

多媒體檔案管理

陳立原

國家圖書館管理師

【摘要】

多媒體技術為目前資訊科技發展重點項目之一，其在圖書館實際應用則為數位典藏系統，本文主要討論當建置及營運一個數位典藏系統時，需如何管理其多媒體檔案，從多媒體檔案製作、命名到資料儲存及數位典藏系統的連結等各方面有系統進行討論，並對目前正在研發的多媒體技術做一介紹。

關鍵詞：多媒體檔案、數位典藏系統、資料儲存、檔案命名機制、系統連結

一、前言

人類文明的急速發展主要是源由智識的不斷累積，就像英國科學家牛頓說過的一句話：「我之所以有今天的成就，是因為我站在巨人之上，看得比別人遠」，故人類的智識保存和廣泛傳播實為人類文明發展的重要因素，而圖書館在此扮演著不可取代的地位。

當科技發明對資料典藏產生進化時，人類的文明更會加速進化。如印刷術的發明使資料能大量簡便複製，使得資訊能廣泛傳播；錄放映技術的發明使資料除平面文字外，更包含了連續動作及聲音的記錄，使得資訊的內涵更為豐富；電腦的發明可將各種資料轉為數位化，而透過數位化的檔案使其資訊複製成本急速下降；網際網路的普遍使用，使資料的傳播能無距離的限制，即時的傳送到需求者手上。而圖書館在面臨科技的進步，為求更能提供原有的服務目的，使圖書館必須走向數位圖書館的道路，而數位圖書館成功與否的關鍵，則在於多媒體檔案管

理機制的良窳。

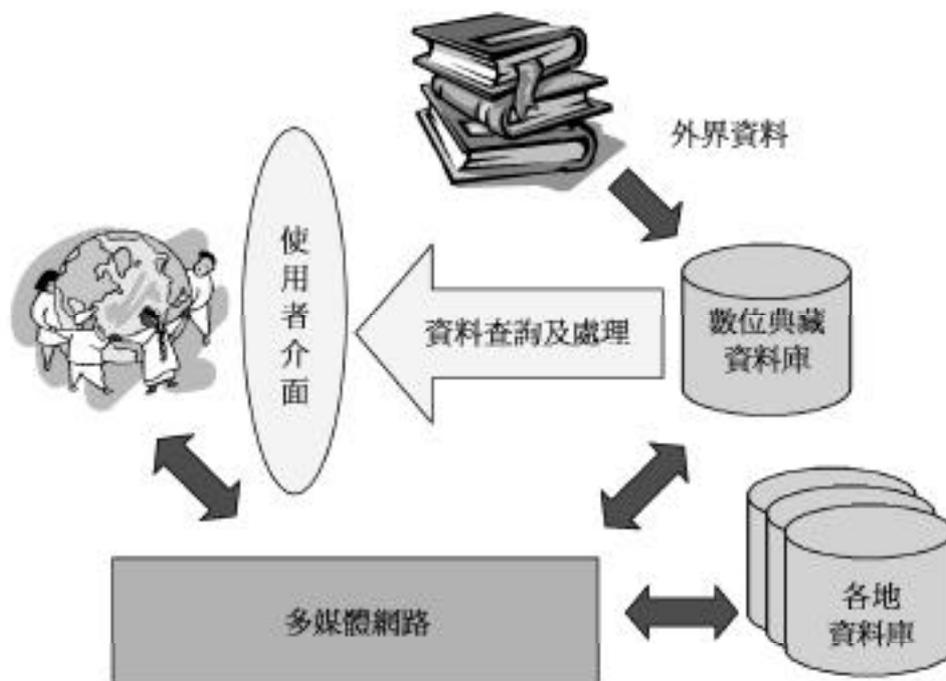
多媒體檔案的廣泛定義可包括圖書館所有館藏，如書本、錄影帶、錄音帶、VCD、DVD、圖畫、照片、數位檔案等各種媒介，而本文所討論的多媒體檔案將限於數位化檔案，至其管理則包含數位典藏系統的應用。

