏 200 種期刊與 6 種報紙(含微縮片轉製)掃描，共約 116 萬頁期刊及報紙之全文影像數位化。工作項目包括期刊典藏影像數位化、報紙典藏數位化、期刊主題分析三項作業。

數位化工作流程簡介：

本計畫的數位化作業，大約分為下列七大項步驟施行，依序為：數位典藏前置作業、訂定相關標準與規範、詮釋資料分析與著錄、進行資料數位化與品管、數位化資料的儲存與管理、智慧財產權管理、數位化資料的利用。



圖一、國家圖書館數位化影像掃描流程圖

一、數位典藏前置作業

在數位典藏的前置作業中，必須先訂定標準與規範、選擇數位化內容、確立施作方式等細項工作規範。就主題而言，各種數位圖書館、數位博物館或文化資料庫所包括的資料類型可能有：音樂、戲劇、舞蹈、美術、文學、電影、建築、傳統文化等各種主題；然而就媒體類型而言，則可分為：文字、影像、聲音、視訊等四種型態。由於原始資料之來源不同，因此也需要不同的數位化方式。

各種資料數位化時，必須考量到使用者的設備、使用的便利性、資訊檢索的需求、網路上資料的傳輸速度、資料的永久保存等問題，分別建立幾個不同使用目的的檔案。各種資料數位化的處理原則、檔案格式及驗收規範，如下：

（一）確立檔案格式

依據媒體資料的不同，所制定的檔案格式亦不同。本計畫為符合檔案命名原則以及詮釋資料格式，總共區分為文字檔及影像檔兩種檔案格式（表一：數位化檔案格式）。

1. 文字檔：

若原始資料是以電腦打字的電子檔，除存一份原始檔外，建議轉成 HTML、PDF 或 RTF 三種檔案格式。

原始資料為手稿或印刷資料，如果要做全文檢索，則需重新打字，其餘處理方式同上；若不做全文檢索，則只須建立詮釋資料，並將原件掃描，並以不壓縮格式，儲存一份永久檔，再視須求轉成其他目的之格式，如網路下載格式及預覽格式。

2. 影像檔：

若原始資料為照片、圖片、地圖等，則需以掃描器掃成影像檔，並以不壓縮格式儲存一份永久檔，再視需求以永久檔轉存成其他目的之格式，網路下載格式及預覽格式。

（二）數位化物件選擇

確立檔案格式之後，隨即進行選擇數位化的物件的作業，先區分資料來源，而後依照典藏目的與使用目的選擇數位化物件。因此依據資料之來源不同、典藏目的與使用目的，也需要不同的數位化方式。

(三) 確立施作方式

各種資料數位化須考量到使用者的設備、使用的便利性、資訊檢索的需求、網路上資料的傳輸速度、資料的永久保存等問題，以分別建立幾個不同使用目的的檔案。以確立各種資料數位化的處理原則、檔案格式及驗收規範。而在實際施行數位化工作時，依據工作內容等項目，區分為兩種方式，即自行製作以及委外作業（依照「九十一年度國家圖書館期刊數位化影像掃描暨回復作業招標規範」、「國家圖書館報紙數位化影像掃描招標規範」、「掃描工作流程表」…等工作文件來執行）進行施作。

檔案格式	建議規格	說明
文字檔		
資料永久保存格式	檔案格式：TIFF 壓縮：不壓縮 色調深度：黑白；灰階-每像素 8-bits；彩色-每像素 24-bits 解析度: 300~600 或更高 dpi（依原始資料品質及重要性選擇適當解析度，一般印刷品可採 300dpi）	將資料數位化典藏，保持原有風貌。提供使用者作為重製、壓縮處理或其他圖像處理交換之用
網路下載格式	檔案格式：JBIG or JBIG2 色調深度：黑白；灰階-每像素 8-bits；彩色-每像素 24-bits 壓縮：JPEG(灰階壓縮比約 10:1，彩色壓縮比約 10:1) 解析度: 150dpi~300 dpi 或影像大小：從 500x400 至 1000x700 pixels	提供使用者網路上觀看及列印用
預覽影像	檔案格式：GIF 色調深度：每像素 8-bits 壓縮：原生影像至 GIF 解析度或影像大小: 72dpi 或影像大小從 150x100 到 200 x 200 pixels	提供使用者預覽及選擇欄位用
影像檔		
資料永久保存格式	檔案格式：TIFF 色調深度：灰階-每像素 8-bits；彩色-每像素 32-bits 壓縮：不壓縮色彩濃度值 4.0D 以上(color)，3.2D(B&W) 解析度: 300~600 或更高 dpi（依原始資料品質及重要性選擇適當解析度，一般印刷品可採 300dpi，美術品供複製畫使用建議採 600dpi，供印刷出版使用採 350dpi）	將資料數位化典藏，保持原有風貌。提供使用者作為重製、壓縮處理或其他圖像處理交換之用。

