

# 紙本檔案數碼化技術規範性指引

## 1. 標的及目的

1.1. 《紙本檔案數碼化技術規範性指引》（下稱“《指引》”）訂定澳門特別行政區公共部門在實施紙本檔案數碼化工作時應遵守的基本要求及指導性規定，包括主要工作環節、管理要求及技術指標等項。

1.2. 《指引》旨在建立紙本檔案數碼化工作的規範，當中包括對數碼化工作的流程、技術指標和圖像格式的規範，以便數碼化工作的質量控制、資訊交流，以及建設和積累澳門特別行政區政府的檔案資料。

## 2. 適用範圍

2.1. 《指引》適用於澳門特別行政區公共部門在下列情況進行的數碼化工作：

2.1.1. 按部門的需要，對永久保存的紙本檔案進行數碼化。

2.1.2. 將紙本檔案經數碼化後的檔案納入到部門運作系統中，以支援日常活動的開展。

2.2. 《指引》不適用於原生電子檔案的處理。

## 3. 參考標準

《指引》的制訂參考下列標準：

3.1. GB/T 20530-2006 文獻檔案資料數字化工作導則（中華人民共和國國家標準）。

3.2. DA/T 31-2005 紙質檔案數字化技術規範（中華人民共和國檔案行業標準）。

3.3. ISO/TR 13028:2010 *Information and documentation – Implementation*

*guidelines for digitization of records*。

3.4. ISO/TR 15801:2009 *Document management – Information stored electronically – Recommendations for trustworthiness and reliability*。

## 4. 技術用詞

### 4.1. 數碼化 (*digitization*)

用電腦技術將模擬訊號轉換為數碼訊號的處理過程。

### 4.2. 紙本檔案數碼化 (*digitization of paper-based records*)

利用掃描器或數碼相機等設備，將紙本檔案轉化為能被電腦識別，並可儲存在磁盤、磁帶、磁片、光碟等電腦載體上的數碼圖像處理過程。

### 4.3. 數碼圖像 (*digital image*)

紙本檔案經過數碼化處理後所形成的圖像。

### 4.4. 黑白二值圖像 (*binary image*)

黑白兩級灰度的數碼圖像。

### 4.5. 連續色調靜態圖像 (*continuous-tone still image*)

以多於兩級灰度的不同濃淡層次或以不同顏色組合成的靜態數碼圖像。

### 4.6. 解像度 (*resolution*)

單位長度內數碼圖像包含的點數或像素數，一般用每英寸點數 (*dpi*) 表示。

### 4.7. 失真度 (*distortion measure*)

對紙本檔案原件進行數碼化後，在同等測試環境下，數碼圖像與紙本檔案原件在色彩、幾何等方面的偏離程度。

### 4.8. 可懂度 (*intelligibility*)

數碼圖像向使用者或設備提供訊息的能力。

#### **4.9. 圖像壓縮 (*image compression*)**

清除數碼圖像內冗餘或對數碼圖像進行近似的處理，其目的是對數碼圖像以更緊密的形式表示。

### **5. 紙本檔案數碼化應遵循的原則**

#### **5.1. 基本原則**

在進行紙本檔案數碼化工作時，應因應第 2.1.點所指的情況，選取適當的技術指標，符合紙本檔案數碼化的基本要求，以便準確、方便、快捷及安全使用檔案資源。

#### **5.2. 評估原則**

在規劃開展紙本檔案數碼化工作前，應先評估數碼化的可行性及風險，考量該等紙本檔案的狀況、所需的資源和設備等因素。

#### **5.3. 鑒定原則**

在規劃開展紙本檔案數碼化工作前，應先對檔案進行分類及價值鑒定，以便從第 2.1.點所指的情況中，確定最為適合的情況。

#### **5.4. 合法原則**

數碼化工作應遵守澳門特別行政區現行法例的規定。

### **6. 紙本檔案數碼化基本要求**

#### **6.1. 工作規劃**

因應第 2.1.點所指的情況，對數碼化工作流程、數碼化影像的儲存載體、技術規格等進行規劃，並對數碼化工作所需投入的時間、人力資源和經費進行安排，以便制訂一個完整的規劃。