二、多媒體檔案管理與數位圖書館

各種文獻對數位圖書館 (digital library) 的定義有不同的說法，而在眾多說法中筆者認為 US-EU 研究小組所定義最為傳神，該小組對數位圖書館的定義為「數位圖書館乃在整合並使用計算、通訊及大量的數位化資料，建立一個新的基礎架構及環境」，在這定義中有幾個重點：

1. 計算：資訊科技促使整個圖書館日新月異，邁向數位圖書館。
2. 通訊：數位圖書館是要提供公眾迅速、方便的使用方式。而通訊技術的蓬勃發展，尤其是網際網路的成長，使數位典藏能廣泛提供給民眾使用。
3. 數位化資料：為數位圖書館的基礎。將各種資料數位化，若無數位化資料則像圖書館無書本一樣，成為一個空殼子。
4. 新的基礎架構及環境：數位圖書館是一個新興事物，需要有政治、法律、習慣等方面來配合才能成功，其與傳統圖書館有三個重要差異：
 - (1) 傳統圖書館的大小一般是依藏書量多寡為判斷的基準，而數位圖書館則是使用率，所以使用者介面的設計非常重要。
 - (2) 圖書館提供的是資料，數位圖書館提供的是資訊，故需要有效的檢索功能設計。
 - (3) 圖書館是被動，數位圖書館是主動，資訊科技如何妥善應用在這方面則為當前的重要課題。

從數位圖書館定義中得知，數位化資料為數位圖書館的基礎，多媒體檔案管理即為管理數位化資料，本文中用多媒體檔案取代數位化資料則是強調數位化資料的來源種類，多媒體係指包含文字 (Text)、視訊 (Video)、動畫 (Animation)、圖形 (Graphics)、影像 (Images)、音訊 (Audio) 等多種媒介 (Media)。多媒體檔案管理若從廣義的定義包含了整個數位圖書館的功能，而狹義而言則包含多媒體檔案典藏、儲存及存取使用，在數位典藏系統概念圖(圖一)中，多媒體檔案管理則包含外界資料和數位典藏資料庫兩大區域。



圖一：數位典藏系統概念圖

故多媒體檔案管理包含三個部分：

1. 多媒體檔案製作：包含數位化程序和多媒體檔案格式，其重點是如何使多媒體檔案能充分保全原實體典藏風貌。
2. 多媒體檔案儲存：依使用目的分為永久保存儲存和系統應用儲存。
3. 多媒體檔案應用：提供讀者使用時的應用程式與多媒體檔案的關聯，故必須考慮檔案的命名及檢索系統存取方式。

三、多媒體檔案製作

在數位化時如何能把原物件忠實記錄是非常重要的課題，在實體和數位檔案的轉換中有兩個重要關鍵，一為數位化程序，另一為數位化後的多媒體檔案格式，兩者是相輔相成、缺一不可，若數位化程序做不好，所保存的多媒體檔案與實體不符，則是對使用者傳達錯誤的資訊；若多媒體檔案格式選擇不適宜，則會使檔案表達缺乏原物件的資訊或造成使用不便。

