

# 檔案管理局

## 「電子檔案長期保存技術平台建置案」

### 電子檔案轉置技術指引

V1.01

一、	簡介.....	3
貳、	適用範圍.....	3
參、	轉置分類.....	3
一、	媒體轉置.....	3
二、	格式轉置.....	5
	(一) 電子檔案是依據某種特定的格式描述的資料，必須經由應用程式方能予以呈現，例如 Microsoft Office 中的 Word 檔案，其中記載了各段落的文字格式，必須依賴能解讀 Word 格式的應用程式才能將其呈現與輸出；各種應用程式也必需執行在作業系統之上，某些作業系統更只能執行在特定的硬體架構之上。.....	5
	(二) 格式轉置的時機如下：.....	5
	(三) 格式轉置注意事項.....	5
肆、	轉置品質.....	6
	(一) 客觀評量.....	6
	(二) 主觀評量.....	11
伍、	轉置作業程序.....	14
	(一) 確認轉置對象.....	14
	(二) 進行媒體複製.....	14
	(三) 進行安全檢測.....	14
	(四) 進行格式檢測.....	14
	(五) 確認轉置格式.....	14
	(六) 格式轉置.....	14
	(七) 品質驗證.....	15
	(八) 詮釋資料.....	15
	(九) 原儲存媒體與暫存媒體消磁或銷毀.....	15

## 壹、簡介

近年來由於資訊系統的普及，電子檔案的數量急遽增加，但由於資訊技術日新月異，電子檔案的儲存格式與儲存媒體幾乎是以三至五年為一個世代快速轉變，有鑑於此，如何確保被保存的電子檔案能持續地被存取與利用，已成為一個非常重要的課題。目前已有許多長期保存方法與策略被廣泛地探討，但仍以系統保存、模擬與轉置等為主要及常見之保存策略，其中又以轉置最廣為接受與使用。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，本指引針對電子檔案轉置技術，提供說明與相關的注意事項。

## 貳、適用範圍

本電子檔案轉置技術指引主要對象為符合機關檔案管理資訊化作業要點所訂之電子檔案格式表之各類型電子檔案。

## 參、轉置分類

### 一、媒體轉置

1. 所有的電子檔案都必須經由支援該電子檔案儲存格式的軟體解讀呈現並藉由媒體(載體)來儲存，如果媒體因故無法取得或讀取，儲存於其上的電子檔案也將無法存取與利用，因此儲存媒體的選擇在電

子檔案長期保存實務上相當重要，故在選擇儲存媒體時，應該針對各種儲存媒體的使用年限、保管條件、軟體支援程度、普及性、儲存空間、需求等等條件妥善評估。

## 2. 媒體轉置的時機如下：

- (1)媒體本身已屆保存期限。
- (2)應用系統改版或更新。
- (3)原媒體儲存空間不足。
- (4)媒體因外力或其他因素面臨損壞的風險，例如光碟片因故受潮嚴重，雖仍可讀取，但存在風險。
- (5)因為管理方式改變或應用系統修改或更新而必須進行資料結構調整，如資料夾異動或合併等。
- (6)維護成本考量，如舊儲存媒體容量較小體積較大，雖未到達保存期限，但為求維護方便，必須進行轉置。

## 3. 媒體轉置注意事項：

- (1)媒體轉置應建立標準作業程序與產出標準作業文件，以備日後查核並作為日後作業改善之依據。
- (2)媒體轉置後應進行檔案比對與總和檢查。
- (3)媒體轉置後之新儲存媒體應即刻依標準保管方式進行保管。
- (4)媒體轉置必須保持原媒體資料夾結構，如無法保持原資料夾結構時，必須詳載調整方式備查。
- (5)媒體轉置總和檢查結果、檔案雜湊值、轉置檔案數量、轉置時間等資料記載於詮釋資料。
- (6)所有與轉置過程中之標準文件與詮釋資料除須列印出紙本保管外，並應保存乙份於轉置後儲存媒體中。

## 二、格式轉置

(一) 電子檔案是依據某種特定的格式描述的資料，必須經由應用程式方能予以呈現，例如 Microsoft Office 中的 Word 檔案，其中記載了各段落的文字格式，必須依賴能解讀 Word 格式的應用程式才能將其呈現與輸出；各種應用程式也必需執行在作業系統之上，某些作業系統更只能執行在特定的硬體架構之上。

(二) 格式轉置的時機如下：

- (1)儲存的電子檔案格式為非開放格式。
- (2)隨著技術的發展，電子檔案格式可能過時，或因普及率不高，應用程式廠商可能停止提供支援或服務。
- (3)支援電子檔案之應用程式相依的作業系統或硬體架構無法繼續獲得。
- (4)由於長期保存的需要，必須將特定類別屬性的檔案轉置成特定格式以符合長期保存規格。
- (5)配合法令規定。

