

第十二章

檔案數位影像製作之流程與管理

Document Digitalization: Procedures and Management

莊樹華

Shu- Hwa Chuang

中央研究院近代史研究所

檔案館秘書

Secretary of the Archives of the Institute of
Modern History, Academia Sinica

張凱達

Kai-Ta Chang

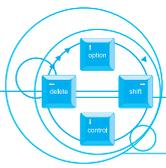
中央研究院近代史研究所

檔案館資訊管理員

T. Manager of the Archives of the Institute of
Modern History, Academia Sinica

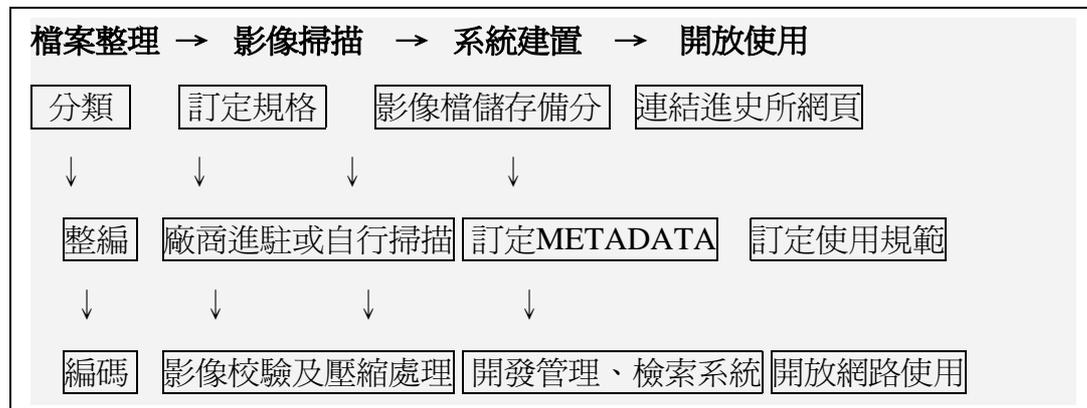
壹、前言

檔案數位典藏計畫其作業包括檔案整理、影像製作、系統建置及開放使用等部分。而其製作核心在於資料庫的建置及數位化影像製作兩大部分，二者結合為影像資料庫。資料庫的建置包括著錄欄位的訂定、Metadata設計、建檔、校對以及資料庫的加值如權威控制及權威檔的建立。影像的製作包括影像掃描及備份兩大部分。本文所討論的數位化影像製作之管理，主要針對影像數位化製作之規格的訂定及流程的管理。



影像數位化模式有幾種選擇：1、將檔案的資訊以人工方式重新輸入，此種方式雖有助於檔案資訊內容的全文檢索，但無法反應檔案文字資訊外的其他資訊如印章、貼條。2、將檔案掃描再以OCR辨識。目前OCR辨識的有效性以排版字為主，而檔案多屬手稿，除少數附件為規格性的打字稿外，不宜使用OCR辨識。3、將檔案掃描以圖檔方式儲存。（註1）這是目前各界處理檔案數位化的主要方法，唯圖檔無法作全文檢索，因此聯結影像的資料庫就成為提供檢索重要資訊來源。

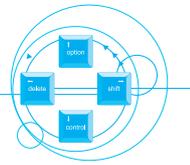
檔案掃描方式有以饋紙式的快速掃描機或以逐頁操作的平台掃描機，前者可用於年代較近、紙質較佳的公文性資料，年代久遠的歷史檔案則必須使用平台式的掃描機。本文所討論的檔案影像數位化即為利用平台式掃描所製作存檔的歷史檔案影像。



貳、影像規格的製訂

影像數位化是件投所謂數位化影像規資浩大的工程，其作業模式有如大量製造的生產線，若不能於開動生產線前審慎確認產品的規格，待量產一段時間後想要再變更，將造成人力物力的重大損失。

所謂數位影像化規格包括影像儲存格式及影像解析度。影像儲存格式一般使用的以TIFF或JPEG兩種格式；解析度係指在一英吋內以多少點來表現一個影像，而組成這個影像的點則稱之為像素（Pixel）。影像規格的訂定關係經費成本，規格愈高，意謂著所掃描影像的操作時間、品檢、校驗的電子檔讀取、處理時間，