(一) 數位化程序

在將物件數位化過程中有三個要注意的重點，分別為數位化媒介、數位化方式及數位化設備。

1. 數位化媒介：選擇要數位化的媒介，除了直接使用原物件外，有時為了保存原物件或因其他因素不能直接將原物件數位化，而需要先轉為其他媒介後再進行數位化轉換，如美術作品數位化時可先行用相機拍照，轉成正片，再用正片去做數位化，而在決策要用正片、負片、原件或用數位相機拍攝時，就要考慮到對原物件本身影響、數位化目的、經費及資訊科技配合等因素。
2. 數位化方式：擬定原物件轉成數位化過程，如欲將貝殼數位化，則可設定將一個貝殼拍成正反兩張照片，再用照片做數位化，而其拍攝的張數、角度、距離及光線都需要明定之。
3. 數位化設備：決定好要使用那一種媒介去做數位化，同時考慮用何種機器設備如掃瞄器、數位相機、數位攝影機等設備，除硬體設備外還應考慮軟體如中文辨識系統、語音合成等技術，才能充分掌握原物件資訊。在選擇機器設備時有多個功能要項要考慮，如以掃瞄器為例，其軟硬體必須考慮項目：
 - (1) 媒介掃瞄可接受種類：透射稿或反射稿。
 - (2) 掃瞄區域大小：A4、A3。
 - (3) 媒介掃瞄方式：平臺式、輪轉式。
 - (4) 色彩濃度大小：濃度範圍表示了影像所包含的「最暗」至「最亮」範圍，所表現就是能夠展示的層次範圍與色彩空間，一般為 3.2D 至 3.4D。
 - (5) 色彩模式：RGB、CMYK 等。
 - (6) 影色修正功能：是否具有濃度、黑點 / 白點、亮度 / 對比、Gamma 曲線等功能。
 - (7) 色彩管理功能：包含色彩管理系統、反射稿校色片、透射稿校色片等。
 - (8) 輔助掃瞄配件：自動文件送紙器、幻燈片掃瞄框、長條底片掃瞄框等。

(二)多媒體檔案格式

各種資料數位化時，必須考量到使用者的設備、使用的便利性、資訊檢索的需求、網路上資料的傳輸速度、資料的保存時限等問題，而分別建立其所需要的

檔案格式。

各種類型數位化後的解析度選擇以其原件的品質與數位化目的為依據，一般印刷品及照片的品質為 250dpi 至 300dpi。若要用數位檔來做重製使用，則須以原件原始尺寸取 300dpi 以上的解析度來計算（即若原件為 8in × 10in 拍成 4in × 5in 則解析度須設為 600dpi 或放大兩倍後掃描解析度為 300dpi）。

民國 90 年國家圖書館修訂圖書館法^[1]，其中有建議各種數位化檔案的規格。檔案種類以媒體類型而言，可分為文字、影像、聲音、視訊等四種型態（美國國會圖書館則多增加一項地圖類型，其檔案壓縮格式採 Mrsid）。

1. 文字資料

文字資料可分兩大類，第一類是以電腦打字的文字檔，如 word、pagemaker 或 HTML 等檔案，因本身就是一個數位檔，不需再將原件數位化，但為能使檔案規格公眾化，可轉存成一般公眾語法，如 TXT、HTML、XML 及 PDF 多種檔案格式，以便流通和閱讀者使用。

第二類為手稿或印刷資料，如需做全文檢索，則要重新打字或 OCR（optical character recognition），其檔案格式可與第一類相同；若不做全文檢索，即可將原件掃描，此時該文字資料經掃描後，可依目的存成下列格式：

檔案目的	說明	建議檔案規格
資料永久保存格式	將資料數位化典藏，保持原有風貌	檔案格式：TIFF 色調深度：灰階—每像素 8-bits；彩色—每像素 24-bits 壓縮：不壓縮或非破壞性壓縮 解析度：300-600 或更高 dpi（依原始資料品質及重要性選擇適當解析度，一般印刷品可採 300dpi）
資料服務 / 參考格式	提供使用者網路上觀看及列印用	檔案格式：JBIG or JBIG2 色調深度：灰階—每像素 8-bits；彩色—每像素 24-bits 壓縮：JPEG（灰階壓縮比約 10:1，彩色壓縮比約 10:1） 解析度：300 dpi
預覽影像	提供使用者預覽及選擇欄位用	檔案格式：GIF 色調深度：每像素 8-bits 壓縮：原生影像至 GIF 影像大小：150 × 100 到 200 × 200 pixels

[1] 圖書館法—數位圖書館資訊系統相關技術規範 (2-2—2-4)

上述色調深度為每一掃描點所存放的資訊大小，如色調深調為8bit 則整張圖檔只能以 256 (2 的 8 次方) 種顏色表示；解析度為每單位長度取點的數目，如解析度為 300dpi 則是在 1 inch 內取點 300 個。

