

目次

壹、簡介.....	1
貳、轉置格式說明.....	2
參、轉置品質驗證方法說明.....	3
一、結構相似性品質指標 SSIM (Structural SIMilarity) ..	3
二、峰值訊噪比 PSNR(Peak Signal to Noise Ratio)	4
肆、轉置程序.....	5
伍、DOCX 格式轉置為 PDF/A 格式.....	6
一、轉置目的.....	6
二、轉置測試環境.....	11
三、檔案轉置結果.....	11
四、遭遇問題.....	18
陸、POSTSCRIPT 格式轉置為 PDF/A 格式.....	19
一、轉置目的.....	19
二、轉置測試環境.....	20
三、檔案轉置結果.....	20
四、遭遇問題.....	24
柒、機關實作電子檔案格式轉置	25
一、DOC 格式轉置為 PDF/A 格式	25
二、WDL 格式轉置為 PDF 格式	35
捌、問題檢討	39
一、DOCX 格式轉置 PDF/A 格式	39
二、POSTSCRIPT 格式轉置 PDF/A 格式	39
三、中央印製廠 DOC 格式轉置 PDF/A 格式.....	40
四、國立師範大學 DOC 格式轉置 PDF/A 格式.....	40
五、財政部臺灣省中區國稅局 WDL 格式轉置 PDF 格式.....	40

壹、簡介

近年來由於電子化政府之推動，機關電子檔案之類型與數量均呈現急遽成長，且因資訊技術日新月異，電子檔案之儲存格式與儲存媒體，幾乎是以 3 至 5 年為一個世代快速轉變。因此，如何確保被保存之電子檔案能持續被存取與利用，已成為一個非常重要之議題。目前已有許多長期保存方法與策略被廣泛地探討，但仍以轉置、封裝、模擬與系統保存等為主要及常見之保存策略，其中又以轉置為最常使用之保存策略。然而電子檔案每經過 1 次轉置都有可能造成資訊的流失，甚至可能是極為珍貴的資料。因此，本報告彙整本年度開發整合之轉置工具與驗證方式，並針對機關試作過程中所面臨問題，提出檢討與解決方式。

貳、轉置格式說明

本專案 98 年度已擇定 3 種類型的電子檔案格式進行轉置作業，其轉置格式如下：影像檔案轉置(TIFF to JPEG)、可攜式文字檔案轉置(WDL to PDF)、視訊檔案轉置(MPEG-2 to H.264)，相關轉置工具業已整合至電子檔案長期運作平台。

99 年度之轉置作業為文字類型轉置(DOC to PDF/A & DOC to ODT)、視訊類型轉置(WMV to MPEG2)及電子檔案封裝檔格式轉置(舊版轉為新版)。

本專案 100 年度已擇定 2 種類型的電子檔案格式進行轉置作業，其轉置格式如下文字類型轉置(DOCX to PDF/A 及 POSTSCRIPT to PDF/A)，將相關轉置工具整合至電子檔案保存工具。

為了驗測電子檔案格式轉置品質，並確認工具符合機關需求，由檔案管理局擇定 3 個機關進行機關電子檔案試轉作業，分別為國立師範大學(DOC 格式轉置為 PDF/A 格式)、中央印製廠(DOC 格式轉置為 PDF/A 格式)及財政部台灣省中區國稅局(WDL 格式轉置為 PDF 格式)。

本年度派員至國立師範大學、中央印製廠及財政部台灣省中區國稅局安裝電子檔案保存工具，進行格式轉置與驗證作業，並已依機關試作結果修正工具。

參、轉置品質驗證方法說明

電子檔案保存工具，針對文字檔類型格式轉置作業之品質驗證方式，係將文字類型 (WDL、DOCX、PS 及 DOC 格式)之來源檔案及轉置後檔案(PDF/A-1b 格式)，分別轉置為 JPEG 檔，再將兩個 JPEG 檔案進行影像結構相似性指標評估法 (Structure Similarity, SSIM)及峰值訊噪比(Pear signal-to-noise ratio, PSNR)檢測轉置前後的檔案品質。

一、結構相似性品質指標 SSIM (Structural SIMilarity)

結構相似性指標是用於測量在兩幅影像之間相似性的一種方法。SSIM 指標係改進以前提出之通用性影像品質指標 (UQI) 模型，而可以被視為一個完美之影像品質量測。

作為結構相似性理論之實現，結構相似度指數從圖像組成的角度將結構信息定義為獨立於亮度、對比度及反映場景中物體結構的屬性，並將失真建模為亮度、對比度和結構三個不同因素的組合。用均值作為亮度的估計，標準差作為對比度的估計，共變異數作為結構相似程度的度量。

例如 $SSIM=1$ 就是和原始視訊檔案完全一致，若 $SSIM \geq 0.98$ 就是難以與原始視訊檔案分辨區別，而 $SSIM=0.95$ 的時候大多數人都會對畫面滿意，這個數值可以認為是及格的畫面。 $SSIM=0.90$ 的時候意味著瑕疵可能要比 0.95 多一倍，肉眼就會察覺到明顯