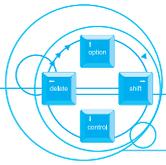
傳輸、備份的等待、錄製時間等均隨著影像規格提高而成倍數增長。反映在影像製作成本，每頁的差異可達數元至數十元，累乘上龐大的檔案數量，將產生數十萬甚至百萬元的成本差異。如何在有限的資源下，完成既定的產量又達到相當品質，評估適當的影像規格，成為典藏單位進行數位化前最重要的工作。

數位化影像可分典藏保存類型與提供網路瀏覽、下載的一般使用類型，後者可直接由典藏保存類型轉換，不需重製，因此在選擇數位化影像規格時，可直接考慮典藏保存之基本要求，以此作為數位化時的參數依據，以避免重製的風險。

（註2）

數位化影像規格的訂定除了考慮高品質的典藏需求外，另一個考量為使用目的。使用的影像規格愈高，單位成本自然愈高，在有限的經費下，所能完成的數量相對較少。數位化的標的物若是器物類，每一件器物、書畫、古籍、標本等均以單品的型式呈現其高度典藏及利用價值。而數位化的標的物若是檔案，則須考慮檔案的數量及完整性。檔案的保存必須尊重全宗原則，即同一個全宗檔案是不允許被拆散的，由於一份全宗的檔案可從數萬張至數十萬張，數量十分龐大，單獨完成任何一部分的數位化，是無助於檔案的利用。因此檔案的數位化必須同時考量品質與產量。若採用高品質因而受限於經費無法完成整批全宗檔案，亦無法達到數位化的目的。如在1998年中央研究院計算中心負責規劃「日治時期台灣總督府公文類纂及專賣局檔案」的數位影像製作時，各單位（史語所內閣大庫檔案、近史所戰後台灣經濟發展相關檔案）所採用的影像規格均為黑白300DPI/TIFF，平均一張價格為5元。若是A3大小之手稿以人工的方式進行全彩掃描，當時市價，每頁掃描單價從30元至50元不等，「臺灣總督府公文類纂」及「臺灣總督府專賣局檔案」之典藏數量約計七百萬頁，如何以有限之預算，兼顧數位化之品質及數量，以便於日後資料運用之可行性，為製訂影像規格之考量。

由於數位化影像製作耗時費力，基於成本及避免對檔案造成二度傷害考量，中央研究院計算中心（以下簡稱計中）以黑白與全彩實作方式測試，發現黑白影像無法清楚顯現檔案中關防印信，確定黑白影像不可行後只能就全彩規格考量。計中分別以100/150/200/300DPI進行測試，結果以150點解析度的JPEG影像格式製



作儲存，檔案影像之清晰度與檔案大小為最佳組合。由於當時國內尚未有大量檔案彩色影像外製之經驗，市場價格偏高，於是計中透過自訂的掃描規範及工作流程，現場模擬作業，估算作業成本，計算出精確的單價，協助廠商降低製作成本。最後以A3每頁4.9元價格完成影像招標。150點解析度全彩的JPEG影像格式也成為日後其他單位進行大量影像製作的規格。

以目前影像製作技術也許會認為150點解析度全彩的JPEG影像格式標準過低，但在1998年，國內尚未有大量全彩數位化影像製作經驗的環境下，由執行單位自行研發出操作模式，協助廠商降低成本，採取全彩的掃描規格，實屬一個成功的案例。

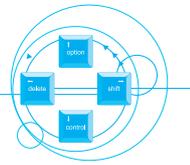
訂定數位化影像規格另一個考量因素為影像的大小，若原件愈小，採用的解析度愈高，影像品質愈清楚。如一張3×5英吋的照片可以用600DPI以上的解析度，底片則可採1500DPI，其影像大小約15MB。一般檔案大小尺寸從A4到A3不等，而其價值在於文本中的文字內容，影像規格的訂定所須考量的是檔案資訊的完整性，而不是影像的精密度。因此檔案影像解析度的訂定，無論從實體影像的大小及影像精密度而言，均與照片、底片的考量不同，反須把握以不流失檔案中任何訊息為原則。

此外，數位影像具有增值或再處理的可能性，為了避免未來因不同需求再進行影像重製，規劃影像規格時除了考量檔案的清晰度外，必須充分思考產出的數位影像規格未來是否可提供增值處理的空間，例如印出、出版、文字辨識、影像接圖處理等。