2. 影像資料

原始資料為照片、圖片、地圖、底片、幻燈片等，則可用掃描器或數位相機等設備，依目的可轉存成下列檔案格式：

檔案目的	說明	建議檔案規格
資料永久保存格式	將資料數位化典藏，保持原有風貌	檔案格式：TIFF 色調深度：灰階—每像素8-bits；彩色—每像素24-bits 壓縮：不壓縮或非破壞性壓縮 解析度：300-600 dpi (依原始資料品質及重要性選擇適當解析度，一般印刷品可採300dpi)
資料服務 / 參考格式	提供使用者網路上觀看及列印用	檔案格式：JFIF (JPEG 交換格式) 色調深度：灰階—每像素8-bits；彩色—每像素24-bits 壓縮：JPEG (灰階壓縮比約10:1，彩色壓縮比約20:1) 解析度：300 dpi
縮圖影像	提供使用者預覽及選擇欄位用	檔案格式：GIF 色調深度：每像素8-bits 壓縮：原生影像至GIF 影像大小：150 × 100 到 200 × 200 pixels

上述影像大小為整個圖檔的長寬點數。

3. 聲音資料

原始資料為錄音帶、膠質唱片等媒介時，則先要找到可播放媒介的器材，然後透過音效卡等工具，轉成數位檔案，而原始資料為CD 時，則可直接由光碟機經軟體轉換，依目的存成下列檔案格式：

檔案目的	說明	建議檔案規格
資料永久保存	將資料數位化	檔案格式：WAVE (Microsoft 格式)

存格式	典藏, 保持原有風貌	取樣率: 44.1 KHZ 樣本位元數: 16 bit/sample 聲道: 立體音
串流 (streaming)	寬頻格式	提供以寬頻方式上網的使用者線上聽取 檔案格式: RA 格式 編碼種類: 雙頻ISDN 立體音 取樣率: 16KHZ 頻寬: 100Kbps
		檔案格式: MP3 取樣率: 44.1KHZ 聲道: 立體音 頻寬: 128Kbps
串流 (streaming)	窄頻格式	提供以窄頻方式上網 (撥接上網) 的使用者線上聽取 檔案格式: RA 格式 編碼種類: ISDN 立體音 取樣率: 8KHZ 頻寬: 50Kbps
		檔案格式: MP3 取樣率: 22.05KHZ 聲道: 立體音 頻寬: 56Kbps

上述取樣率為每單位時間的樣本數，如取樣率 16KHZ 為每秒鐘對音波取 16,000 個資料點；樣本位元數為每一資料點的資料大小，如樣本位元數 16bit/sample 為每個資料點用 16bit 來表示；聲道可分為單音、立體音、多重音道；串流 (streaming) 檔案提供即時播放，不需要檔案全部下載才能執行，但系統須架設串流伺服器。

4. 視訊資料

原始資料為錄影帶等媒介，先找到可播放媒介的器材，然後透過視訊擷取軟體將之轉為數位視訊檔，並可存成下列檔案格式。

檔案目的	說明	建議檔案規格
資料永久保存格式 (1)	將資料數位化典藏，保持原有風貌，VCD 的品質，資料量較少	檔案格式：mpg 壓縮方式：MPEG-1 圖片大小：352 × 240 像素以上 視訊解碼畫面變化率：29 個畫面 / 秒以上 音效解碼為立體雙聲道 CD 音質 資料傳輸率 (Data rate)：約 1.2 megabits / 秒 (約 150 kilobytes / 秒) Jitter 標準畫面時間 9mSec 或更優 平均同步位移 1mSec 或更優 標準裝置同步位移 15mSec 或更優 音效資料串流 nChannels：2 或更優 音效資料串流 nSamplesPerSec：44,100 或更優
資料永久保存格式 (2)	將資料數位化典藏，保持原有風貌，DVD 的品質，資料量為格式 (1) 的 3-100 倍，適合需高解析保存物件	檔案格式：mpg2 壓縮方式：MPEG-2 圖片大小：720 × 480 像素以上 視訊解碼畫面變化率：29 個畫面 / 秒以上 音效解碼為立體雙聲道 資料傳輸率 (Data rate)：約 6-8 Megabits / 秒 (約 150 kilobytes / 秒) Jitter 標準畫面時間 9mSec 或更優 平均同步位移 1mSec 或更優 標準裝置同步位移 15mSec 或更優 音效資料串流 nChannels：2 或更優 音效資料串流 nSamplesPerSec：44,100 或更優
串流 (streaming) 寬頻格式	提供以寬頻方式上網的使用者線上觀看	檔案格式：ASF Bit-rate (位元傳輸率)：250Kbps Audio Format：22 KHZ, stereo 檔案格式：RM Bit-rate (位元傳輸率)：250Kbps