檔案格式	建議規格	說明
資料服務 / 參考格式	檔案格式：JFIF(JPEG 交換格式) 色調深度：灰階-每像素 8-bits；彩色-每像素 24-bits 壓縮：JPEG(灰階壓縮比約 10:1，彩色壓縮比約 20:1) 解析度: 150dpi~300 dpi 或影像大小：從 500x400 至 1000x700 pixels	提供使用者網路上觀看及列印用
縮圖影像	檔案格式：GIF 色調深度：每像素 8-bits 壓縮：原生影像至 GIF 解析度或影像大小: 72dpi 或影像大小從 150x100 到 200 x 200 pixels	提供使用者預覽及選擇欄位用

表一、數位化檔案格式

二、訂定相關標準與規範

邀請館外圖書資訊界專家十餘人以及協同國家圖書館館內同仁(特藏組、資訊組、參考組、政府出版品部門)，成立「文獻分析機讀格式計畫小組」，共同修訂期刊文獻資源建檔之 metadata；並參與館內之「自動化及網路諮詢委員會」、「中國機讀編目格式計畫小組會議」、「詮釋資料(metadata)格式標準研訂小組會議」、「資料數位化標準研訂小組會議」、「分散檢索標準研訂小組會議」五項會議，一起訂定數位化作業的相關標準與規範。

訂定的標準與規範是本計畫數位化作業人員在執行時所遵循的準繩，可以幫助各項作業有所依據，包括有：

(一) metadata：文獻分析機讀格式。



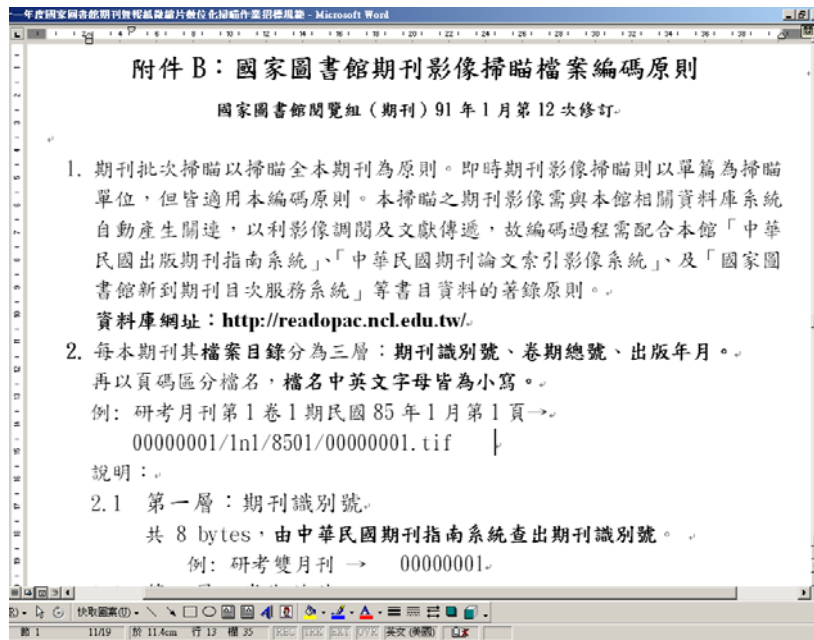
圖二、文獻分析機讀格式

(二) 產出規格：

1. 掃描影像檔：Tiff 檔。
2. 影像製作規格：300 dpi G4 壓縮 Tiff 檔或 PDF 檔。在原稿清楚的前提下，影像掃描品質力求清晰，內容完整。部分藝術類雜誌掃描成全彩影像，早期報紙掃描成灰階影像。
3. 已電子化之雜誌：以 HTML 或 SGML，XML 格式儲存。
4. 微縮捲片：35mm。