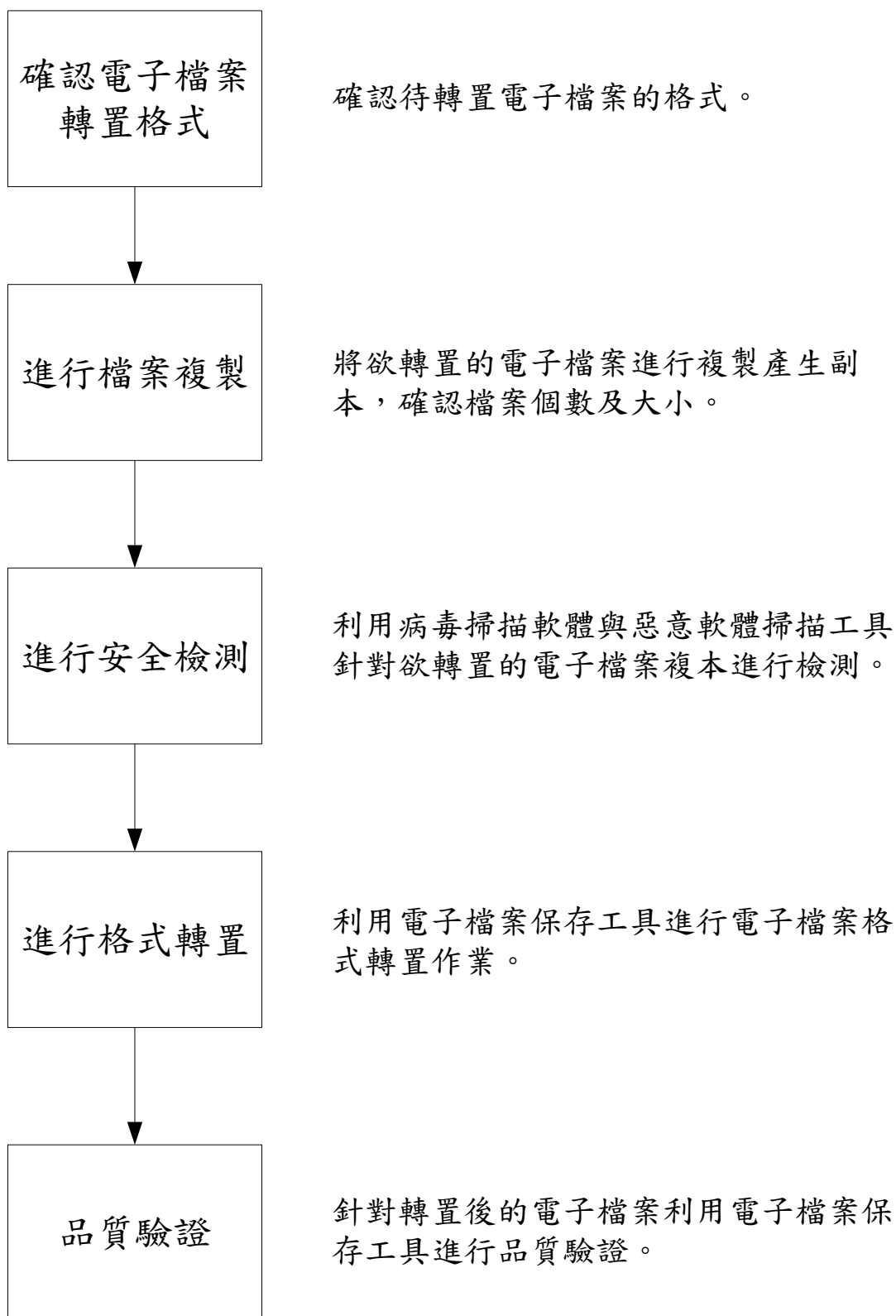
的畫面劣化，因此低於這個數值時，就可以判定為沒有實際觀賞價值。

二、峰值訊噪比 PSNR(Peak Signal to Noise Ratio)

峰值訊噪比經常用作圖像壓縮等領域中信號重建質量的測量方法，主要是利用影像信號的最大值與影像中雜訊的比值作為評估的標準。對於每點有的 RGB 三個值的彩色圖像來說峰值訊噪比的定義類似，只是均方差是所有方差之和除以圖像尺寸再除以 3。

通常 PSNR 值越高表示品質越好，一般而言當 $PSNR < 30\text{db}$ 時，代表以人的肉眼看起來是不能容忍的範圍。因此大部分 PSNR 值皆要 $> 30\text{db}$ 。但 PSNR 高，並不代表影像品質一定好，有時候還是必須靠人的肉眼輔助來判斷影像的品質才較為正確。

肆、轉置程序



伍、DOCX 格式轉置為 PDF/A 格式

一、轉置目的

Microsoft Office 目前是被廣為接受之辦公室軟體，許多電子公文檔案與附件皆以 Microsoft Office 之檔案格式儲存(DOC、DOCX、XLS 及 PPTX 等)，但由於 Microsoft Office 檔案格式並沒有對外公開，在歷次 Microsoft Office 改版過程中也曾經發現過各版本呈現內容不一致之情況，因此該格式並不適合做為長期保存格式。

經本專案評估後，建議文字檔類型採用 PDF/A 作為電子檔案格式轉置之目的格式，PDF 歸檔有下列幾個優點：

(一)PDF 儲存的是結構化之物件

結構化的物件(例：文字，向量圖形，點陣圖形)，這允許 PDF 檔案可以進行全文檢索，相對之下，TIFF 檔案格式保存的檔案如果要達到相同的功能，則在可以被搜尋之前必須先經過文字辨識(OCR)，但文字辨識結果並非百分之百，人為的校正是必要的，針對大量的檔案來說，這樣的工作量是相當可觀的。

(二)PDF 檔案的大小

通常比 TIFF 檔案來得小，也通常提供較好之呈現品質，較小的檔案在電子傳遞上相當具有優勢。

(三)PDF 檔案裡可以儲存各式的詮釋資料

其詮釋資料(標題、作者、建立日期、修改日期、主旨、關鍵字等等)，這些資料將使得 PDF 檔案能夠自動化分類，完全無需人為介入。

(四)PDF 檔案之內容通常可以與裝置無關

PDF 檔案可以不依賴特定之解析度與色彩系統，這使得日後的輸出不會因為新的輸出設備不支援舊式規格而產生問題。

PDF 檔案格式同時也是國際上民間或政府單位廣泛被接受的標準電子文件格式，許多網際網路上應用的表單與文件亦以 PDF 檔案格式來儲存，大多數的電腦也已經安裝免費 PDF 閱覽程式 Adobe Acrobat Reader。

近年來電子公文與附件的呈現、交換、保存上普遍地採用 PDF 檔案格式，「文書及檔案管理電腦化作業規範」附錄 8 中已明訂 PDF 格式為適合保存之檔案格式。

(五)PDF 的四種主要國際標準：

由於 PDF 在各行各業的廣泛應用，現今已經衍生出下列國際標準：

1. PDF/A：A 是 Archive，是考量檔案歸檔用之 PDF 標準，只要符合此標準之 PDF 格式檔案，即使過了數十年以後，依舊能確保可以被正常讀取並開啟，而且文件呈現格式與儲存時一模一樣，非常適用於電子公文檔案長期保存使用。
2. PDF/X：X 是 eXchange，用於交換用的 PDF 標準，屬印刷上運用之檔案交換格式。
3. PDF/H：H 是 Health，醫療用設備可讀取之 PDF 標準，屬目前醫療體系之標準檔案交換格式。
4. PDF/E：E 是 Engineering，工程工業上運用的 PDF 標準，屬工業上交換檔案時的標準格式。

PDF/A 是 ISO(International Standard Organization)標準 document management-Electronic document file format for long-term preservation-Part 1: Use of PDF 1.4 (PDF/A-1) (ISO 19005-1:2005)，因此，如果考慮長期保存之因素，所保存之 PDF 檔案格式應該盡可能符合 PDF/A 標準。

