

檔案管理局

102 年度電子檔案保存管理機制 委託服務案

電子檔案保存工具 轉置實作報告 (1020709_01)

英福達科技股份有限公司 謹呈

民國 102 年 07 月

版本紀錄

| 版序 | 實施日期 | 修改內容 |
|------------|---------|--|
| 1020628_01 | 1020628 | 初版 |
| 1020709_01 | 1020709 | JHOVE 僅針對轉置後檔案做格式檢測，依建議修改測試表格內容。PCD 轉 JPEG 時，檔案大小大幅下降，依建議增加原始及轉置後檔案圖示。為避免文字重複出現，將結論與問題檢討修改為結論。 |
| | | |
| | | |

目錄

| | | |
|----|------------------------|--------|
| 壹、 | 簡介 | - 1 - |
| 貳、 | 沿革說明 | - 2 - |
| 參、 | 轉置驗證方法說明 | - 3 - |
| 一、 | JHOVE 格式介紹 | - 3 - |
| 肆、 | 轉置程序 | - 4 - |
| 伍、 | LEH 格式轉置 TIFF 格式 | - 5 - |
| 一、 | 轉置目的 | - 5 - |
| 二、 | 轉置測試環境 | - 5 - |
| 三、 | 檔案轉置結果 | - 6 - |
| 陸、 | PCD 格式轉置 JPEG 格式 | - 10 - |
| 一、 | 轉置目的 | - 10 - |
| 二、 | 轉置測試環境 | - 10 - |
| 三、 | 檔案轉置結果 | - 11 - |
| 柒、 | 結論 | - 15 - |
| 一、 | LEH 格式轉置 TIFF 格式 | - 15 - |
| 二、 | PCD 格式轉置 JPEG 格式 | - 15 - |

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，且資訊技術發展日新月異，電子檔案之格式與儲存媒體，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，提出檢討與解決方式。

貳、沿革說明

98 年度之轉置作業為影像類型轉置(TIFF 格式轉置 JPEG 格式、JPEG 格式轉置 TIFF 格式)、文件檔轉置(WDL 格式轉置 PDF/A 格式)、視訊類型轉置(MPEG-2 格式轉置 H.264 格式)。

99 年度之轉置作業為文件類型轉置(DOC 格式轉置 PDF/A 格式及 DOC 格式轉置 ODT 格式)、視訊類型轉置(WMV 格式轉置 MPEG-2 格式)及電子檔案封裝檔格式轉置。

100 年度之轉置作業為文件類型轉置(DOCX 格式轉置 PDF/A 格式及 POSTSCRIPT 格式轉置 PDF/A 格式)、影像類型轉置(TIFF 格式轉置 PNG 格式、GIF 格式轉置 PNG 格式)。

101 年度之轉置作業為文件類型轉置(PPT 格式轉置 PDF/A 格式、XLS 格式轉置 PDF/A 格式及 EML 格式轉置 PDF/A 格式)及聲音類型轉置(MP3 格式轉置 WAV 格式)。

102 年度之轉置作業為影像類型轉置(LEH 格式轉置 TIFF 格式、PCD 格式轉置 JPEG 格式)。98 年度至 102 年度相關轉置工具已整合至電子檔案保存工具。

參、轉置驗證方法說明

電子檔案保存工具針對影像檔類型格式轉置作業之品質驗證方式，係將轉置後檔案利用 JHOVE 格式檢測工具進行檢測，檢測是否為 TIFF 及 JPEG 影像檔的格式。

一、JHOVE 格式介紹

由 JSTOR 與哈佛大學圖書館共同發展電子檔案型態描述 (Characterization) 工具，功能摘述如下。

- (一) 識別(Identification)：經由特徵比對，推定電子檔案格式。
- (二) 確認 (Validation)：判斷電子檔案之格式完好 (Well-formed) 且具有有效性 (Valid)。
- (三) 提供 Module：AIFF、ASCII、Bytestream、GIF、HTML、JPEG、JPEG2000、PDF、TIFF、UTF-8、WAVE、XML 等格式辨識模組。
- (四) 屬性擷取：格式、版本、作者、主題、產生時間、修改時間、字型、TrueType 字體、頁數。