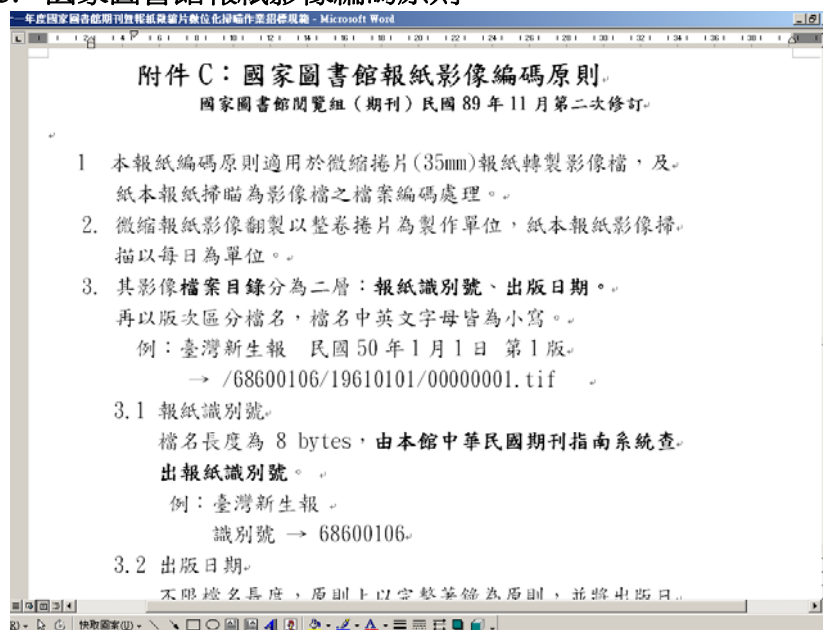
(三) 作業規範：

1. 資料數位化標準—檔案數位化與命名原則（見圖二）。
2. 國家圖書館期刊影像編碼原則。



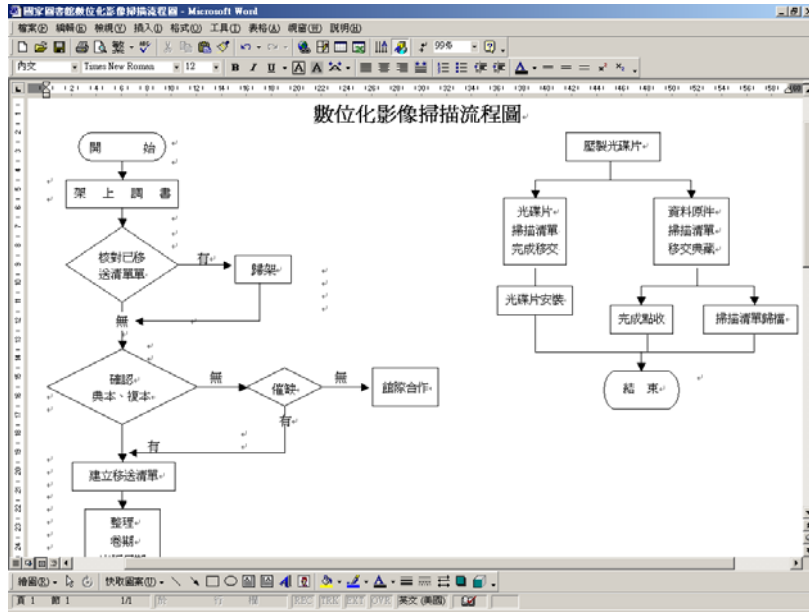
圖三、國家圖書館期刊影像編碼原則

3. 國家圖書館報紙影像編碼原則。



圖四、國家圖書館報紙影像編碼原則

(四) 工作程序：國家圖書館數位化影像掃描流程圖。



圖五、國家圖書館數位化影像掃描流程圖