(六)PDF/A 標準包含了下列幾點規範：

1. 檔案不能包含音訊或視訊資料：如果 PDF 檔案中含有音訊或視訊資料，可能必須額外之影音解碼器才能開啟，因此會與影音解碼器相依而導致保存困難。
2. 不能執行 JavaScript 與其他夾帶附件的功能：夾帶附件的檔案使得 PDF 檔案需要額外程式處理附件檔；JavaScript 動態內容可能會使呈現行為失去控制。
3. 所有文字都必須內嵌，以便在不同電腦讀取：完全杜絕不同電腦上文字編碼與字體所帶來之視覺上差異。
4. 顏色必須是 device-independent（與設備無關）：於任何顯示設備上都可以正確顯示原本顏色（必須符合 ICC(International Color Consortium)國際色彩協會之顏色內嵌規範，會嵌入符合 ICC 規範的色彩描述檔），完全杜絕不同電腦上色彩演繹實作所帶來之視覺差異。
5. 圖像之透明物件會被平面化：透明物件透明程度無法絕對化，因此必須避免使用透明物件。
6. 不可以加密。
7. 詮釋資料：必須包含詮釋資料（Metadata），以方便搜尋。

(七)根據 ISO 標準，PDF/A 又可細分為以下：

1. PDF/A -1a：PDF/A-1a 確保文件之邏輯架構，並規定文字內容必須依讀取順序排列，以便日後文字萃取，文字萃取對於日後文件在移動裝置上的讀取是相當重要的，因為多數的移動裝置會受限於畫面大小，其文字必須重新排列。要符合此標準除了要符合 PDF/A-1b 所有規範外，PDF 文件還需要加入 Tag(標籤)，設定標籤化的 PDF 文件是為了方便各種裝置（如手持裝置或是閱讀輔助裝置等）閱讀 PDF/A 檔案。
2. PDF/A-1b：基本上完全符合 PDF/A 標準，同時也是符合標準之最小要求，達到本類要求屬 B 等級相容，可以確保以後文件開啟時保持一模一樣之呈現格式。
3. PDF/A-2：此標準加入了 PDF1.5、1.6、1.7 中的新功能，具有向後相容性，所有符合 PDF/A-1 的檔案都符合 PDF/A-2，但符合 PDF/A-2 的檔案不一定符合 PDF/A-1。

二、轉置測試環境

項目	規格
中央處理器	Pentium (R)Dual-Core CPU P8700 @ 2.6GHz
硬碟	110GB
記憶體	3.0GB
作業系統	Microsoft Windows XP Professional Version 2002 Service Pack 3
顯示晶片	Intel G33/G31 Express ChiPSet Family

三、檔案轉置結果

(一) 公文樣式之檔案轉置個案

1. 測試檔案詳細資料

(1) 測試檔案 1

電子檔案基本資料			
檔案名稱		台北市函(系統文件).DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.4secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		47KB	20KB
品質驗證	PSNR	34.3381	
	SSIM	0.997586	

(2)測試檔案 2

電子檔案基本資料			
檔案名稱		台北市函(展示).DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.6secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		48KB	20KB
品質驗證	PSNR	34.1717	
	SSIM	0.997468	

(3)測試檔案 3

電子檔案基本資料			
檔案名稱		台北市函(配合合約).DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		49KB	20KB
品質驗證	PSNR	33.8595	
	SSIM	0.997246	

(4)測試檔案 4

電子檔案基本資料			
檔案名稱		台北市函(電子檔案).DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		1.3secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		47KB	20KB
品質驗證	PSNR	34.3381	
	SSIM	0.997586	

(5)測試檔案 5

電子檔案基本資料			
檔案名稱		台北市函(颱風).DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		59KB	20KB
品質驗證	PSNR	32.9699	
	SSIM	0.996583	

2. 轉置前後檔案容量比較：轉置前後檔案容量如下表，測試結果顯示轉置後檔案 PDF/A 格式之平均容量比原始檔案 DOCX 格式大。

序號	檔案名稱	轉置後檔案	原始檔案
1.	台北市函(系統文件).DOCX	47KB	20KB
2.	台北市函(展示).DOCX	48KB	20KB
3.	台北市函(配合合約).DOCX	49KB	20KB
4.	台北市函(電子檔案).DOCX	47KB	20KB
5.	台北市函(颱風).DOCX	59KB	20KB
平均		50KB	20KB

3. 轉置前後檔案品質驗證：利用 PDFCreator 工具將測試來源檔案 DOCX 格式及轉置後檔案 PDF/A 格式，分別轉置為 JPEG 格式，再將 2 個 JPEG 格式檔案進行 PSNR 及 SSIM 品質驗證，轉置前後檔案品質驗證比較如下表。若 PSNR 值為 30 左右，表示失真越少，2 張圖相似度高，而 SSIM 值大於 0.95 代表可接受。5 個測試檔 PSNR 平均值約 33.93546，SSIM 平均值為 0.997294，由此可知，PSNR 值和 SSIM 值皆已符合標準。以肉眼觀察其 DOCX 格式與 PDF/A 格式差異並不大。

序號	檔案名稱	PSNR	SSIM
1.	台北市函(系統文件).DOCX	34.34	0.99
2.	台北市函(展示).DOCX	34.17	0.99
3.	台北市函(配合合約).DOCX	33.86	0.99
4.	台北市函(電子檔案).DOCX	34.34	0.99
5.	台北市函(颱風).DOCX	32.97	0.99
平均		33.94	0.99

(二) 非公文樣式之檔案轉置個案

1. 測試檔案詳細資料

(1) 測試檔案 1

電子檔案基本資料			
檔案名稱		不重疊文字方塊.DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		57KB	13KB
品質驗證	PSNR	33.4059	
	SSIM	0.997801	

(2) 測試檔案 2

電子檔案基本資料			
檔案名稱		重疊文字方塊.DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		57KB	13KB
品質驗證	PSNR	31.1112	
	SSIM	0.99634	

(3)測試檔案 3

電子檔案基本資料			
檔案名稱		彩色圖片測試檔.DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOC
Size 大小		33KB	236KB
品質驗證	PSNR	46.8031	
	SSIM	0.9987	

(4)測試檔案 4

電子檔案基本資料			
檔案名稱		黑白圖片測試檔.DOCX	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.8secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	DOCX
Size 大小		1981KB	2636KB
品質驗證	PSNR	63.8328	
	SSIM	0.999999	

4. 轉置前後檔案容量比較：轉置前後檔案容量如下表，測試結果顯示，轉置後檔案 PDF/A 格式內含圖片檔，其檔案容量會比原始檔案小。

序號	檔案名稱	轉置後檔案	原始檔案
1.	不重疊文字方塊.DOCX	57KB	13KB
2.	重疊文字方塊.DOCX	57KB	13KB
3.	彩色圖片測試檔.DOCX	33KB	236KB
4.	黑白圖片測試檔.DOCX	1981KB	2636KB
平均		532KB	724.5KB