(三) 格式轉置注意事項

- (1)格式轉置前評估相關應用系統是否需要一併調整或更新。

- (2)格式轉置應建立標準作業程序與產出標準作業文件，以備日後查核並作為日後作業改善之依據。
- (3)轉置後之目的格式必須符合長期保存要求。
- (4)轉置品質檢驗。
- (5)應將轉置工具名稱、轉置工具版本、轉置參數、品質驗證結果及歷史版本資訊等轉置相關資訊納入詮釋資料中與轉置後檔案一併保存。

## 肆、轉置品質

無論是任何形式的轉置都存在資訊流失的風險，因此必須建立一個評量機制以確保轉置的品質。

評量媒體轉置品質可藉由以下幾個方式進行：

- (1)原始儲存媒體與轉置後儲存媒體檔案總數比對。
- (2)原始儲存媒體與轉置後儲存媒體檔案雜湊值比對。
- (3)是否產出標準作業文件。
- (4)如果有媒體封裝檔則以清查工具進行清查。

評量格式轉置品質的方式可分為客觀(Objective)與主觀(Subjective)兩種準則：

- (1)客觀準則：以計算轉置前後的量化指標來評估影像品質。
- (2)主觀準則：以人類肉眼的觀察來比較轉置前後影像品質。

以下為針對主客觀方式做詳細介紹：

### (一) 客觀評量

客觀評量著重現實，可提供準確量化值且可重複的評量結果，

如同電子閱卷，標準定義很清楚，也很容易識別對錯，但無法用來評估作文或申論。

常見的客觀評量方法有影像方面的 MSE、SNR、PSNR、S-CIELAB 色差評估模式、Structure Similarity (SSIM) 及語音的 SSNR 的評估方式，其中最廣泛被使用的客觀評量方法是 Peak Signal to Noise Ratio (PSNR)。

客觀的影像方面的評估方式如下：

### 1. 均方差 (Mean Square Error, MSE)

均方差為一般影像評估常用的方式之一，求出來的 MSE 值越小，表示輸入影像與複製影像之間的差異少，也就代表品質較好。均方差主要計算方式是將輸入影像以及複製影像的每一個像素點相減，再將其差值的平方加總起來取平均值，取得 MSE 值，其計算公式如方程式下所示。

$$MSE = \frac{1}{MN} \sum_{i=1}^m \sum_{j=1}^n [f(i,j) - \hat{f}(i,j)]^2$$

### 2. 訊噪比 (Signal to Noise Ratio, SNR)

訊噪比是指影像中信號與雜訊的比值，而雜訊的取得則是由複製影像以及輸入影像相減之後的差值而來，一樣是針對影像的每一個像素點來計算，其計算公式如下方程式。計算出來的比值，數值越大，表示雜訊越少，複製影像與輸入影像越相似，代表影像品質也越好。

$$SNR = \frac{\sum_{i=1}^m \sum_{j=1}^n [f(i,j)]^2}{\sum_{i=1}^m \sum_{j=1}^n [f(i,j) - \bar{f}]^2}$$

### 3. PSNR (Peak Signal to Noise Ratio)

PSNR 也就是峰值訊噪比，經常用作圖像壓縮等領域中信號重建質量的測量方法，它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，其計算公式如下方程式，其中  $(2^n - 1)^2$  代表是表示圖像點顏色的最大數值，如果每個像素點用 8 位表示，那麼就是 255。在利用 PSNR 算出來的比值越大，代表複製影像與輸入影像越接近，影像品質良好。

$$PSNR = 10 \log \frac{(2^n - 1)^2}{MSE}$$

### 4. S-CIELAB 色差評估模式

所謂色差公式，是在相同環境下，將人眼對於兩種色彩的視覺差異，以量化形式表示的計算公式。CIELAB 是為 CIE 協會所認可的標準色差公式之一。在這個色域空間內的色彩座標值可標示為  $L^*a^*b^*$ ，是經由 CIE 色彩三刺激值 XYZ 所轉換過來者。

在 CIELAB 色彩空間內，兩點之距離代表人眼感知對於色彩的色差。分別以  $\Delta L^*$ 、 $\Delta a^*$ 、 $\Delta b^*$ ，代表兩點在人眼感知中對於彩度軸以及非彩度軸之物理距離，但其計算方式是仍以整體的平均量來計算。然而人眼對於每個分量的色差敏感度若不相



同，便無法得知兩色彩真正的色差，因此不斷的有不同之不同  
 色差公式的發展。其 CIELAB 色差公式如下公式：

$$\Delta E_{ab}^* = [(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2]^{1/2}$$

$$\Delta E_{ab}^* = [(\Delta L^*)^2 + (\Delta C_{ab}^*)^2 + (\Delta H_{ab}^*)^2]^{1/2}$$