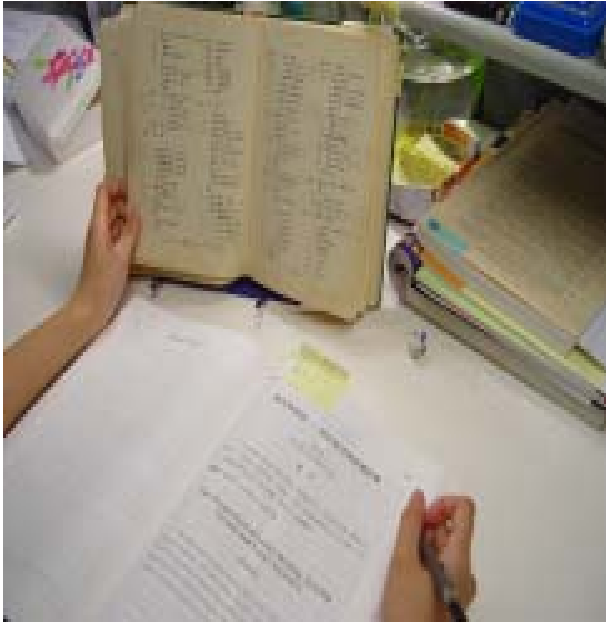
三、詮釋資料分析與著錄

詮釋資料分析與著錄部分，包括索引編選、建檔、校訂等作業，甚至部分採委外處理，另外再加 7 名工作人員支援處理。此一步驟所依循的標準包括：

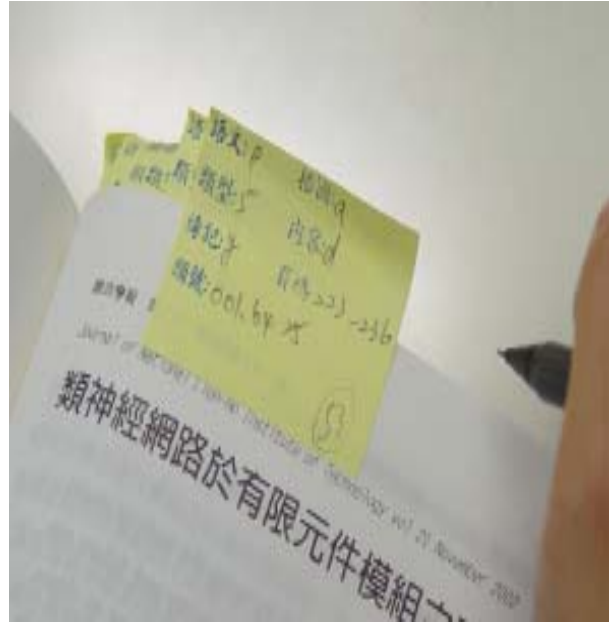
1. 資料數位化標準—檔案數位化與命名原則草案（見圖二）。
2. 國家圖書館期刊影像編碼原則（見圖三）。
3. 國家圖書館報紙影像編碼原則（見圖四）。