5. 轉置前後檔案品質驗證：利用 PDFCreator 工具將測試檔案 DOCX 格式及轉置後檔案 PDF/A 格式分別轉置為 JPEG 格式，再將 2 個 JPEG 格式檔案進行 PSNR 及 SSIM 之品質驗證，轉置前後檔案品質驗證比較如下表。若 PSNR 值為 30 左右，表示失真越少，2 張圖相似度高，而 SSIM 值大於 0.95 代表可接受。4 個測試檔 PSNR 平均值約 43.78825，SSIM 平均值為 0.99821，由此可知，PSNR 值和 SSIM 值皆已符合標準。以肉眼觀察其 DOCX 格式與 PDF/A 格式差異並不大。

序號	檔案名稱	PSNR	SSIM
1.	不重疊文字方塊.DOCX	33.41	0.99
2.	重疊文字方塊.DOCX	31.11	0.99
3.	彩色圖片測試檔.DOCX	46.80	0.99
4.	黑白圖片測試檔.DOCX	63.83	0.99
平均		43.79	0.99

四、遭遇問題

將 Microsoft Word DOCX 檔案利用 PDFCreator 工具將 DOCX 格式轉置為 PDF/A 格式，以符合保存建議格式。

針對非點陣圖形檔案的品質驗證，必須先將來源與轉置後檔案皆轉置成為點陣圖形，由於驗證過程中會牽涉到另外一個轉置程序，亦會影響執行效能，建議持續研究更直接有效之驗證方法。

陸、POSTSCRIPT 格式轉置為 PDF/A 格式

一、轉置目的

POSTSCRIPT 為一種文件描述語言，是由美國 Adobe 公司於 1985 年所發表的技術。POSTSCRIPT 可描述(或定義)字型邊框或者是任意的平面設計的圖型。目前被普遍使用在印刷領域，例如：數位印刷機、螢幕顯示及雷射印表機等輸出設備。

POSTSCRIPT 文件是指以 POSTSCRIPT 所撰寫之程式檔案，目前大部分的人會藉由應用程式進行文字編排及圖形繪製，再由該應用程式輸出 POSTSCRIPT 文件。例如：Microsoft Word 及 Adobe Illustrator 都可以輸出 POSTSCRIPT 文件。科學界常用 TeX 或 LaTeX 處理科學文獻的排版工作，排版後之輸出檔案叫做 dvi(device independent)格式。而今天。大部分的使用者會將 dvi 格式之檔案轉換成 POSTSCRIPT 文件。

科學文獻常需使用數學符號及圖表，而 POSTSCRIPT 提供可高品質之呈現畫面，因而大部分之科學文獻常以 POSTSCRIPT 文件作為儲存媒介。對長期保存來說，重要之科學文獻是需要被保存。

二、轉置測試環境

項目	規格
中央處理器	Pentium (R)Dual-Core CPU P8700 @ 2.6GHz
硬碟	110GB
記憶體	3.0GB
作業系統	Microsoft Windows XP Professional Version 2002 Service Pack 3
顯示晶片	Intel G33/G31 Express ChiPSet Family

三、檔案轉置結果

(一) 測試檔 1

電子檔案基本資料			
檔案名稱		Test1.PS	
檔案個數		1	
掃毒工具名稱/版本		OfficeScan/6.3.1	
病毒掃描結果		一切正常沒有病毒	
轉置花費時間		0.5secs	
比較			
		轉置後檔案	原始檔案
電子檔案格式		PDF/A	POSTSCRIPT
Size 大小		6KB	5KB
品質驗證	PSNR	48.1754	
	SSIM	0.999888	

(二) 測試檔 2

電子檔案基本資料		
檔案名稱		Test2.PS
檔案個數		1
掃毒工具名稱/版本		OfficeScan/6.3.1
病毒掃描結果		一切正常沒有病毒
轉置花費時間		0.5secs
比較		
	轉置後檔案	原始檔案
電子檔案格式	PDF/A	POSTSCRIPT
Size 大小	8KB	7KB
品質驗證	PSNR	40.7649
	SSIM	0.9993

(三) 測試檔 3

電子檔案基本資料		
檔案名稱		Test3.PS
檔案個數		1
掃毒工具名稱/版本		OfficeScan/6.3.1
病毒掃描結果		一切正常沒有病毒
轉置花費時間		0.5secs
比較		
	轉置後檔案	原始檔案
電子檔案格式	PDF/A	POSTSCRIPT
Size 大小	14KB	14KB
品質驗證	PSNR	34.0467
	SSIM	0.997812

(四) 測試檔 4

電子檔案基本資料		
檔案名稱		Test4.PS
檔案個數		1
掃毒工具名稱/版本		OfficeScan/6.3.1
病毒掃描結果		一切正常沒有病毒
轉置花費時間		0.5secs
比較		
	轉置後檔案	原始檔案
電子檔案格式	PDF/A	POSTSCRIPT
Size 大小	22KB	18KB
品質驗證	PSNR	31.0813
	SSIM	0.995686

(五) 測試檔 5

電子檔案基本資料		
檔案名稱		Test5.PS
檔案個數		1
掃毒工具名稱/版本		OfficeScan/6.3.1
病毒掃描結果		一切正常沒有病毒
轉置花費時間		0.5secs
比較		
	轉置後檔案	原始檔案
電子檔案格式	PDF/A	POSTSCRIPT
Size 大小	30KB	23KB
品質驗證	PSNR	30.7954
	SSIM	0.995186

1. 轉置前後檔案容量比較：轉置前後檔案容量如下表，測試結果顯示，轉置後檔案 PDF/A 格式之平均容量比原始檔案 POSTSCRIPT 格式大。

序號	檔案名稱	轉置後檔案	原始檔案
1.	Test1.PS	6KB	5KB
2.	Test2.PS	8KB	7KB
3.	Test3.PS	14KB	14KB
4.	Test4.PS	22KB	18KB
5.	Test5.PS	30KB	23KB
平均		16KB	13.4KB

2. 轉置前後檔案品質驗證：利用 PDFCreator 工具將測試檔案 POSTSCRIPT 格式及轉置後檔案 PDF/A 格式，分別轉置為 JPEG 格式，再將 2 個 JPEG 格式檔案進行 PSNR 及 SSIM 之品質驗證，轉置前後檔案品質驗證比較如下表。若 PSNR 值為 30 左右，表示失真越少，2 張圖相似度高，而 SSIM 值大於 0.95 代表可接受。5 個測試檔 PSNR 平均值約 36.97274，SSIM 平均值為 0.997574，由此可知 PSNR 值和 SSIM 值皆已符合標準。以肉眼觀察其 POSTSCRIPT 格式與 PDF/A 格式差異並不大。