#### **6.2. 基本環節**

紙本檔案數碼化的基本環節主要包括：檔案整理、目錄建庫、檔案掃描、圖像處理、圖像儲存、數據質檢、數據掛接、數據驗收、數據備份及成果管理。

### **6.3. 程序管理**

紙本檔案數碼化程序各個環節進行的所有工作，應作出記錄，以獲得該程序的完整紀錄。

### **6.4. 安全措施**

6.4.1. 為確保檔案原件和數碼化檔案訊息的安全，須建立並完善紙本檔案數碼化各環節的安全保密管理機制。

6.4.2. 用於紙本檔案數碼化的設備應盡可能與外界分隔。

## **7. 檔案整理**

為確保數碼化工作的質量，在進行紙本檔案數碼化前，須按下列步驟對檔案進行整理。

### **7.1. 拆除裝訂**

當檔案裝訂物影響掃描工作進行或影響掃描品質時，應拆除裝訂物；在拆除裝訂物時，須注意保護檔案不受損害。

### **7.2. 頁面修整**

7.2.1. 對破損嚴重、無法進行數碼化的檔案，應先進行技術修復。

7.2.2. 對皺摺不平或影響掃描品質的檔案頁面，在進行掃描前，應先進行適當處理，將有關頁面壓平或熨平。

### **7.3. 裝訂還原**

7.3.1. 掃描工作完成後，已被拆除裝訂物的檔案應按檔案保管的要求重新裝訂。

- 7.3.2. 在恢復裝訂過程，應保持檔案頁面的排列順序不變，並確保有關處理工作安全、準確、無遺漏。

## **8. 目錄建庫**

### **8.1. 目的**

目錄數據庫的建立旨在加強對數碼圖像的管理，以便日後使用。

### **8.2. 數據格式選擇**

在建立目錄數據庫時，應選擇通用且能直接或間接通過 XML 文檔進行數據交換的數據格式。

### **8.3. 資料輸入**

- 8.3.1. 數碼化的檔案應記錄在目錄數據庫。

- 8.3.2. 對於尚未數碼化的紙本檔案，應按照訂定的輸入規範，在目錄數據庫中輸入相關的描述資料，以便日後掛接使用。

### **8.4. 品質檢查**

- 8.4.1. 應對建立的目錄數據庫進行品質檢查，並核對描述的項目是否完整、有關內容是否符合既訂要求及準確。
- 8.4.2. 上點所指的檢查可透過人工校對或軟件自動校對為之。
- 8.4.3. 如發現有不準確的數據，應作出必要的修改或重新輸入數據。