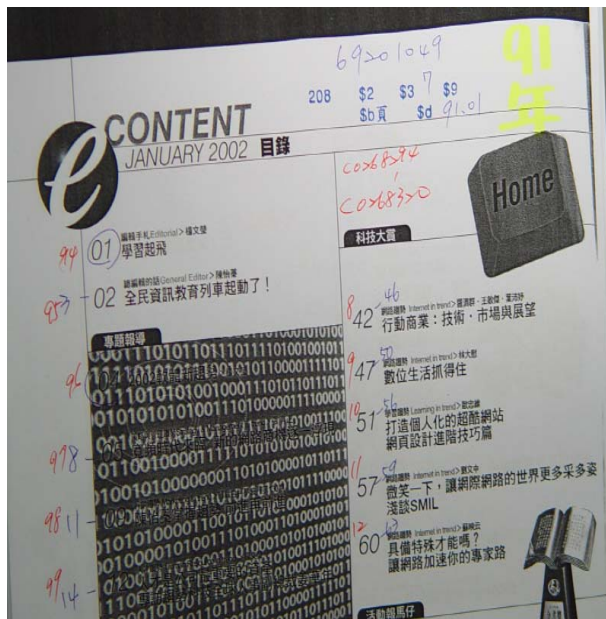
圖六、區分目次、開架、直接掃描期刊三種類別進行數位化作業



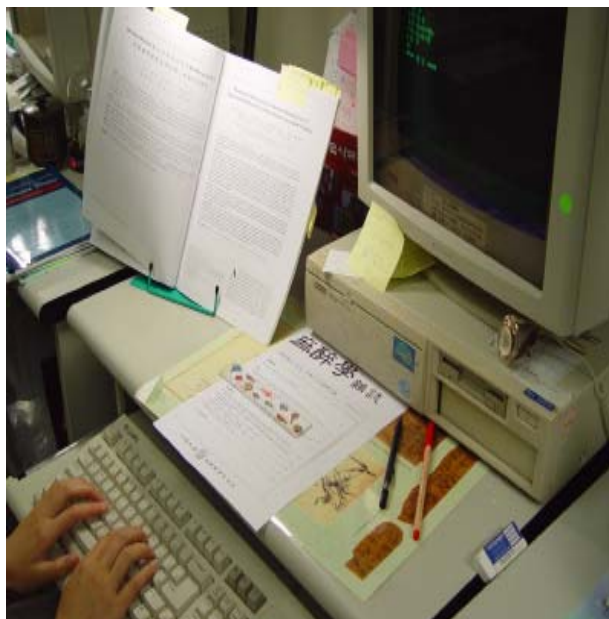
圖七、給予期刊篇章分類號與標題



圖八、期刊篇章之註明標籤



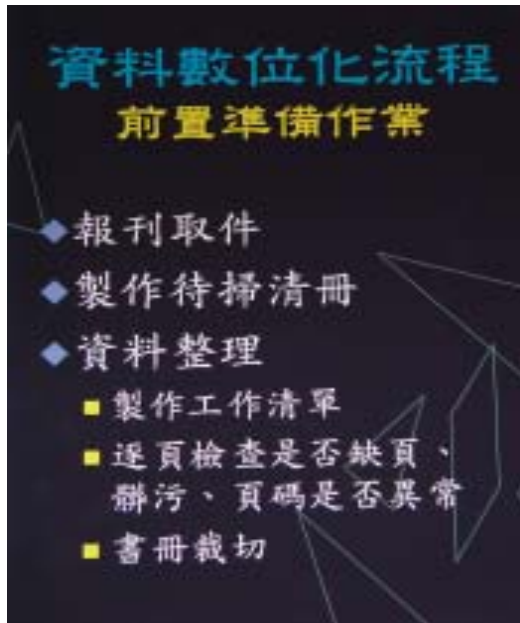
圖九、期刊目次之分類註明



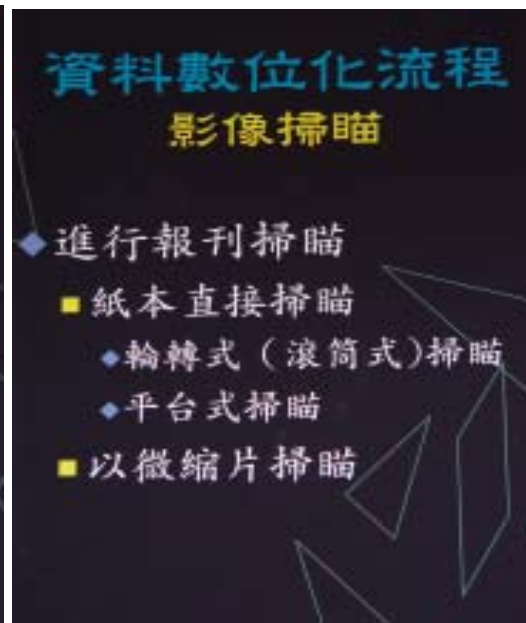
圖十、輸入期刊目次

四、進行資料數位化與品管

待標準與規範皆訂定好之後，開始進行資料掃描作。資料數位化與品管的進行步驟包括：取件、製作掃描清單、送件、資料整理、進行掃描、品質檢驗、數位成果產出、資料驗收、影像儲存以及索引更新等 10 項子步驟。數位化影像掃描流程概要，如下：