串流 (streaming) 窄頻格式	提供以窄頻 方式上網的 使用者線上 觀看	Audio Format : 22 KHZ, stereo
		檔案格式 : mpg4
		壓縮方式 : MPEG-4
		音效解碼為MP3
		Bit-rate (位元傳輸率) : 250Kbps
		檔案格式 : ASF
		Bit-rate (位元傳輸率) : 56Kbps
		Audio Format : 11 KHZ, mono
		檔案格式 : RM
		Bit-rate (位元傳輸率) : 56Kbps
		Audio Format : 11 KHZ, mono
		檔案格式 : mpg4
		壓縮方式 : MPEG-4
		音效解碼為MP3
		Bit-rate (位元傳輸率) : 56Kbps

數位化的檔案格式依目的可從上擇其一至多種格式做數位化，而不需要每種都做。其次資訊科技日新月異，可能多年後此檔案格式已經不適用或無可閱覽的軟體，故其檔案格式要時常注意其實用性，為避免因常更動檔案規格而造成管理與使用上困擾，建議可每隔 5 至 10 年重新審定整個作業流程與標準，而要選定新標準時有幾點注意事項：

- (1) 標準國際化程度
- (2) 閱覽軟體普遍性
- (3) 格式簡單及效率性
- (4) 可適用平臺種類多寡

四、多媒體檔案儲存

儲存管理為多媒體檔案管理的基礎，若儲存管理做不好，一切則會化為虛有。儲存管理最重要的有兩個考慮要素，一為成本，另一為資料安全性，成本的考量應包含一次建置成本和未來逐年擴增成本，而資料的安全性則從資料對整個

系統的重要性及典藏的重要性，依不同使用目的而有不同考量。兩個要素為互補，資料安全性越高所需成本則越大，反之成本越少則資料安全性將越低，故儲存系統的設計是在追求其兩者平衡的最大效用。

資料的儲存依目的與實用性上可區分為三種，一為應用儲存，主要提供應用系統所使用；另一為備份儲存，當應用儲存發生問題時能夠在短時間內回復；最後一種為永久儲存，資料最原始最基本儲存位置，其儲存媒體能保存最久遠。

1. 應用儲存

多媒體檔案的應用儲存主要是將資料存放在適當位置，使應用軟體能快速的存取使用，在規劃應用儲存時除資料安全與經費外，還要考慮下列幾點要項：

- (1) 儲存容量大小：建置一個多媒體應用系統的儲存空間，必須估計系統初次建置時所需儲存容量和未來每年內所增加的資料量，然後依經費和其他要素規劃系統初次建置所需的容量大小。
- (2) 儲存系統架構：在過去儲存系統架構為直接內建在伺服器內部，而隨著儲存技術的進步，目前市面上有多種不同儲存系統架構，在規劃系統時就要依資料特性和使用傳輸方式來找尋最適當的技術架構。
- (3) 應用軟體讀取資料方式：除自己開發的系統可自由選擇讀取方式外，若系統有包含一些套裝軟體，則要考慮到套裝軟體的特性，才不會當系統建立後才發現應用軟體無法讀取資料情形發生。
- (4) 擴充方式：數位典藏系統的儲存設備建置不是一次就可以完成，儲存設備會隨資料成長而一起成長，故儲存設備除系統建置初所需之外，會依每年資料量成長而作設備擴充，故儲存設備的未來擴充性是非常重要的，在系統規劃時就要詳細的考慮。