## **9. 檔案掃描**

### **9.1. 掃描方式**

- 9.1.1. 應根據紙本檔案原件幅面的大小，選擇適當規格的掃描

設備。

9.1.2. 對大幅面的紙本檔案，可採用分區掃描及圖像拼接方式處理，或可採用大幅面數碼平台拍攝。

9.1.3. 紙張狀況較差、過薄、過軟或超厚的紙本檔案，採用的掃描方式應避免損壞原件。

9.1.4. 紙張狀況良好的紙本檔案方可採用高速掃描方式，以提高工作效率。

## **9.2. 掃描色彩模式**

9.2.1. 掃描色彩模式一般有黑白二值、灰度、彩色。

9.2.2. 頁面為黑白兩色，並且字跡清晰、不帶插圖的紙本檔案，應採用黑白二值模式進行掃描。

9.2.3. 頁面為黑白兩色，但字跡清晰度差或帶有插圖的紙本檔案，以及頁面為多色文字的紙本檔案，應採用灰度模式掃描。

9.2.4. 頁面中有印章或插有黑白照片、彩色照片或彩色插圖的紙本檔案，應採用彩色模式進行掃描。

## **9.3. 掃描解像度**

9.3.1. 掃描解像度的選擇，不應影響圖像的清晰度、完整性及使用。

9.3.2. 如字體偏小、密集、清晰度較差等特殊情況下，可適當提高解像度。

## **9.4. 掃描記錄**

應記錄檔案掃描及頁數，核對該頁數與原件頁數是否一致，而在不一致時，應註明具體原因和處理方法。

## **10. 圖像處理**

### **10.1. 方向調整**

對方向不正確的數碼圖像進行旋轉還原，以符合閱讀習慣。

### **10.2. 糾偏**

對出現偏斜的數碼圖像進行糾偏處理。

### **10.3. 去污**

對數碼圖像中出現的影響數碼圖像品質的雜質，如黑點、黑線、黑框、黑邊等進行去污處理。

### **10.4. 畫質調整**

對掃描後畫面品質不理想的數碼圖像進行相應參數的調整，如亮度、對比度、飽和度、色階、色彩平衡等，處理過程中應遵循不影響可懂度的前提下展現檔案原貌的原則。

### **10.5. 圖像拼接**

10.5.1. 對大幅面的紙本檔案文件進行分區掃描時，各分區的大小尺寸應相近，而解像度和曝光量應一致。

10.5.2. 相鄰數碼圖像之間必須有部分重疊，重疊面積不少於分區面積的二十分之一。

10.5.3. 掃描後形成的多幅數碼圖像應進行拼接處理，合併為一個完整的數碼圖像，以保證數碼圖像的完整性。

### **10.6. 裁邊處理**

對數碼圖像進行裁邊處理，去除多餘的白邊，有效減少數碼圖像的容量，節約儲存空間。

### **10.7. 數碼圖像處理許可及記錄**

第 10.3.至 10.6.點所指的圖像處理須預先經部門領導許可後，方可執行，且應即時進行記錄，並註明具體原因、已作出的處理和處理人

員等內容。

## **11. 圖像儲存**

### **11.1. 儲存格式**

11.1.1. 數碼圖像應儲存為 TIFF 格式，並選擇不壓縮或無損壓縮模式。

11.1.2. 供網絡查閱的數碼圖像可儲存為 PDF、JPEG 或其他格式。

### **11.2. 數碼圖像的命名**

每份紙本檔案應有唯一對應的文件號，並應以該文件號作為基礎，為該紙本檔案掃描後產生的數碼圖像命名，且數碼圖像的命名應具一致性。

## **12. 數據質檢**

### **12.1. 目錄數據庫品質檢查**

應檢查目錄數據庫內容的完整性及該內容是否符合相關規範，並對不準確的數據進行修改或重新輸入。

### **12.2. 數碼圖像品質檢查**

12.2.1. 對方向不正確，偏斜度、清晰度、失真度等不符合既訂質量要求的數碼圖像，應重新處理或重新掃描。

12.2.2. 當發現漏掃任一頁時，應即時補掃並正確插入遺漏的數碼圖像。

12.2.3. 掃描產生的數碼圖像的排列順序與檔案原件不一致時，應即時進行必要的調整。

### **12.3. 質檢記錄**



應記錄數據質檢，指出質檢結果和處理建議。

### **13. 數據掛接**

13.1. 為實現目錄數據庫與數碼圖像的掛接，應透過目錄數據庫中的文件號與數碼圖像名稱的一致性和唯一性，建立對應的關聯關係。

13.2. 當目錄數據庫和數碼圖像符合既訂質量要求，應編製或使用適當的軟件，實現目錄資料與相關聯的數碼圖像的自動掛接。

13.3. 應記錄數據掛接，加入數據掛接後的頁數，核對該頁數與掃描時填寫的頁數是否一致，如不一致時，應註明具體原因和處理方法。

### **14. 數據驗收**

倘數碼化工作由外判商執行或對數據有特定驗收要求，有關驗收須按下列規定進行：

#### **14.1. 數據抽檢**

14.1.1. 應以抽檢的方式檢驗已完成數碼化的資料的品質。

14.1.2. 數據抽檢以種類為單位進行，抽檢率不得低於 5%，並按以下公式計算：

$$\text{抽檢率} = \text{抽檢文件數} / \text{待驗收種類的文件總數} \times 100\%$$

#### **14.2. 驗收標準**

14.2.1. 如抽檢的目錄數據庫、數碼圖像不符合既訂品質要求，或目錄數據庫與數碼圖像的掛接存有瑕疵，有關檢驗視為不合格。

14.2.2. 以種類為單位進行抽檢的數據，合格率達 95% 或以上者，有關檢驗視為合格，並可接收數碼化工作。

14.2.3. 沒有通過檢驗的數據則退回質檢環節進行檢查及修正。

14.2.4. 第 14.2.2.點所指的合格率按以下公式計算：

$$\text{合格率}=\text{抽檢合格的文件數}/\text{抽檢文件總數}\times 100\%$$

### **14.3. 數據驗收記錄**

應詳細記錄數據驗收，尤其是合格率和驗收結果，而對驗收不合格者，應註明不合格原因和後續的處理。

## **15. 數據備份**

### **15.1. 備份範圍**

對目錄數據庫和數碼圖像應適時進行備份。

### **15.2. 備份方式**

為保證數據安全，備份載體的選擇應多樣化，可採用在線、離線相結合的方式實現多套備份，如需要時可進行異地備份。

### **15.3. 備份數據校驗**

備份數據須定期進行校驗，校驗內容主要包括：備份數據能否讀取、數據訊息的完整性及數碼圖像數量的準確性。

### **15.4. 備份標籤**

數據備份後，應在相應的備份介質上作出標籤，以便查找和管理。

### **15.5. 備份記錄**

應記錄數據備份，指出備份數量、備份日期和存放位置。

## **16. 成果管理**

16.1. 應加強對紙本檔案數碼化成果的管理，確保其安全、完整和長期可讀。

16.2. 為保證數據的長期可讀，應訂定數據備份政策，內容應包括與之相關連的元數據。

16.3. 數據的保管期限應與其原件的保管期限相同。

16.4. 數據的處置應按照紙本檔案的處置程序執行。

16.5. 當數碼化檔案提供網上檢索使用時，可根據具體情況對數碼圖像作出標識，並限制其下載或列印。

## 17. 技術指標

17.1. 如屬第 2.1.1.點所指的情況，應採用表一的技術指標，並允許衍生等於或低於該技術指標的複製品。

表一

掃描色彩模式	解像度	色深	格式	壓縮
黑白二值	最低 300 <i>dpi</i>	1 <i>bit</i>	TIFF	無損壓縮
灰度	最低 300 <i>dpi</i>	8 <i>bit</i> 灰階	TIFF	無損壓縮
彩色	最低 300 <i>dpi</i>	最低 24 <i>bit</i> 彩色	TIFF	無損壓縮

17.2. 如屬第 2.1.2.點所指的情況，應採用表二的技術指標。

表二

掃描色彩模式	解像度	色深	格式
黑白二值	最低 200 <i>dpi</i>	1 <i>bit</i>	- TIFF - PDF - JPEG
灰度	最低 200 <i>dpi</i>	8 <i>bit</i> 灰階	- TIFF - PDF - JPEG
彩色	最低 200 <i>dpi</i>	最低 8 <i>bit</i> 彩色	- TIFF - PDF - JPEG