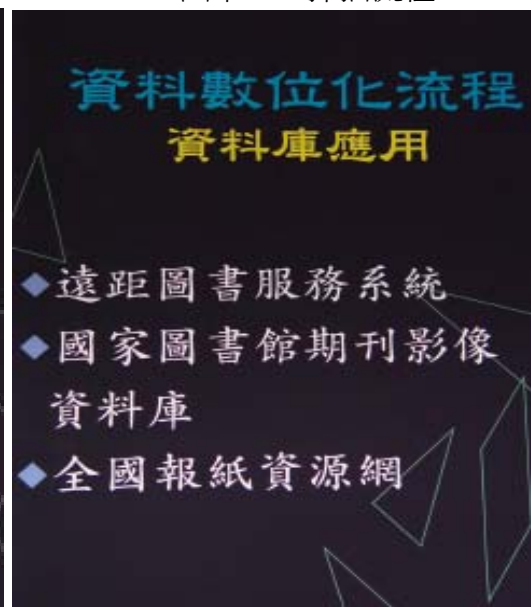
圖十一、掃描流程 1



圖十二、掃描流程 2



圖十三、掃描流程 3



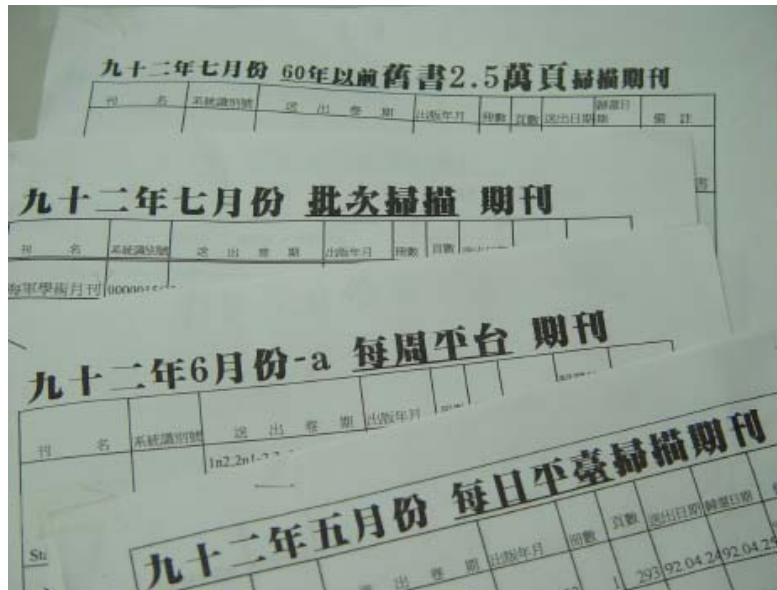
圖十四、掃描流程 4

（一）取件

館內工作人員從國家圖書館的封閉式書架上「直接」調出欲進行掃描數位化作業的期刊、報紙或微縮片，先內部進行『核對已移送清單』，若已經移送隨即『歸架』；若尚未移送則進行確認書架上是否有『典藏本與複本』，若無則進行『催缺』動作，若已絕版或無法取得，則與他館進行『館際合作』，盡可能在數位化執行期間取得該件期刊或報紙。

(二) 製作待掃描清單

取件後根據狀況挑選可進行掃描數位化之期刊與報紙微縮片。首先進行『整理卷期、出版日期、系統號』，之後再『核對資料庫資料』中之卷期、出版日期、系統號，若有錯誤之處，即進行『修改』，直至正確後開始製作『平台待掃描期刊清單』以及『待製影像報紙微縮捲片清單』，完成『製作待掃描清單』作業。