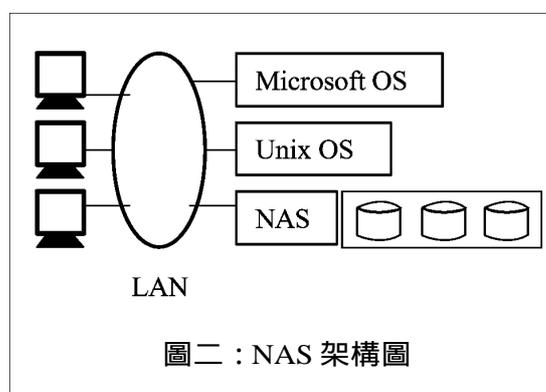
系統在面臨資料儲存量的增加需求時，若只從單純擴充儲存容量去考量，則易造成將來儲存管理整合困難，且其儲存基礎架構建置速度無法符合儲存需求成長等問題。現今儲存架構技術進步迅速，目前最普遍有三種儲存架構，包括直接存取儲存 DAS (Direct Attach Storage)、網路存取儲存 NAS (Network Attach Storage)、儲存區域網路 SAN (Storage Area Network)。

DAS 是直接儲存設備，在伺服器上直接加裝磁碟陣列控制卡 (RAID Control Card)，然後用此界面卡連接多顆硬碟，構成磁碟陣列提供伺服器使用。其架構是由電腦主機硬碟架構演進，安裝使用方式如一般電腦上加裝一顆硬碟一

樣，只是它可外接一個機殼內存放多顆硬碟擴充。故 DAS 就是針對伺服器儲存裝置擴充，使能不受主機內部空間和匯流排的限制，而其他伺服器要存取於 DAS 內的資料，就要透過原伺服主機的網路卡、記憶體、處理器等，因而減緩資料的存取效率。

為了避免 DAS 的缺點，能消除各個伺服器儲存空間的孤島，將原本分散放在各個主機儲存資料釋放出來，集中於一地方管理、資料的安全和分享，使各伺服器能專心於其原本工作，並能快速提供各個伺服器所需的資料，故有 NAS 和 SAN 的架構產生。

NAS 是網路儲存伺服器，其包含主機與硬碟，但與一般電腦不同，其硬體架構及軟體作業系統則是針對網路儲存特別設計，可讓多個伺服器透過網路連結傳輸資料，存取相同檔案並提高其效率。NAS 的使用方式與一般 Microsoft Windows 作業系統的網路磁碟大致相同，其 NAS 架構圖如圖二，從圖中可知 NAS 是透過 LAN 用 TCP/IP 協定與其他伺服器傳送資料，而其他伺服主機的作業平臺可包含 Microsoft 和 Unix 等作業系統，其 NAS 的特性：

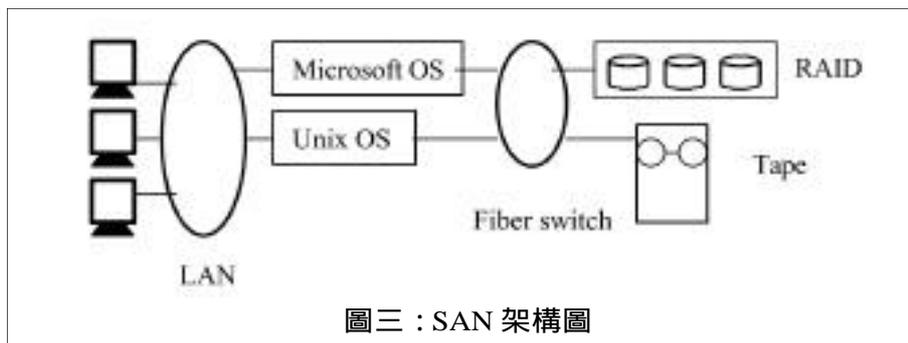


從圖中可知 NAS 是透過 LAN 用 TCP/IP 協定與其他伺服器傳送資料，而其他伺服主機的作業平臺可包含 Microsoft 和 Unix 等作業系統，其 NAS 的特性：