總之，影像儲存成本逐年有明顯降低的趨勢，影像規格的選擇必須考量其品質的永久性與未來的增值可能。

參、備份的選擇

影像規格選定後，直接影響到儲存備份。影像規格愈高，儲存空間愈大，全彩24bit、300DPI的解析度的TIFF存檔格式，平均一張A3大小的影像約50MB，不適



於用一般640MB的光碟片作備份。檔案的影像數量龐大，必須多元性考量備份的型式。一分完整數位影像的備份容量為：媒體單位容量×數量×備份份數。選擇何種載體作為備份，除了考量備份規格是否標準化外，載體的容量無疑是最主要的考量因素。

數位資料的備份載體常見的有CD光碟、DVD光碟、磁帶、硬碟、磁碟陣列。
表一為各類備份媒體性質比較：

表一 常見備份媒體性質比較

性質 媒體	單位容量 (GB)	容量比	耗材單價 (元)	讀寫器	讀寫器單價 (元)	備份時間 (分)
CD光碟	0.7	1	10	CD燒錄器	3,000	5
DVD光碟	4.7	7	100	DVD燒錄器	15,000	30
磁帶	100	143	3,200	LTO磁帶機	130,000	120
硬碟	120	171	5,500	-	-	120
磁碟陣列	>1,000	> 1,429	-	250,000	-	-

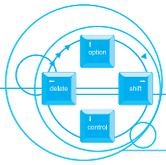
說明：至2003年第1季，磁帶已有單卷200GB，硬碟已有單個200GB容量的產品問世。耗材及讀寫器價格以2002年第4季市場價格為準。

若以300DPI全彩A3影像存為TIFF檔20萬頁估算，將產生10,000GB（10TB）的影像資料，以此進行鉅量備份的總成本估算如表二：

表二 不同媒體備份 10TB成本估算

性質 媒體	備份時間 (小時)	耗材數	耗材成本	器材成本 (元)	備份成本 (元)
CD光碟	1190	14,285	142,850	3,500	146,350
DVD光碟	1,064	2,128	212,800	15,000	227,800
磁帶	200	100	320,000	130,000	450,000
硬碟	180	90	495,000		495,000
磁碟陣列				2,500,000	2,500,000

比較上表可以看出，一般常用的CD光碟無疑是所有備份媒體中成本最低的。不過必須指出，除了耗材及設備成本外，也應同時考量人力成本。備份作業的人力需求由於實際作業模式不同而難以估算，譬如CD及DVD光碟燒錄器由於單價較



便宜，可以規劃一人操作多台燒錄機方式進行錄製作業。磁帶及硬碟備份方式由於操作時間較長，工作人員可在兼顧其他業務情況下同步作業，但大致而言，備份人力與備份所需時間是成正比。再者，不同載體單位容量、保存條件、錄製及讀取時間、資料格式世代相容性、讀寫機具設備價格及普遍性等性質各有利弊情況下，備份媒體的選擇無法僅止於成本考量。

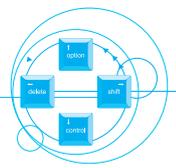
一般最被廣泛使用的備份媒體為光碟及磁帶。CD光碟具有成本最低、相容性最高等優點，但單位容量低且錄製時間長。此外，備份完成的大量光碟片一旦需要全盤讀出，將耗費非常多的時間及人力進行讀取動作。（註3）

DVD光碟較CD光碟耗材成本高，製作備份時間相當。由於單位容量為CD的7倍，故可節省存放空間及讀寫作業的抽換片動作，然而最大的問題是現階段DVD光碟尚未形成統一的工業標準，可能產生讀取資料時因不同廠牌而有讀寫器格式相容性的顧慮。

磁帶也有數種規格，不同規格的磁帶各自使用不相容的磁帶機，且各廠牌均不斷在容量和速度上競相成長，單卷容量已可達200GB，製作備份時間則接近硬碟。相較於光碟媒體，磁帶容量大，具有節省儲存空間、便於管理及高速錄製、讀取的優勢。但由於技術新穎，價格相對較其他載體為高，至於磁帶保存期限則尚難確知。

就各項儲存設備言，磁碟陣列無疑是成本最高，因此大型容量的備份採用磁碟陣列進行鉅量電子檔案備份的可能性極低。採用硬碟進行備份亦不理想，一方面成本過高，一方面硬碟的電氣特性是否適宜長期保存尚不確知。