序號	檔案名稱	PSNR	SSIM
1.	Test1.PS	48.18	0.99
2.	Test2.PS	40.76	0.99
3.	Test3.PS	34.05	0.99
4.	Test4.PS	31.08	0.99
5.	Test5.PS	30.80	0.99

序號	檔案名稱	PSNR	SSIM
平均		36.97	0.99

四、遭遇問題

針對非點陣圖形檔案之品質驗證，必須先將來源與轉置後檔案皆轉置成為點陣圖形，由於驗證過程中會牽涉到另外一個轉置程序，亦會影響執行效能，建議持續研究更直接有效之驗證方法。

柒、機關實作電子檔案格式轉置

為驗證電子檔案保存工具之轉置品質，並確認工具符合機關需求，檔案管理局擇定中央印製廠、國立師範大學及財政部臺灣省中區國稅局進行試轉作業。中央印製廠及國立師範大學係進行 DOC 格式轉置為 PDF/A 格式之試作，財政部臺灣省中區國稅局係進行 WDL 格式轉置為 PDF/A 格式之試作。

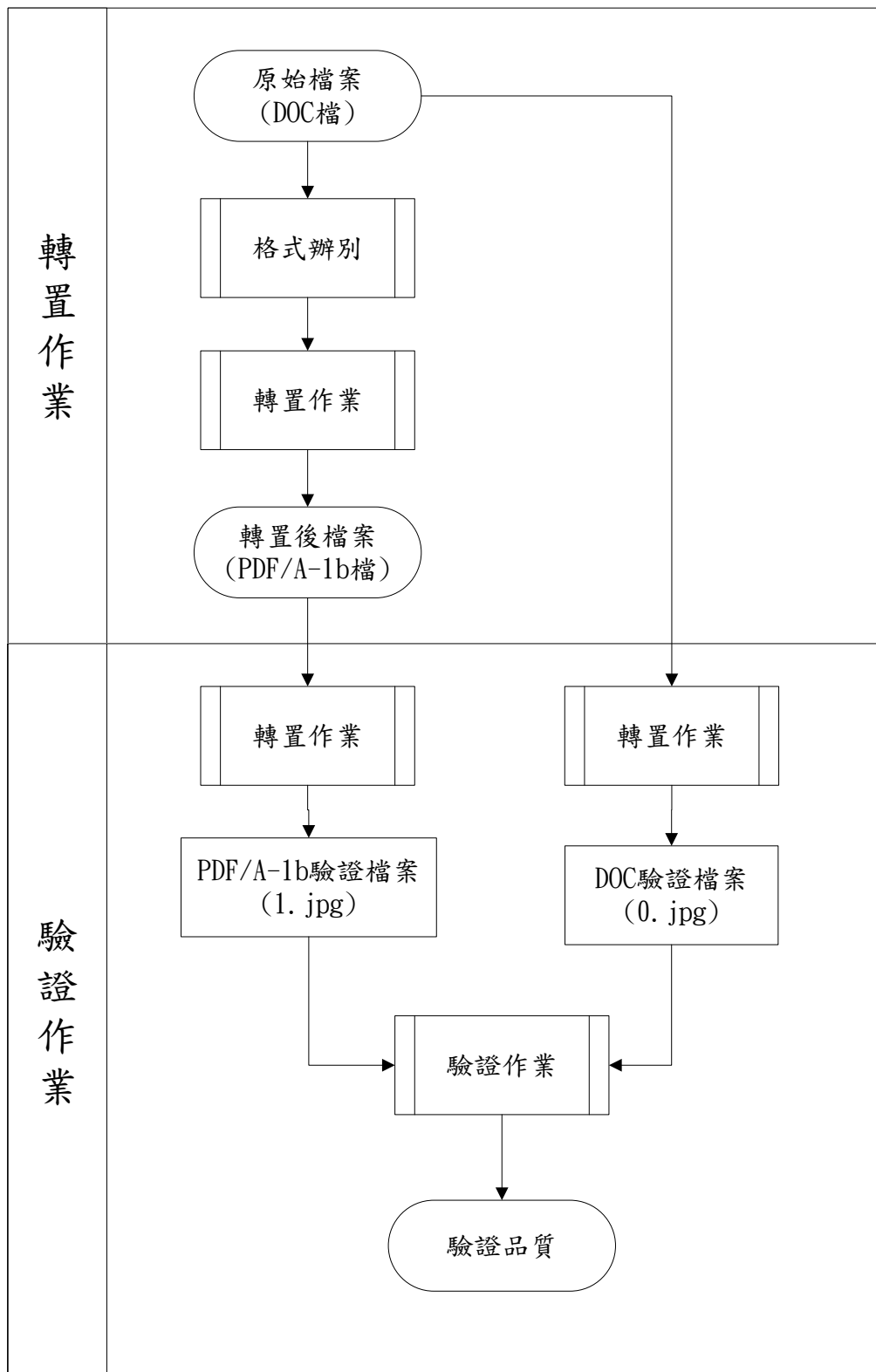
一、DOC 格式轉置為 PDF/A 格式

(一) 轉置程序

1. 轉置程序說明：

- (1) 匯入 DOC 檔案格式。
- (2) 執行 DOC 檔案格式轉置為 PDF/A 檔案格式。
- (3) 執行 DOC 檔案格式轉置為 JPEG 檔案格式。
- (4) 執行 PDF/A 檔案格式轉置為 JPEG 檔案格式。
- (5) 執行驗證。
- (6) 呈現驗證結果。

2. 轉置程序流程圖：



(一) 中央印製廠檔案轉置結果

1. 品質檢測結果

本次試轉共進行 2 次測試轉置作業，共 32 筆資料。

- (1) 測試一：在第一次測試中，執行了 32 個檔案的轉置作業，轉置結果顯示出有 32 個檔案都發生轉置失敗，其失敗原因為進行驗證之前，系統會將 DOC 檔和 PDF 檔各自轉為 JPG 圖檔，發現轉置出來的 JPG 圖檔檔名中多了一個%d 的字元，以致於系統無法正常進行驗證。