- (1) 安裝簡單
- (2) 與現有網路相容
- (3) Web 的管理介面
- (4) 跨平臺的檔案分享
- (5) 網路安全集中管理
- (6) 與現有備援機制相容

NAS 提供網路上共享儲存資料功能，主要用途為提供以檔案為單元的服務，和一般檔案伺服器最大的不同，在於僅需要簡單的安裝便可以在區域網路環境下使用，並支援 UNIX 作業系統的 NFS 標準、Windows 作業系統的 CIFS 標準、FTP 和 HTTP 等標準，可提供跨平臺的服務，減少對同一平臺建置儲存設備的支出，並可統一集中管理資料，降低系統管理人員的負荷。

SAN 提供一個專屬網路，讓存取儲存單元透過網路與服務伺服器連接傳送資料，而存取儲存單元則可被服務伺服器視為本機儲存設備使用。SAN 的架構圖如圖三，整個 SAN 系統是採光纖通道傳輸，因此伺服主機和儲存設備都需要相關的連接設備才能互連，系統建置時伺服主機內須安裝一片光纖網路卡，另外需要光纖通道的交換器和傳輸線路，最後其儲存單元也必須有光纖存取介面。



圖三：SAN 架構圖

SAN 是一個高效率的資料專屬通道，提供伺服主機快速存取資料，其特性：

- (1) 儲存設備集中管理
- (2) 高效能的備份及還原
- (3) 大量資料傳送不會影響到主網路
- (4) 建置技術較困難
- (5) 不同作業平臺的檔案不易分享

SAN 可大致分為專屬型與開放型兩大類，目前整個市場趨向開放型的 SAN，SAN 因有專屬的光纖網路，提供較高可靠性的資料傳輸，適合應用在資料庫及需要速率保證系統單元，例如資料倉儲、顧客關係管理等應用。

DAS、NAS 和 SAN 各有各的特點與優劣，沒有任何一種儲存架構可完全替代其他架構，因此在一個儲存系統中可能同時存在這三種架構，所以統合管理這三種架構就非常重要，現在市面上就有多家廠商提出整合方案。另外除這三種架構外還有其他架構在發展，如 iSCSI、IPSAN 等新架構，故在設計儲存架構時必需多參考市場現有技術。

2. 備份儲存

所有應用儲存設備都不可能保證永不故障，因此如何制定一個完善的資料備份與回存管理制度非常重要。除要安排備份排程外，還須制定備份設備及儲存媒

體規格，使當應用儲存發生問題時能夠在短時間內回復。

備援的目的^[2]是為了災害回復，依回復時間和回復策略的不同，其備援的架構就不同，備援方式及途徑大致可分為四種：

(1) 磁碟陣列容錯機制

磁碟陣列提供高效率高安全的容錯機制，採取額外硬碟空間來互相備援，依技術種類可分為：Level 0、1、3、5、0+1、10、30、50 等不同的容錯計算方式，最常用的是 Level 5 容錯方式，而在磁碟陣列還提供可熱抽換及熱備援的磁碟，使系統可以線上即時更換磁碟備援。

(2) 以磁帶備份

磁帶或光碟備份所需的復原時間較長，磁帶規格如 DAT、DLT、SuperDLT、LTO 等，磁帶的單位成本比其他儲存媒體較低，但在資料的讀取速度則有待解決。

(3) 電腦叢集備援

由兩套相同的電腦系統組成，當一電腦設備發生問題時，另一組的相同電腦將會自動接手，使整個系統不因電腦發生問題而停頓。系統在平常時有兩種運作方式，一種方式為一套正常工作，另一套閒置等待，另一種方式為兩套系統都正常工作，工作量互為負載平衡，當另一系統發生問題時則將工作量移轉過來。