圖十八、依據掃描類別製作待掃描清單

(三) 送件

根據之前完成正確無誤之『待掃描清單』，由期刊或報紙掃描工作人員將清單，移送給委外廠商（駐國圖館內），準備進行下一步影像掃描數位化作業。

(四) 資料整理

此作業由國圖根據招標規範與契約書委外處理。委外廠商之期刊或報紙掃描工作人員(委外 5 名人員)先進行『資料整理』，包括『批次待掃描期刊』整理、製作『工作清單』（逐頁檢查有否缺頁、髒污或頁碼異常）、進行『書冊裁切』…等作業。



圖十九、依據送交清單檢查期刊並註明頁數



圖二十、去釘



圖二十一、手動裁切過厚期刊



圖二十二、裁切器裁切適當厚度期刊

(五) 進行掃描

此作業由國圖根據招標規範與契約書委外處理。期刊或報紙掃描工作人員（委外 10 名人員）依照招標規範與契約書之工作項目、作業規範與作業程序進行掃描數位化。



圖二十三、平台式掃描設備



圖二十四、平台式掃描

(六) 報紙微縮片轉製

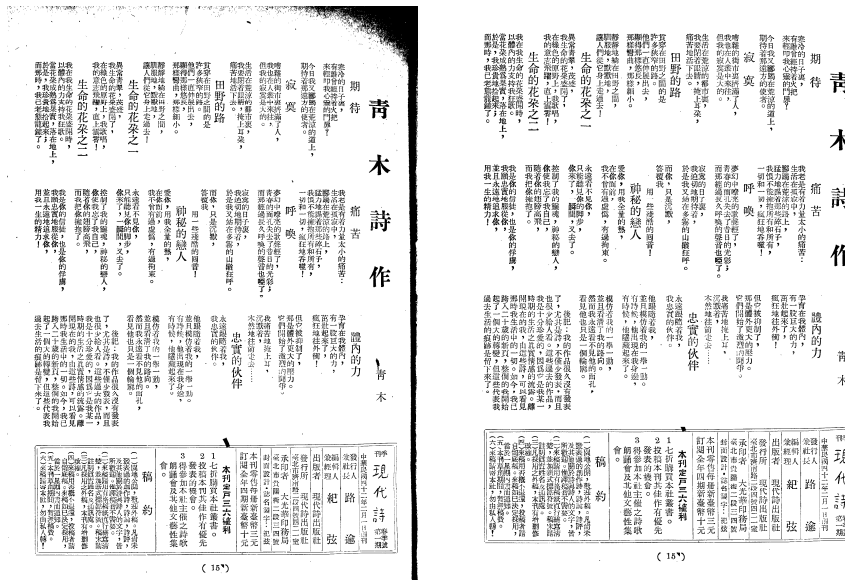
此作業由國圖根據招標規範與契約書委外處理。報紙微縮片轉製承包商，依照招標規範與契約書之工作項目、作業規範與作業程序進行掃描數位化。



圖二十五、委外掃描報紙微縮片轉製圖

(七) 品質檢驗

國家圖書館根據招標規範與契約書交予廠商處理。委外廠商應逐頁檢查掃描後的文件是否有太濃、太淡、黑邊污點、歪斜或是影像內容不合格…等情形發生。檢驗出不佳品質之數位化影像檔後，則須進行修改至合乎品質規格要求。



圖二十六、影像校正（修正前1張／修正後1張）

(八) 數位成果產出

國家圖書館根據招標規範與契約書交予廠商處理。委外廠商將通過品質檢驗之數位化物件，以『壓縮成光碟片』的方式完成數位成果呈現。並製作「光碟片掃描清單」、「資料原件掃描清單」連同數位成果移交國圖辦理『資料驗收』。



圖二十七、成果產出

（九）資料驗收

於國圖館內自行施作。由國圖工作人員按月以抽驗或整批檢驗…等方式，依據「年度即時期刊影像掃描量統計表」、「期刊影像掃描檔案編碼原則」、「報紙影像編碼原則」、「期刊影像掃描暨報紙微縮資料轉製影像規格」、「掃描期刊回復裝訂規格」、「掃描期刊移送清單」、「掃描期刊資料整理紀錄單」、「掃描期刊問題整理紀錄單」、「期刊影像掃描交驗清單」、「掃描期刊回覆裝訂清單」、「待製影像報紙微縮捲片清單」…等清單，配合合約之規格以及項目，對委外廠商製作完成之數位成品，進行影像品質之『資料驗收』。所製作之光碟片與微縮捲片凡有品質不佳或內容錯誤等情形發生，皆依照招標規範、契約書辦理之。