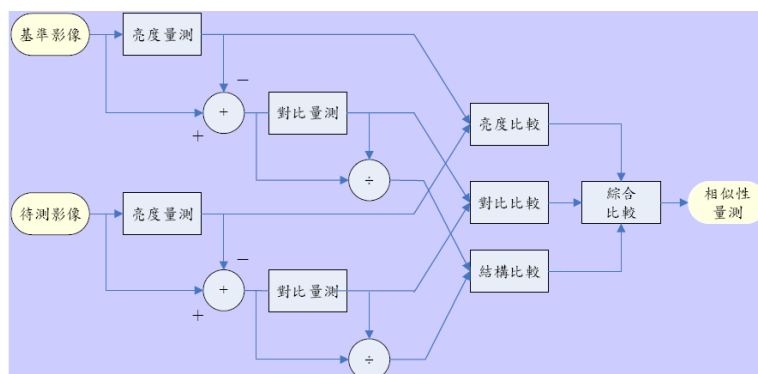
利用 CIELAB 色差公式預測單一色塊與單一色塊的色差預測是非常適用的。但是在色差公式發展成熟之餘，如何將現有色差公式應用於複雜影像則所要考慮的因素也就更多。因此，CIELAB 色差公式僅適用於單一均勻色塊的色差評估，因此加入空間域分佈的概念，將 CIELAB 色差公式延伸推展，而推導出 S-CIELAB 的色差評估模式。

## 5. 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的通用性影像品質指標 (UQI) 模型，而可以被視為一個完美的影像品質量測。

此一指標模型可以透過比較誤差敏感性理論 (Error Sensitivity Philosophy) 而得到瞭解。首先，誤差敏感性接近估計察覺錯誤確定影像降低的數量，考慮影像降低而察覺在架構性訊息變化的變化。其次，錯誤差敏感性理論由下而上的方式，相關功能類似早期 HVS 系統。而此型模是由上至下，模仿 HVS 的功能性，避開在以前部分提及臨界值的問題，因為它不倚賴臨界值確定察覺的變形數量。並降低相互作用問題，因此結構相似性指標主要在評估兩張影像間的結構變化。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$



結構相似性量測系統圖

傳統客觀品質指標是以統計的方法計算出有關於整張影像整體灰度值誤差的總合，而結構相似性指標不單僅是利用計算灰度值誤差的方法，並以模擬人眼視覺系統(HVS)，來進行影像品質評估結果，而由結構相似性指標計算結果卻可得到不同的指標值，並且得到符合人眼視覺的評估結果。

## 6. 通用性影像品質指標 UQI(Universal Quality Index)

通用性影像品質指標，對於各類型影像處理均可適用，設計上是以任何影像失真模型均含有三種因素：相關係數降低、亮度改變和對比改變。

模型評估影像的最佳及最差的量化值均介於-1 至+1 之間，因此可得到比較具體的評估值。

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

客觀的語音評估方式：

## 1. SSNR

由於語音壓縮的方法都是將一長段語音切成一段小段的音框，因此最常被使用在語音品質評估方法就是 SSNR(segmental signal-to-noise ratio)，這個方法是以音框為單位，先計算出每一音框的 SNR(signal-to-noise ratio) 值，最後再把各個音框的 SNR 值加總平均出來即為 SSNR 值。假設原始的語音訊號為  $s(k)$ ，經過運算處理後的語音訊號為  $\hat{s}(k)$ ，且一個音框有  $N$  個語音取樣訊號 (sample)、共有  $M$  個音框的話，則 SSNR 的計算如下中的  $i$  為音框的計數器， $k$  為音框中的語音取樣訊號計數器。

$$SSNR = \frac{1}{M} \sum_{i=1}^M \left( 10 \log_{10} \left( \sum_{k=1}^N \frac{s(k)s(k)}{(s(k) - \hat{s}(k))^2} \right) \right)$$

SSNR 方法的好處是簡單，並且可以明確判斷語音波形的相似度，SSNR 值越高語音品質越好。

## (二) 主觀評量

主觀評量考量人類主觀感受，而影響人類主觀感受的因素相當地多且難以歸納與量化統計，因此通常比較複雜且結果可能會被認為不太精確；如前例，主觀評量就是類似尋求一個電子批閱作文或論文的方法。

最廣泛被使用的客觀評量方法是影像方面的心理物理實驗及語音的 MOS(Mean Opinion Score)、PSQM (Perceptual Speech Quality Measurement)

主觀的影像評估方式：

### 1. 心理物理實驗

心理物理實驗是一種影像對於人眼感知的刺激，是透過觀測者以人眼觀測影像來評比影像的品質，其目的在於研究色彩的「物理刺激值」與「感知」之間的關係，色彩影像是人眼視覺的一種感知，一般對於影像色彩的都是以數值為依據，而色彩除了量化的數值外，也應該考慮到人眼視覺的感受，如何將人眼視覺的感知轉換成量化的數值，也就是心理物理實驗的目的。