若歸納不同備份媒材的差異性可以看出，單位容量愈大的設備價錢愈高，備份時間愈短；單位容量愈小的媒材價錢愈低，備份時間愈長。就保存風險而言，單位容量愈高的媒材，每個單位一旦損壞則流失的資料愈多，風險也就愈高。就保存期限而言，無論是光碟、磁帶、硬碟等備份媒體的保存年限至今尚未有絕對的數據，因此，選用愈新的設備或耗材，風險自然就愈大。



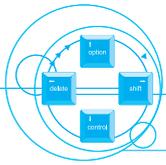
肆、影像製作

檔案由於數量十分龐大，數位化影像大都採委外製作。委外方式或採單純以價格高低定案的價格標。若為防止廠商以低價競標，承包後卻無法履行合約或達到品質要求的情形發生，可採「技術標案」，由執行單位事先訂定包含建議書審核、簡報及答詢、實作程序等客觀的評分標準，邀請檔案界及資訊界的專家擔任評審委員，以遴選適當的合作廠商。惟廠商所提供的掃描機具品質和價格差異性極大，實作測試時由於機具新，操作數量少，無法看出長期大量掃描後機具的耐用度及品質穩定度，因此，在正式招標前，執行單位必須對市面上各類型掃描機具的種類、價格與性能作一了解，提供評審參考，否則短時間的技術評審無法作出客觀之評估。

廠商建議書中會提出人力及機具配置，必須留意的是使用的人力多、機具多，雖反應出廠商的投資成本，但機具性能的好壞、操作人員的素質以及廠商的管理才是掃描工作能否順利進行之關鍵，否則人多、機具多徒增場地空間的浪費以及管理的困難。

檔案一旦進入掃描階段，從掃描、品檢到備份，就如同量產的生產線，每天有固定的產量，任何一個環結若銜接不好，將影響產量；反之只顧大量生產，沒做好品質控管，則生產得愈多，製作出有問題的影像也就愈多。因此，如何做好影像製作過程時量與質的控管，是數位化計畫成敗重要的關鍵。

影像製作的流程，從掃描到備份過程如下：影像掃描、影像裁切、影像校正、備份。為了節省作業時間，廠商以一人操作多機的配置方式進行掃描，掃描時通常會先預掃，同時校正色彩，掃描存檔後再由後端人員負責裁切。裁切人員須先拉框再將影像四邊的空白切勻，由於一個人平均一天的掃描量約2,000頁，裁切人員一天也須負責近2,000頁的數量，才能使流程順利進行。為避免工作人員誤裁影像內容，最好規定檔案周圍須留0.1公分的空白。若顧慮四周留白，導致掃描時影像容易歪斜，可容許檔案的一邊靠齊掃描機，其餘三邊於拉邊裁切時均須保留空白處，以示資訊的完整性。

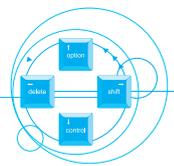


伍、大於A3尺寸的檔案

一般平台掃描器皆為A3尺寸，而檔案的尺寸大都在A3以下，因此影像的掃描發包都以A3以下尺寸為單位。若檔案的長度大於A3，如奏摺，則可以A3長度為標準分段掃描。分段掃描時每一張影像的末段與下一張影像的首行必須重覆掃描，除可避免分段時不小心造成資訊的流失外，若未來有接圖的必要，才可作後續處理。另一種大尺寸的文件可能長寬均大於A3，如地圖，若採切割掃描，無法完整顯示檔案資訊，必須使用更大尺寸的掃描器進行數位化。此類文件無法規格化處理，因此必須另訂特殊文件的製作管理規範，如抽出的文件如何標記？如何掃描？影像檔如何歸位？而影像的完整性關係著影像資料的備份時間，因此特殊文件在檔案群中固然屬少數，但處理時反費時費功，而也因為文件數量少，在大量生產流程中，若管理不好反亦容易掉失。