（十）影像儲存

於國圖館內自行施作。將通過驗收之光碟片等數位成果，轉入資料庫，資料原件回復裝訂，微縮捲片歸回庫房，各種清單歸檔。待所有的成果皆完成點交後，光碟片則準備移交典藏，由國圖購買光碟櫃儲存數位成果資料，另為便於提供服務，甚至購買影像伺服器儲存完成之數位化資料。

（十一）索引更新

於國圖館內自行施作。由國圖工作人員將完成點收之數位化成果，進行光碟安裝與轉入資料庫，更新索引檔資料，供使用者利用。

五、數位化資料的儲存與管理

數位化資料的儲存與管理部分，由於數位化的格式種類多，在儲存空間上必須加大與更多的經費，依據不同的目的，儲存與備份設備包括：TAPE、DVD、CD-R、磁碟陣列、光碟櫃…等多種形式。

本項作業由業務單位及資訊組人員 2 名，依循上述硬體處理數位化資料的儲存與管理。



圖二十八、磁碟陣列



圖二十九、光碟櫃

六、智慧財產權管理

將已經完成的數位化成品，進行使用權限的管理。在資料進行數位加值與利用時，尚未取得著作人授權的文獻，遭遇著作權法的限制，無法提供線上瀏覽或列印，頗多不便，需要積極爭取授權。另外一方面，目前數位成品的浮水印部分，尚未製作相關加密處理的作業。

七、數位化資料的利用

國家圖書館事先擬定相關規格與招標規範，以進行公開招標委外處理，並且按月檢驗品質辦理驗收，而後再轉入資料庫中，提供使用者利用。由國圖系統推廣營運（委外），利用 Sun5000 硬體，建置資料庫、提供館內讀者線上檢索、結合電子商務。但遭遇著作權法之限制，尚未取得著作人授權之文獻，無法提供使用者利用。目前數位成果包括以下三種資料庫：

(一) 期刊影像資料庫檢索與應用：<http://readopac.ncl.edu.tw/periodical/hypage.cgi>



圖三十、期刊影像資料庫檢索與應用

(二) 報紙影像資料庫檢索與應用：http://192.192.58.101/cgi/ncl9/m_ncl9_news



圖三十一、報紙影像資料庫檢索與應用

(三) 遠距圖書服務系統檢索與應用：<http://readopac.ncl.edu.tw/>



圖三十二、遠距圖書服務系統檢索與應用

以此三種資料庫來作系統推廣營運，包括資料庫的建置、提供館內讀者線上檢索以電子商務的結合…等作業。但是在進行中，仍是以著作權法的關係，限制文獻的使用，因此需取得著作權人的授權為重要的作業。如此才能將成品呈現給使用者來進行利用，達成資料交換與推廣利用的目的。

※以上數位化工作流程簡介之文字，主要是參考「國家圖書館期刊報紙數位典藏工作流程調查表」、【國家數位典藏國家型計畫中程綱要計畫書】。

※製作單位：數位典藏國家型科技計畫－內容發展分項計畫
國家圖書館期刊報紙典藏數位化計畫

※文字撰寫：國家圖書館期刊報紙典藏數位化計畫 編輯 林淑芬小姐
數位典藏國家型科技計畫－內容發展分項計畫
新聞主題小組助理 汪怡慧

※圖文編輯：數位典藏國家型科技計畫－內容發展分項計畫
新聞主題小組助理 汪怡慧、曾欣怡

※圖片拍攝：數位典藏國家型科技計畫－內容發展分項計畫
新聞主題小組助理 汪怡慧、陳美智

※感謝：國家圖書館「期刊報紙典藏數位化計劃」之計畫主持人俞小明主任、聯絡人林淑芬編輯撥冗指導及提供實地拍攝與簡介編寫。並感謝國家圖書館等相關計畫人員的協助。