1927 年 Thurstone 提出配對比較法定律，此模式是針對物理刺激強度的比較，將心理物理實驗分成兩種方式，第一種為「偏好色」評比模式(Preference)，此模式是讓觀測者依自己的喜好選擇影像偏好的影像或色彩，這種方式進行觀測的人員不一定需要具備色彩訓練的基礎即可進行實驗，第二種為「正確色」評比模式(Accuracy)，此模式則是在實驗過程中，根據原稿之正確性來選出自己認為最接近原稿的複製影像或色彩，這種方式的觀測者必對於色彩學有基本的認知與訓練，才能針對原稿與複製影像或色彩做出更為精準的評估標準及結果。在實驗過程中，需記錄各實驗所測試的數值，最後使用統計分析 z-score 來評估影像的品質。如下圖所示。



圖 心理物理實驗

主觀的語音評估方式：

## 1. MOS(Mean Opinion Score)

平均意見分數 MOS(Mean Opinion Score)是由 ITU-T P. 800 所提出的一種對語音品質好壞的主觀評定方法。根據 ITU-T 建議的要求，特定的發話者與聽話者在特定的環境下，透過收集測試者在各種不同情景下的主觀感受，再根據建議的分析法則得出該語音的品質。ITU-T 對測試的要求非常嚴格，所有的操作都要嚴格地服從操作流程，對錄音系統、語音採樣、語音輸入級別、聽者級別、不同發話者、多發話者、差錯處理、不同語音編碼模式的兼容性、過失、環境噪音、音樂等等，都作出了詳細嚴格的規定。測試者的主觀感受結果也被分為很多不同的範疇，如聽者感受的網路品質結果、品質降級結果、比較結果等。

## 2. PSQM (Perceptual Speech Quality Measurement)

MOS 方法是一種模糊的評估方法，完全是依據人的感覺來評斷語音的好壞，其測試結果很難對 VoIP 系統的改進和不同 VoIP 設備之間性能的比較作出有實際意義的判別。因此，有人提出 PSQM (Perceptual Speech Quality Measurement)方法，用來作為客觀語音度量與評估。

PSQM 仍以 MOS 的五個級別作為標準，所不同的是其對每一個級別都以百分比的模式做出了差或最差（%PoW =Percent Poor or Worse）和好或最好（%GoB = Percent Good or Better）的進一步描述。而 PSQM 方法並未擺脫原始的人類主觀評估，只是作了進一步的說明，但目前有些通話品質的測試儀器，他就是利用電腦產生波型檔案（Wave File），透過比較透過 VoIP 網路傳輸前後聲音波形的變化，計算出與 PSQM 中相對應的級別及好壞程度，以此作

為評估語音品質的方法。

## 伍、轉置作業程序

### (一) 確認轉置對象

藉由人工分類或系統自動比對搜尋轉置對象，整理轉置清單並擬定轉置計畫。

### (二) 進行媒體複製

將欲轉置資料進行複製產生複本，利用檔案個數比對、總和檢查、檔案雜湊值比對等方法檢測轉置資料複本是否與原始檔案相同。

### (三) 進行安全檢測

使用病毒掃描軟體與惡意軟體掃描工具針對轉置資料複本進行檢測，以確保系統安全與資料安全。

### (四) 進行格式檢測

針對轉置資料複本檔案使用電子檔案長期保存運作平台進行格式檢測以確定原始檔案格式；如來源檔案為數位內容封裝檔則同時建行封裝檔格式檢測，以檢測封裝檔版本與正確性。

### (五) 確認轉置格式

參考電子檔案長期保存知識庫系統以選擇適當的轉置格式，或利用電子檔案長期保存運作平台取得相關說明。

### (六) 格式轉置

使用電子檔案長期保存運作平台進行電子檔案轉置作業。

## (七) 品質驗證

如為媒體轉置則使用檔案個數比對、總和檢查、檔案雜湊值比對或媒體封裝檢查進行品質驗證；如為格式轉置則使用電子檔案長期保存運作平台進行品質驗證。

## (八) 詮釋資料

在轉置的過程中擷取來源檔案屬性、病毒掃描結果、轉置工具、轉置工具版本、轉置時間、轉置檔案個數、格式檢測結果、原始檔案格式、轉置後檔案格式及工作執行紀錄等相關資料做成詮釋資料並與轉置後資料一併封裝保存。

## (九) 原儲存媒體與暫存媒體消磁或銷毀

完成轉置並通過品質驗證後，原儲存媒體與轉置過程中使用之暫存媒體應予以消磁，如原媒體過保存期限或無法繼續使用應予以銷毀。