國內目前處理大量大於A3檔案文件數位化單位應屬國史館台灣文獻館的台灣總督府及專賣局檔案。該檔案小於A3的部分已於文獻館與中央研究院於1997年至2002年五年半的合作計畫，完成了約計六百七十餘萬張的影像，惟大於A3檔案，文獻館另申請國家數位典藏計畫單獨執行。文獻館對於大尺寸文件的作法是將大於A3文件抽出，依大小再分為四類，每一類給予9001~9004的代號，若文件號001000010010001的下一頁為A3以上至A2之大圖文件，它的編號就成為0010000100190010002m。（註4）此種命名方式必須假設文件上的編碼即是影像檔的編碼，而非原文件上由人工所作的編碼，因為手工編頁難免有誤，影像頁碼才是將來聯結資料庫系統的依據。由於大圖製作時間與原檔案掃描分開處理，且時間間隔太長，期間原檔案經由校正或亦更動過影像檔名，而大圖檔名若未立即隨文件檔名更正調整其原編的號碼，將無法作為歸位的依據。而且影像頁數是由機器自動產生的流水號，若未預留圖檔的影像編號，從系統中無法得知大圖的影像位置。未來如何將這六萬多張的大圖歸回六百七十多萬張檔案中正確的位置，勢必會是一個難題。

近史所檔案館對於大尺寸規格特殊文件的處理，於整編檔案時即將大於A3文



件抽出，為使掃描時影像流水號不因抽出一頁文件而受影響，必須於抽出文件的位置放入一張註記“A0”的白紙，若抽出文件為兩張，則放入兩張註記“A0”的白紙，依續編上流水號，並將抽出的檔案登錄在清單上。掃描時必須重新核對每一頁的影像頁次及原文件編號，務求二者一致。掃描時遇註記“A0”的白紙依舊掃描，並於該檔名後加“！”，以便在系統中作為辨識之用。大圖以大型掃描機掃描後，從影像目錄檔找出有“！”的檔名，即可將大圖檔影像插入，圖文一旦完整，即可進行磁帶備份。以下為近史所農商部的影像檔清單，從檔名可一目瞭然大於A3文件的位置。

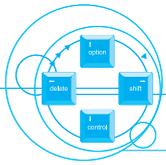
大於A3的特殊文件處理，可依各單位文件的性質及管理的方便需求，訂定不同的處理方式，但必須把握的原則是，除了考慮如何將大圖等特殊文件抽出單獨處理，還要考慮如何歸回正確的影像位置，而且處理時間愈短，成本愈低，正確性也愈高。

表三農商部影像檔屬性清單

檔名 (File Name)	存檔格式 (Format)	色 階 (bits per pixel)	檔案長度 (Bytes)	像 素 (width x height)	解析度 (dpi)	頁數
082400008010001.jpg	JPEG	24bpp	749,871.0	2076× 3 132	200 × 200	1
082400008010002.jpg	JPEG	24bpp	837,471.0	2098× 3 109	200 × 200	1
082400008010003.jpg	JPEG	24bpp	729,897.0	2091× 3 129	200 × 200	1
082400008010004.jpg	JPEG	24bpp	839,969.0	2075× 3 134	200 × 200	1
082400008010005 !.jpg	JPEG	24bpp	195,620.0	1680× 2 346	200 × 200	1
082400008010006.jpg	JPEG	24bpp	830,301.0	2075× 3 134	200 × 200	1
082400008010007.jpg	JPEG	24bpp	745,531.0	2075× 3 134	200 × 200	1
082400008010008.jpg	JPEG	24bpp	847,571.0	2075× 3 134	200 × 200	1
082400008010009.jpg	JPEG	24bpp	744,108.0	2075× 3 134	200 × 200	1
082400008010010.jpg	JPEG	24bpp	880,313.0	2075× 3 134	200 × 200	1

陸、影像校驗

影像校驗分為廠商自行作初步校正及執行單位的檢驗。兩階段檢驗均無誤後，才能作備份。影像檢驗項目包括影像的歪斜、清晰度、色彩、檔名等。在招標規範書上會訂定影像可容許的歪斜度，而一般廠商在大量製作的經驗下，對於影像的歪斜、清晰度均有相當程度的掌握，而歪斜與清晰度可由肉眼判斷，爭議



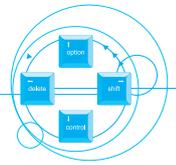
性較低。檔名由手工輸入，難免有誤，且不易校正；影像的色差除非很明顯，否則不易由肉眼辨別；至於影像品質參數，則要由影像處理軟體逐筆校正。以下分別說明之。