中央印製廠電子檔案轉置-測試計畫與成果表	
轉置說明：	(1) 啟動電子檔案保存工具。 (2) 匯入轉置檔案。 (3) 進行轉置。(DOC->PDF/A) (4) 等待程式完成或結束，記錄「工作報告」。工作報告中記錄原始檔案檔名、轉置後檔名、轉置開始時間、轉置結束時間、花費時間(秒)、驗證(PSNR)和驗證(SSIM)。
驗證說明：	電子檔案保存工具中包含自動驗證的程序。利用 PDFCreator 工具將測試檔案 DOC 及轉置後檔案 PDF/A 格式分別轉置為 JPEG 格式，再將兩個 JPEG 格式檔案進行 PSNR 及 SSIM 的品質驗證，但此方法並不能保證轉置內容的正確性，因而需要搭配主觀的方式(肉眼觀看)來檢驗轉置後的品質。
測試資料說明： (選定的資料區間)	近一個月的公文檔案。
轉置開始時間	日期：民國 100 年 9 月 19 日 時間：11:00:00

中央印製廠電子檔案轉置-測試計畫與成果表				
轉置完成(或結束)時間		日期：民國 100 年 9 月 19 日 時間：11:34:00		
測試總筆數		32 筆		
完成總筆數		32 筆		
轉置失敗筆數		32 筆		
平均每筆工作時間		60 秒		
完成後檔案總筆數		0 筆(共 0 個檔案)		
轉置後檔案檢查：		共檢查 0 筆(人工抽檢至少 2%，不包含 36 筆轉置失敗)		
品質與正確度檢查(主觀)	檢查原先文字可閱讀性(若有)	<input type="checkbox"/> 可辨視	<input type="checkbox"/> 不可辨	註記：
	檢查原先影像檔轉置後長寬比(若有)	<input type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查轉置頁數是否正確	<input type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查是否有應轉未轉之檔案	<input type="checkbox"/> 有	<input type="checkbox"/> 無	註記：
	檢查是否有不應轉卻轉之檔案	<input type="checkbox"/> 有	<input type="checkbox"/> 無	註記：
可否順利完成		<input type="checkbox"/> 可 <input checked="" type="checkbox"/> 否，結束原因(或錯誤訊息)：驗證發生問題。		

(2) 測試二：

中央印製廠電子檔案轉置-測試計畫與成果表	
轉置說明：	(1) 啟動電子檔案保存工具。 (2) 匯入轉置檔案。 (3) 進行轉置。(DOC->PDF/A) (4) 等待程式完成或結束，記錄「工作報告」。工作報告中記錄原始檔案檔名、原始檔案大小、轉置後檔名、轉置後檔案大小(KB)、轉置開始時間、轉置結束時間、花費時間(秒)、狀態、驗證(PSNR)和驗證(SSIM)。
驗證說明：	電子檔案保存工具中包含自動驗證的程序。利用 PDFCreator 工具將測試檔案 DOC 及轉置後檔案 PDF/A 格式分別轉置為 JPEG 格式，再將兩個 JPEG 格式檔案進行 PSNR 及 SSIM 的品質驗證，但此方法並不能保證轉置內容的正確性，因而需要搭配主觀的方式(肉眼觀看)來檢驗轉置後的品質。
測試資料說明： (選定的資料區間)	一個月的公文檔案
轉置開始時間	日期：民國 100 年 10 月 14 日 時間：13:00:00
轉置完成(或結束)時間	日期：民國 100 年 10 月 14 日 時間：13:34:00
測試總筆數	32 筆
完成總筆數	32 筆
轉置失敗筆數	0 筆
平均每筆工作時間	60 秒
完成後檔案總筆數	32 筆(共 32 個檔案)
轉置後檔案檢查：	共檢查 5 筆(人工抽檢至少 2%)

中央印製廠電子檔案轉置-測試計畫與成果表				
品質與正確度檢查(主觀)	檢查原先文字可閱讀性(若有)	<input checked="" type="checkbox"/> 可辨視	<input type="checkbox"/> 不可辨	註記：
	檢查原先影像檔轉置後長寬比 (若有)	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查轉置頁數是否正確	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查是否有應轉未轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
	檢查是否有不應轉卻轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
可否順利完成		<input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否。		

(二) 中央印製廠遭遇問題

針對 JPEG 圖檔檔名中多了一個%d 的字元問題，經研究後發現，電子檔案保存工具是使用 PDFCreator 的軟體進行轉置，在 PDFCreator 設定中，將「允許檔名中有特殊 Ghostscript 字元」的選項設定不勾選後，再進行轉置測試，產生出的圖檔檔名就不會多出%d 的字元了。

在中央印製廠的電腦中，發現該公文管理系統也利用 PDFCreator 相關版本進行檔案轉置作業，但 PDFCreator 版本較

為早期，且該軟體的設定值會有所不同。如要使用電子檔案保存工具的話，必須移除其他版本 PDFCreator 軟體，才能使電子檔案保存工具正常運作，但相對的將會造成該公文管理系統無法正常的運作。

(三) 國立師範大學檔案轉置結果

1. 品質檢測結果

本次試轉共進行 2 次測試轉置作業，共 103 筆資料。

(1) 測試一：在第一次測試中，執行了 103 個檔案的轉置作業，轉置結果顯示出有 36 個檔案發生轉置失敗，其失敗原因有 2 點，以下述之：

A. 當 DOC 檔的檔名中包含空白字元（例如：82 record.DOC），會導致電子檔案保存工具無法正常轉置。

B. 電子檔案保存工具預設轉置最長執行時間為 1 分鐘，當轉置時間到達 1 分鐘尚未轉置完成時，系統將會強制結束轉置作業。當檔案容量愈大或檔案頁數愈多時，轉置時間相對會愈長。因此，轉置時間若大於系統所預設的時間，檔案將無法正常轉置及驗證。

國立師範大學電子檔案轉置-測試計畫與成果表				
轉置說明：		(1) 啟動電子檔案保存工具。 (2) 匯入轉置檔案。 (3) 進行轉置。(DOC->PDF/A) (4) 等待程式完成或結束，記錄「工作報告」。工作報告中記錄原始檔案檔名、轉置後檔名、轉置開始時間、轉置結束時間、花費時間(秒)、驗證(PSNR)和驗證(SSIM)。		
驗證說明：		電子檔案保存工具中包含自動驗證的程序。利用 PDFCreator 工具將測試檔案 DOC 及轉置後檔案 PDF/A 格式分別轉置為 JPEG 格式，再將兩個 JPEG 格式檔案進行 PSNR 及 SSIM 的品質驗證，但此方法並不能保證轉置內容的正確性，因而需要搭配主觀的方式(肉眼觀看)來檢驗轉置後的品質。		
測試資料說明： (選定的資料區間)		不同年度的會議紀錄		
轉置開始時間		日期：民國 100 年 9 月 16 日 時間：10:30:09		
轉置完成(或結束)時間		日期：民國 100 年 9 月 16 日 時間：16:01:54		
測試總筆數		103 筆		
完成總筆數		103 筆		
轉置失敗筆數		36 筆		
平均每筆工作時間		177 秒		
完成後檔案總筆數		67 筆(共 103 個檔案)		
轉置後檔案檢查：		共檢查 10 筆(人工抽檢至少 2%，不包含 36 筆轉置失敗)		
品質與正確度	檢查原先文字可閱讀性(若有)	<input checked="" type="checkbox"/> 可辨視	<input type="checkbox"/> 不可辨	註記：