肆、轉置程序

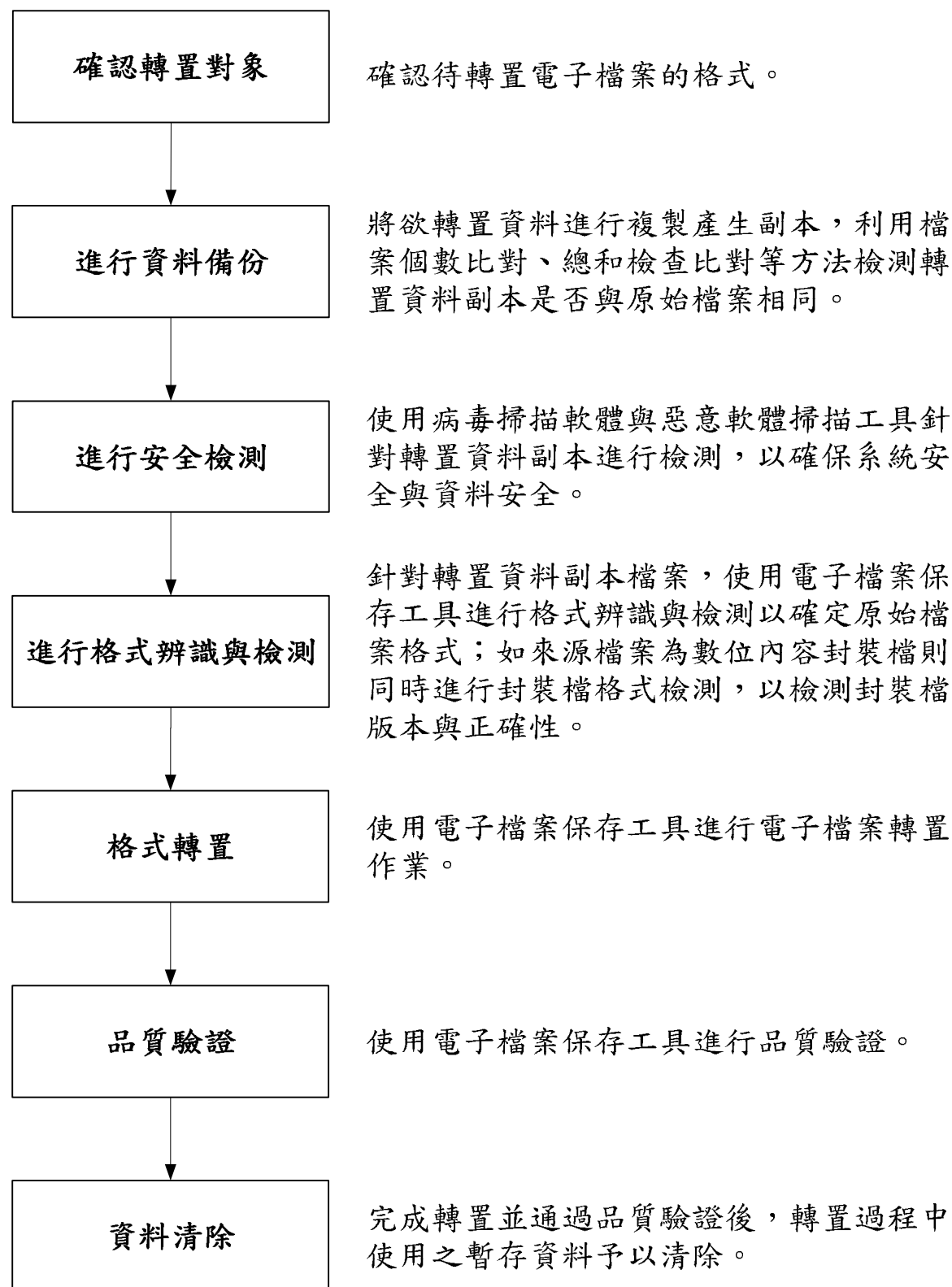


圖 1 轉置程序圖

伍、LEH 格式轉置 TIFF 格式

一、轉置目的

LEH 檔案為一種自製編碼加密過的格式，無法用一般瀏覽器開啟，須做解碼的動作，故無法進行品質的驗證。

二、轉置測試環境

表 1 轉置測試環境表

| 項目 | 規格 |
|-------|--|
| 中央處理器 | Intel(R) Core(TM) i7 CPU 950 @ 3.07GHz |
| 硬碟 | 500GB |
| 記憶體 | 4.0GB |
| 作業系統 | Microsoft Windows 7 Professional Version 32 位元 |
| 顯示晶片 | NVIDIA GeForce GT 220 |

三、檔案轉置結果

(一) 測試檔 1

表 2 測試檔 1 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|-------|
| 檔案名稱 | 00001B6A.leh | |
| 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 | 15 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | TIFF | LEH |
| Size 大小 | 195KB | 195KB |
| 格式驗證 | 正確 | |

(二) 測試檔 2

表 3 測試檔 2 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|------|
| 檔案名稱 | 00001B6B.leh | |
| 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 | 15 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | TIFF | LEH |
| Size 大小 | 82.8KB | 83KB |
| 格式驗證 | 正確 | |

(三) 測試檔 3

表 4 測試檔 3 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|-------|
| 檔案名稱 | 00001B6C.leh | |
| 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 | 15 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | TIFF | LEH |
| Size 大小 | 179KB | 179KB |
| 格式驗證 | 正確 | |

(四) 測試檔 4

表 5 測試檔 4 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|-------|
| 檔案名稱 | 00001B6D.leh | |
| 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 | 15 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | TIFF | LEH |
| Size 大小 | 502KB | 502KB |
| 格式驗證 | 正確 | |

(五) 測試檔 5

表 6 測試檔 5 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|-------|
| 檔案名稱 | 00001B6E.leh | |
| 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 | 15 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | TIFF | LEH |
| Size 大小 | 893KB | 893KB |
| 格式驗證 | 正確 | |