系統中影像的取得是靠影像檔名的聯結，影像檔的命名是影像製作過程中除了掃描外最重要的步驟。由於每一筆文件均由人工建置，除事先規範檔名的屬性如檔名的層次、碼數、純文字或文數字，最好能設計檢查程式，由程式自動排除錯誤的命名，如碼數錯誤、屬性錯誤時則無法完成建檔，以降低檔名的錯誤率。即使如此，亦難免手民之誤，因此檢驗者必須再逐件核對檔名，並檢查影像資料是否放於正確之資料夾。

檢驗過程中爭議性最大的為色彩，單張影像無法看出色彩的正確度，大量影像同時校正時便可看出影像的色差。色差的產生可能來自多部掃描機有多種色調，也可能來自於同一台掃描機不同時間所產生的掃描色差。

由於掃描時多機同時進行，為避免不同的掃描機掃出不同的色彩，必須定時校正不同掃描機的掃描色度是否一致。此外，同一台機器可能因量產之故，色彩銳度逐漸降低，因此，除了各機之間要作統一的校色外，每一台機器也須定時作校色。校正方式為掃描色彩導表，一來比較不同機器的色度是否一致，二來每一台機器所掃描的色表須與前次所掃描的色表作比較，以看出同一台機器前後的色差度。

影像經過歪斜、清晰、色彩、檔名等校正後，必須再以影像處理程式檢查影像縮放比、影像壓縮比、影像解析度。在招標規範書中除了訂定影像儲存規格如TIFF或JPEG等外，會另訂影像解析度、影像縮放比、影像壓縮比，而相關參數廠商在掃描前於每一台機器亦作好設定，惟掃描者有時不諳操作，更動了設定值而不自覺，影像解析度、縮放比、壓縮比是肉眼檢查不出來，必須將所有影像放入硬碟由影像處理程式檢查每一個影像的屬性資料，包括檔名、存檔格式、色階、影像大小、解析度等。以中央研究院近代史研究所檔案館農商部檔案為例，其掃描規格為全彩、200DPI/JPEG存儲格式，經由影像處理程式產生每一筆屬性資料如



下：

082400008010001.JPG.....Ok

file format = JPEG File Format,

file length = 749871 bytes;

width = 2076 pixels,

height = 3132 pixels

color depth = 24 bpp;

image resolution: horizontal =200 dpi,

vertical = 200 dpi;

pages number = 1

Processing.....Ok

082400008010002.JPG.....Ok

file format = JPEG File Format,

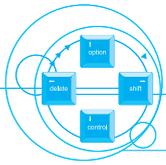
file length = 837471 bytes;

width = 2098 pixels,

height = 3109 pixels

color depth = 24 bpp;

image resolution: horizontal = 200 dpi,



vertical = 200 dpi;

pages number = 1

Processing.....Ok

082400008010003.JPG.....Ok

file format = JPEG File Format,

file length = 729897 bytes;

width = 2091 pixels,

height = 3129 pixels

color depth = 24 bpp;

image resolution: horizontal = 200dpi,

vertical =200 dpi;

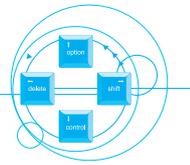
pages number = 1

Processing.....Ok

再將以上資料轉成Excel則可得出的一覽無遺的表列：

表四 農商部影像檔屬性清單

檔名 (File Name)	存檔格式 (Format)	色階 (bits per pixel)	檔案長度 (Bytes)	像素 (width x height)	解析度 (dpi)	頁數
082400008010001.jpg	JPEG	24bpp	749,871.0	2076x 3 132	200 x 200	1
082400008010002.jpg	JPEG	24bpp	837,471.0	2098x 3 109	200 x 200	1
082400008010003.jpg	JPEG	24bpp	729,897.0	2091x 3 129	200 x 200	1
082400008010004.jpg	JPEG	24bpp	839,969.0	2075x 3 134	200 x 200	1
082400008010005.jpg	JPEG	24bpp	737,251.0	2075x 3 134	200 x 200	1
082400008010006.jpg	JPEG	24bpp	830,301.0	2075x 3 134	200 x 200	1
082400008010007.jpg	JPEG	24bpp	745,531.0	2075x 3 134	200 x 200	1
082400008010008.jpg	JPEG	24bpp	847,571.0	2075x 3 134	200 x 200	1
082400008010009.jpg	JPEG	24bpp	744,108.0	2075x 3 134	200 x 200	1
0824000080100010.jpg	JPEG	24bpp	880,313.0	2075x 3 134	200 x 200	1