國立師範大學電子檔案轉置-測試計畫與成果表				
檢查 (主觀)	檢查原先影像檔轉置後長寬比(若有)	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查轉置頁數是否正確	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查是否有應轉未轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
	檢查是否有不應轉卻轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
可否順利完成		<input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否。		

(2) 測試二：第一次測試發生的問題進行修正後進行第二次試轉作業。

國立師範大學電子檔案轉置-測試計畫與成果表	
轉置說明：	(1) 啟動電子檔案保存工具。 (2) 匯入轉置檔案。 (3) 進行轉置。(DOC->PDF/A) (4) 等待程式完成或結束，記錄「工作報告」。工作報告中記錄原始檔案檔名、原始檔案大小、轉置後檔名、轉置後檔案大小(B)、轉置開始時間、轉置結束時間、花費時間(秒)、狀態、驗證(PSNR)和驗證(SSIM)。
驗證說明：	電子檔案保存工具中包含自動驗證的程序。利用PDFCreator 工具將測試檔案 DOC 及轉置後檔案 PDF/A 格式分別轉置為 JPEG 格式，再將兩個 JPEG 格式檔案進行 PSNR 及 SSIM 的品質驗證，但此方法並不能保證轉置內容的正確性，因

國立師範大學電子檔案轉置-測試計畫與成果表				
		而需要搭配主觀的方式(肉眼觀看)來檢驗轉置後的品質。		
測試資料說明： (選定的資料區間)		不同年度的會議紀錄		
轉置開始時間		日期：民國 100 年 10 月 17 日 時間：10:18:00		
轉置完成(或結束)時間		日期：民國 100 年 10 月 17 日 時間：17:24:07		
測試總筆數		103 筆		
完成總筆數		103 筆		
轉置失敗筆數		0 筆		
平均每筆工作時間		177 秒		
完成後檔案總筆數		103 筆(共 103 個檔案)		
轉置後檔案檢查：		共檢查 20 筆(人工抽檢至少 2%)		
品質與正確度檢查(主觀)	檢查原先文字可閱讀性(若有)	<input checked="" type="checkbox"/> 可辨視	<input type="checkbox"/> 不可辨	註記：
	檢查原先影像檔轉置後長寬比(若有)	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查轉置頁數是否正確	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查是否有應轉未轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
	檢查是否有不應轉卻轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
可否順利完成		<input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否。		

(四) 國立師範大學遭遇問題

針對 DOC 檔之檔名中包含有空白字元問題，已修正工具，於第二次測試時已能正常的運作。

針對實際的轉置時間若大於系統所預設的最長執行轉置時間，檔案將無法正常轉置及驗證問題，目前系統預設轉置時間為 10 分鐘，並且提供設定畫面，使用者可自行調整轉置最長執行時間。