(六) 轉置前後檔案容量與格式驗證比較

- 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，原始檔案 LEH 格式之平均容量與轉
置後檔案 TIFF 格式幾乎一樣。

表 7 LEH 格式轉置前後檔案容量比較表

| 序號 | 檔案名稱 | 轉置後檔案 | 原始檔案 |
|----|--------------|----------|---------|
| 1. | 00001B6A.leh | 195KB | 195KB |
| 2. | 00001B6B.leh | 82.8KB | 83KB |
| 3. | 00001B6C.leh | 179KB | 179KB |
| 4. | 00001B6D.leh | 502KB | 502KB |
| 5. | 00001B6E.leh | 893KB | 893KB |
| 平均 | | 370.36KB | 370.4KB |

2.轉置前後檔案品質驗證：利用 JHOVE 格式檢測工具
來檢測轉置後檔案是否為 TIFF 格式。由下表可知 5
個測試檔的檢測結果皆為正確。

表 8 LEH 格式轉置後檔案格式驗證表

| 序號 | 檔案名稱 | 格式驗證 |
|----|--------------|------|
| 1. | 00001B6A.leh | 正確 |
| 2. | 00001B6B.leh | 正確 |
| 3. | 00001B6C.leh | 正確 |
| 4. | 00001B6D.leh | 正確 |
| 5. | 00001B6E.leh | 正確 |

陸、PCD 格式轉置 JPEG 格式

一、轉置目的

PCD (Photo CD) 為柯達一種獨家自製的相片光碟格式，無法使用一般瀏覽器開啟，必須使用支援 PCD 格式的軟體才能檢視，因 PCD 格式採用了柯達精確顏色管理系統 (KPCMS)，故無法將任何檔案儲存成 PCD 格式，也無法進行品質驗證。

二、轉置測試環境

表 9 轉置測試環境表

| 項目 | 規格 |
|-------|--|
| 中央處理器 | Intel(R) Core(TM) i7 CPU 950 @ 3.07GHz |
| 硬碟 | 500GB |
| 記憶體 | 4.0GB |
| 作業系統 | Microsoft Windows 7 Professional Version 32 位元 |
| 顯示晶片 | NVIDIA GeForce GT 220 |

三、檔案轉置結果

(一) 測試檔 1

表 10 測試檔 1 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|---------|
| 檔案名稱 | IMG0001.PCD | |
| 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 | 3 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | JPEG | PCD |
| Size 大小 | 88KB | 4,400KB |
| 格式驗證 | 正確 | |



圖 2PCD 測試檔 1 原始檔案



圖 3PCD 測試檔 1 轉置後檔案

(二) 測試檔 2

表 11 測試檔 2 資料表

| 電子檔案基本資料 | | |
|-----------|------------------|---------|
| 檔案名稱 | IMG0002.PCD | |
| 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 秒 | |
| 比較 | | |
| | 轉置後檔案 | 原始檔案 |
| 電子檔案格式 | JPEG | PCD |
| Size 大小 | 56KB | 4,266KB |
| 格式驗證 | 正確 | |



圖 4PCD 測試檔 2 原始檔案



圖 5PCD 測試檔 2 轉置後檔案

(三) 轉置前後檔案容量與格式驗證比較

- 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，原始檔案 PCD 格式之平均容量比轉
置後檔案 JPEG 格式大。

表 12 PCD 格式轉置前後檔案容量比較表

| 序號 | 檔案名稱 | 轉置後檔案 | 原始檔案 |
|----|-------------|-------|---------|
| 1. | IMG0001.PCD | 88KB | 4,400KB |
| 2. | IMG0002.PCD | 56KB | 4,266KB |
| 平均 | | 72KB | 4,333KB |

- 2.轉置前後檔案品質驗證：利用 JHOVE 格式檢測工具
來檢測轉置後檔案是否為 JPEG 格式。由下表可知 2
個測試檔的檢測結果皆為正確。

表 13 PCD 格式轉置後檔案格式驗證表

| 序號 | 檔案名稱 | 格式驗證 |
|----|-------------|------|
| 1. | IMG0001.PCD | 正確 |
| 2. | IMG0002.PCD | 正確 |

柒、結論

本專案 102 年度已於電子檔案保存工具擴充相關格式轉置技術，包含 LEH 格式及 PCD 格式。以下整理於轉置過程中主要遭遇問題。

一、LEH 格式轉置 TIFF 格式

目前驗證的方式是採用 JHOVE 檢測工具來檢測檔案的格式是否正確，以確保符合 TIFF 的格式。但若要檢測原始檔案及轉置檔案的品質差異，目前並沒有適合的工具可進行檢測。

二、PCD 格式轉置 JPEG 格式

目前驗證的方式是採用 JHOVE 檢測工具來檢測檔案的格式是否正確，以確保符合 JPEG 的格式。但若要檢測原始檔案及轉置檔案的品質差異，目前並沒有適合的工具可進行檢測。