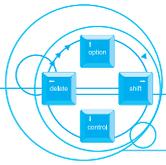
以上為利用影像程式協助進行影像規格參數：存檔格式、色階、長度、像素、解析度、一檔單頁或一檔多頁等資訊的檢驗。至於原件的尺寸與若干圖檔格式的壓縮比率是否正確，則可依據影像程式查核的結果換算得知。規格要項的檢查雖可事後由適當的軟體取得精細至每一個影像的資訊，加以比對分析，但費功耗時。若檢核出的問題無法後製處理時，則需重新掃描，所以廠商清楚掌握數位影像品質要項的標準，在掃描過程中訂定嚴謹的作業規範，並能確實執行，讓每一頁掃描的影像能按既定的標準流程產出，才是最正確有效的作法。

柒、備份的製作

備份的製作，可由廠商製作，也可由廠商交付硬碟由執行單位自行製作。一旦備份必須透過網路進行遠端備份作業時，應與儲存伺服器管理者事前清楚規劃每日傳輸的時段、資料量、備份媒體等，才不致於影像備份的流程導致硬碟塞爆的現象。

影像的備份，須同時考慮保存與利用。一張A4尺寸的檔案原件，經300DPI/24bit的規格掃描存成不壓縮的TIFF檔後，電子檔約為25MB，若為A3尺寸則約50MB。一批十萬頁的檔案照上述規格掃描後，將產生2,500至5,000GB的電子檔，除需解決備份媒體與容量成本的難題外，如何進一步處理如此龐大的影像檔也需先行規劃。較好的辦法是在掃描時同步進行壓縮轉檔，並儲存成另一份壓縮版的備份，未來可直接以此份壓縮版做為上線或再處理之用。換言之，除了保存一份不失真的原始影像檔外，為方便後置的影像處理或傳輸，會再另存一份downsize或壓縮檔，以降低儲存容量。如原始TIFF檔轉存一份JPEG格式的備份時，要訂定多少downsize比例的影像或壓縮參數，必須兼顧影像在網路的傳輸速度與呈現品質。

影像的儲存無論採用何種載體，基本上同性質檔案必須錄製在同一份載體上，即同一冊文件若可錄於同一份備份上，則必須燒錄在同一份備份上，同一冊的影像容量若超過備份容量，必須燒錄在二份的光碟片上時，在光碟編號上須註明1/2及2/2，以利光碟的管理。



至於備份品實體管理，除了每一份備份要有標籤註明該備份的編號及備份內容外，必須列有備份清單。每一個影像的備份型式資訊亦須著錄在該文件的 Metadata 備份欄位中。至於備份實體的保存必須符合該備份媒材的典藏環境與方式。

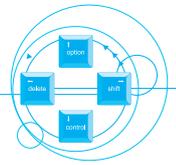
捌、影像的驗收

除了每日例行性影像檢驗外，由於影像大都為委外廠商製作，因此必須定期驗收。驗收時除了核對影像報表與影像總量外，亦須作影像品質的抽驗。影像儲存載體若為光碟、DVD，隨機抽驗讀取較不是問題；影像儲存載體若為磁帶，則必須先作好事前準備工作。由於一卷100GB的磁帶讀取時間為約120分，檢驗人員隨機抽驗，影像必分散在不同磁帶中，若不事先將驗收影像讀錄出來，檢驗人員必須等候每一卷磁帶的讀取時間，十分耗時。因此最好事先準備驗收清冊，由驗收人員隨機勾選影像，由相關人員將每一筆抽檢影像先從磁帶錄製出來，作為檢驗之用，如此才可以節省驗收時間。

玖、圖文比對

數位化製程中由於專業分工的需要，目錄與影像往往採取分別製作的方式產生。數位影像的命名原則通常由館藏號而來，故檔案目錄中的館藏號資訊一方面是影像掃描製作的依據，另一方面則是影像備份及系統整合時的校正依據。

進行圖文比對時首先須訂定比對的基準，通常以檔案的最小項目為單位。數位化的全宗若以「件」為單位，每一筆目錄即為一件檔案，掃描存檔後應產生一筆相應的影像檔。同理，數位化的全宗若以「冊」為單位，每一筆目錄即為一冊檔案，掃描存檔後應產生一筆該冊的影像檔。我們可自實體檔案的電子目錄中擷取出館藏號欄位，製成「應掃檔案文號索引」，再自掃描後的影像檔名中產生「掃描檔案文號索引」，兩個索引檔的產生方式如下表：