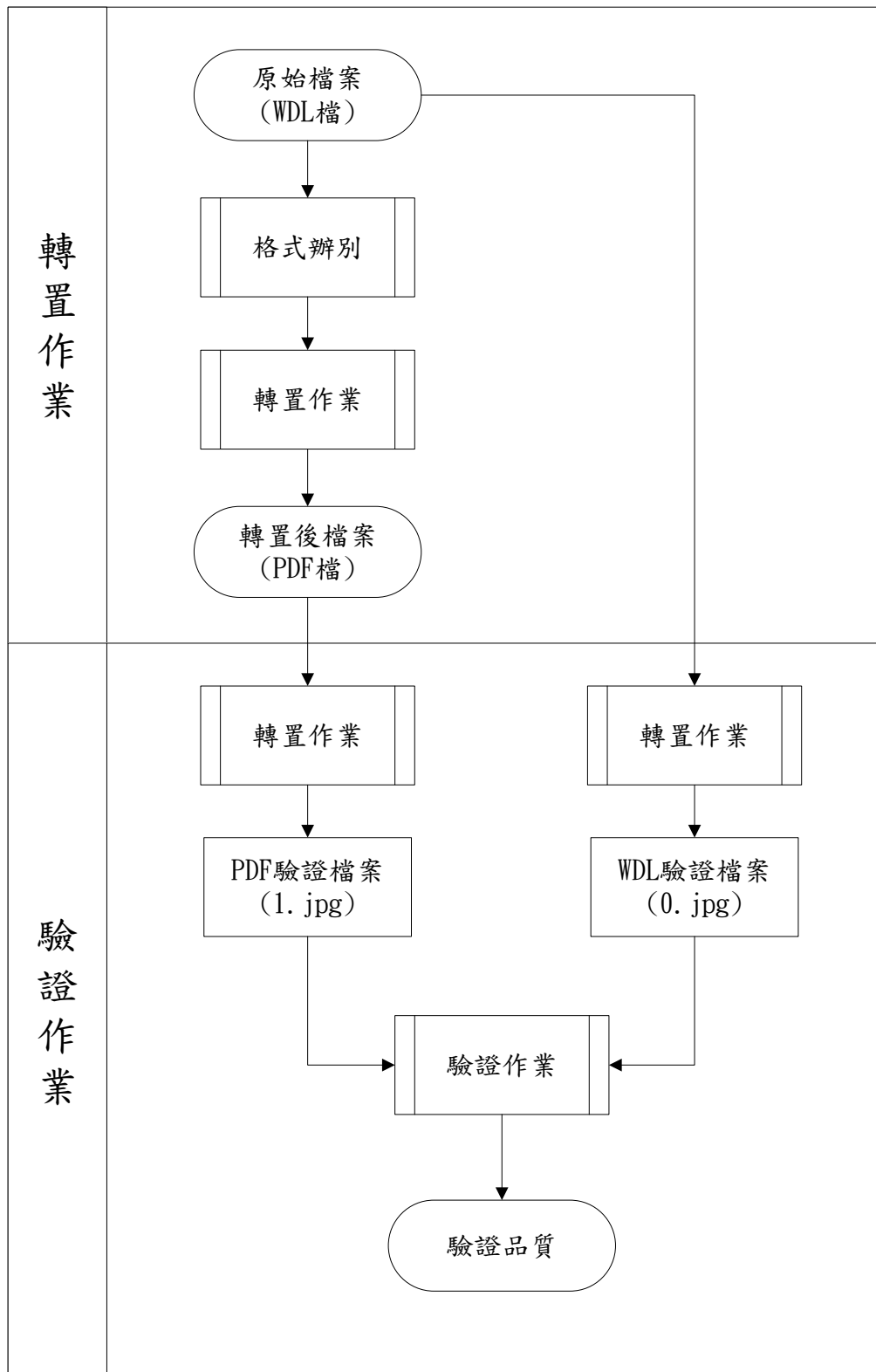
二、WDL 格式轉置為 PDF 格式

(一) 轉置程序

1. 轉置程序說明：

- (1) 匯入 WDL 檔案格式。
- (2) 執行 WDL 檔案格式轉置為 PDF 檔案格式。
- (3) 執行 WDL 檔案格式轉置為 JPEG 檔案格式。
- (4) 執行 PDF 檔案格式轉置為 JPEG 檔案格式。
- (5) 執行驗證。
- (6) 呈現驗證結果。

2. 轉置程序流程圖：



(二) 檔案轉置結果

1. 品質檢測結果

本次試轉共進行 1 次測試轉置作業，檔案筆數為 169 筆資料。本次的測試轉置作業進行順利，每筆轉置時間大約 0.3 秒左右，但驗證時間大約每筆約 2 分鐘左右，其原因在於結構相似度(SSIM)的品質評估方法所花費的時間較長，再加上電腦設備的效能影響，增加驗證的時間成本。

財政部臺灣省中區國稅局電子檔案轉置-測試計畫與成果表	
轉置說明：	(1) 啟動電子檔案保存工具程式。 (2) 匯入轉置檔案。 (3) 進行轉置。(WDL->PDF)。 (4) 等待程式完成或結束，記錄「工作報告」。工作報告中記錄原始檔案檔名、原始檔案大小、轉置後檔名、轉置後檔案大小(KB)、轉置開始時間、轉置結束時間、花費時間(秒)、狀態、驗證(PSNR)和驗證(SSIM)。
驗證說明：	電子檔案保存工具中包含自動驗證的程序。利用 PDFCreator 工具將測試檔案 WDL 及轉置後檔案 PDF 格式分別轉置為 JPEG 格式，再將兩個 JPEG 格式檔案進行 PSNR 及 SSIM 的品質驗證，但此方法並不能保證轉置內容的正確性，因而需要搭配主觀的方式(肉眼觀看)來檢驗轉置後的品質。
測試資料說明： (選定的資料區間)	一年度的公文檔
測試資料總容量：	1.36MB
轉置開始時間	日期：民國 <u>100</u> 年 <u>10</u> 月 <u>18</u> 日

財政部臺灣省中區國稅局電子檔案轉置-測試計畫與成果表				
		時間：10:19:09		
轉置完成(或結束)時間		日期：民國 <u>100</u> 年 <u>10</u> 月 <u>18</u> 日 時間：16:18:43		
測試總筆數		169 筆		
完成總筆數		169 筆		
轉置失敗筆數		0 筆		
平均每筆工作時間		120.00 秒		
完成後檔案總筆數		169 筆(共 169 個檔案)		
完成後檔案總容量		9.98MB		
轉置後檔案檢查：		共檢查 <u>20</u> 筆(人工抽檢至少 2%)		
品質與正確度檢查(主觀)	檢查原先文字可閱讀性(若有)	<input checked="" type="checkbox"/> 可辨視	<input type="checkbox"/> 不可辨	註記：
	檢查原先影像檔轉置後長寬比(若有)	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查轉置頁數是否正確	<input checked="" type="checkbox"/> 正確	<input type="checkbox"/> 不正確	註記：
	檢查是否有應轉未轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
	檢查是否有不應轉卻轉之檔案	<input type="checkbox"/> 有	<input checked="" type="checkbox"/> 無	註記：
可否順利完成		<input checked="" type="checkbox"/> 可 <input type="checkbox"/> 否。		

捌、問題檢討

本專案 100 年度已於電子檔案保存工具整合相關轉置工具，進行 DOCX 格式及 POSTSCRIPT 格式，並擇定 3 個機關進行電子檔案保存工具試作，針對轉置品質說明如下：

一、DOCX 格式轉置 PDF/A 格式

以客觀標準評估轉置品質，SSIM 值大於 0.95 即為可接受範圍。經以轉置後檔案進行比對，其 SSIM 值平均約 0.99，轉置品質相當好；PSNR 值平均為 33，趨近於 40，其品質也相當不錯。在轉置時間部分，平均約 1 秒可完成轉置。

二、POSTSCRIPT 格式轉置 PDF/A 格式

在轉置品質部分，通常 SSIM 值大於 0.95 即為可接受範圍。經以轉置後檔案進行比對，其 SSIM 值平均約 0.99，轉置品質相當好；PSNR 值平均為 36.97，趨近於 40，其品質也相當不錯。在轉置時間部分，平均約 1 秒可完成轉置。但針對此類非點陣圖形檔案的品質驗證，必須先將來源與轉置後檔案皆轉置成為點陣圖形，由於驗證過程中會牽涉到另外一個轉置程序，當驗證品質不佳時，不能確定是格式轉置本身造成或是驗證之轉置程序造成，亦會影響執行效能，建議

持續研究更直接有效之驗證方法。

三、中央印製廠 DOC 格式轉置 PDF/A 格式

在這次的試轉過程中，發現當 PDFCreator 之設定不同時，會影響電子檔案保存工具轉置及驗證之運作。如使用電子檔案保存工具，須先確認該電腦是否已使用 PDFCreator 相關軟體，如有使用上之衝突，建議另外安裝至其他電腦上，才不會影響原系統之正常運作。

四、國立師範大學 DOC 格式轉置 PDF/A 格式

在轉置的過程中，主要遭遇到以下幾個問題：

序號	問題	影響	備註
1	DOC 檔的檔名中包含有空白字元。	導致系統無法完成轉置作業。	已解決。
2	實際的轉置時間若大於系統所預設的最長執行轉置時間。	導致系統無法完成轉置作業。	已解決。

五、財政部臺灣省中區國稅局 WDL 格式轉置 PDF 格式

WDL 格式轉置試轉作業可正常轉置及驗證，可以顯示電子檔案保存工具中 WDL 格式轉置功能屬穩定以符合機關需求。針對 WDL 格式轉置，目前所提供的 WDL 閱讀軟體可以正常開啟該局所提供之檔案。針對 WDL 格式轉置部分，必須在閱讀軟體可支援檔案之情況下，方能正常運作。