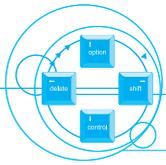
表五 目錄、影像圖文比對產出表

1.目錄中的館藏號欄位：						I .掃描後產出之影像檔 (\表資料夾)：
全宗	系列	宗號	冊號	件號	...	\03-32-001-01-001
03	32	001	01	001	...	03-32-001-01-001-001.JPG
03	32	001	01	002	...	03-32-001-01-001-002.JPG
03	32	001	01	003	...	03-32-001-01-001-003.JPG
2.修改格式制成「應掃檔案文號索引」：						\03-32-001-01-002
03-32-001-01-001						03-32-001-01-002-001.JPG
03-32-001-01-002						03-32-001-01-002-002.JPG
03-32-001-01-003						\03-32-001-01-003
						03-32-001-01-003-001.JPG
						03-32-001-01-003-002.JPG
						03-32-001-01-003-003.JPG
						2.選用件層級製作「掃描檔案文號索引」：
						03-32-001-01-001
						03-32-001-01-002
						03-32-001-01-003

將兩個索引檔利用ACCESS或EXCEL等資料庫管理程式，進行交叉比對，即可得出文字目錄與掃描影像的吻合狀況。若為「有目無圖」則須檢查原檔是否為缺頁或破損未掃，且是否已清楚註記於備註欄；若確定為漏掃漏編，則立即進行補掃作業。若為「有圖無目」則檢查是否為文字目錄漏建或打字錯誤。圖文比對的過程凡有修改文字目錄或影像檔名的情況，應於修改後再重複進行，直至比對無誤為止。比對後則應清楚記錄結果，以為資料庫檢索系統上線運作的檢核依據，整體缺漏情況尚可於使用範例中說明，以提供讀者參考。

拾、結語

將檔案數位化典藏除了可節省備份的保存空間，在利用上經由電腦檢索系統的調閱讀取，不受空間及時間的影響，使用上較調實體檔案方便。然備份保存的安全機制，除了備份實體的妥善管理外，必須將備份作異地的儲存，以防天災人禍。異地的安全距離為多少，並無標準數據，惟目前數位典藏國家型科技計畫亦考量國際備份的可能性。數位化典藏產生巨量的數位影像的存儲、安全及管理，



將是另一個專業且複雜的領域。此外，數位典藏另一個重點為網路的開放利用。數位化檔案於網路開放使用前，可於影像加嵌浮水印，以確保於網路上傳播之電子檔具可追蹤、認證的特性。其次開放後必須對使用群資料的使用行為作追蹤調查，作為管理改善資料庫的依據。

數位典藏從檢索資料庫的建置、影像製作到查詢系統每一個環結環環相扣，尤其檔案數量龐大，作業中除了要確保影像的品質外，更要做好影像系統與資料庫聯結的正確性，否則大量製作出一堆數位影像，一旦進系統將影響影像的有效聯結率，屆時於數十萬張或數百萬張的影像中去查核、修正，只會徒增資料庫修正及影像重製備份的成本。因此，數位化影像製作過程從掃描到備份均要做好每一個流程的管理與品檢，其過程雖可借助電腦等機具，但操作人員及管理人員之素質才是整個數位影像品質好壞之關鍵。

【原刊載於檔案管理局出版之檔案季刊（九十二年三月）第二卷第一期所撰】

註釋：

註1：肖俊。「檔案的數字化處理」，海峽兩岸檔案暨縮微學術交流會論文集。臺北，中國檔案學會編印，頁24，民國91年。

註2：「數位典藏國家型科技計畫」，數位典藏技術彙編，2002年版第五冊，頁2-12，民國91年9月。

註3：實務上一個工讀生搭配一組配備4台高速CD-ROM的電腦進行讀取，一天約可讀出150片(以650MB計)相當於100GB的資料，讀出10,000GB資料約需單人100個工作天。

註4：蕭明治。「國史館臺灣文獻館日治時期檔案大圖掃描實務概述」，國家數位典藏通訊第二卷第一期，民國92年1月1日。