(4) 異地備援

另一機房架設一相同的系統設備，在正常作業時則透過網路以同步或非同步達成兩套系統資料同步，同時可防止因機房或建築物遭遇不可抗拒因素時所造成的作業停止。

在規劃備份方式和策略^[3]時，有許多事項必須先行考量：

- (1) 依資料重要性來決定多久備份一次。
- (2) 依資料類型不同來決定使用何種備援方式。
- (3) 依備份速度及回存時間決定備份類型及頻率。
- (4) 依資料異動性決定備份資料保存期限。
- (5) 依上述備份方式及週期決定所需備份空間和設備。

若要想有效管理和保存資料，須明訂整個備份策略及工作規範，除指定負責

[2] Jan 2002 網路通訊—如何做好資料備份管理26。

[3] Jan 2002 網路通訊—如何做好資料備份管理 28。

執行人員外，還包含備份頻率、備份類型、資料保存期間、定期回存資料演練等，確保資料備份機制有效及可靠性，一個系統其備份資料會隨著時間增加而改變，故應定期審核其備份策略，考慮市面上的資訊產品、系統所產生的資料及其重要性，使其符合系統未來的需求。

3. 永久儲存

數位化資料做為永久典藏用途時，要考慮如何將數位化資料能保持永永久久，為能將資料保持長久就須要考慮儲存媒體的使用年限，而目前使用永久儲存的媒體大部分為光碟片，光碟片的好處是除可保存年限長久外，若碟面上有污點尚可用水洗，但要預防刮傷和電磁波破壞。

光碟片目前可分為CD光碟和DVD光碟兩種。兩種光碟的考量，DVD光碟可儲存較大空間，但成本較高且使用機器設備也較貴，目前DVD光碟讀取設備有DVD-ROM、DVD-R、DVD-RW、DVD+RW、DVD-RAM等設備。而CD光碟讀取設備目前為一般電腦的基本配備，且單片價格便宜為DVD光碟的二十分之一價格，但考慮未來資料增加及館藏空間有限條件之下，以DVD光碟為永久儲存媒體較合適，挑選DVD光碟片種類時要注意自己DVD讀取設備是否能讀取，因一般DVD讀取設備較易挑片。

資料存放在光碟片中，還要幫光碟片找一儲存的位置，如圖四為市面上有銷售控管溼度的光碟片保存箱，國家圖書館和故宮博物院等單位正在使用。

永久保存除考慮資料儲存外還應包括資料格式的讀取，若多年之後缺乏可讀取之應用軟體，則資料雖然存在仍不能應用，像現在有的膠質唱片保持很好，但可讀取的機器設備卻很難尋找到，所以手上有資料在卻無法使用，建議在考慮永久典藏策略時有幾點要注意：

- (1) 選擇合適儲存媒體，選擇重點為保存時間、成本、儲存容量、媒體大小。
- (2) 安置合宜的儲存位置，選擇重點能避免災害、溫度、溼度。
- (3) 儲存檔案格式要能普遍、簡單、國際化及閱讀軟體要廣泛。
- (4) 定期審查永久典藏策略，資料要為永久典藏故時間不宜太短，但為確保資料能永久使用又不宜太長，建議約五至十年審查儲存媒體和檔案格



圖四：光碟保存箱

式，並可將過去資料重新檢驗是否有損壞。

五、多媒體檔案命名^[4]

對使用者而言，將多媒體檔案安全完整的保留下來，其更重要是如何去得到這些多媒體檔案，要能讓使用者找到所想要的檔案，除透過詮釋資料 (metadata) 輔助外，在系統內部最重要的是多媒體檔案的位址。

要如何才能找到多媒體檔案？目前電腦檔案大多數採一致性資源定位 (Uniform Resource Locator, 簡稱 URL) 方式，其包括：電腦名稱 (host name)、儲存路徑 (path) 和檔案名稱 (file name)，以及取用這個檔案的通訊協定 (protocol)，此三者組成了 URL，它最大的問題是無法做為多媒體檔案的永久辨識碼，因為它是採用直接定址方式，一旦數位資源由原來儲存的電腦移到另一臺電腦，或改變儲存路徑和檔案名稱，原來的 URL 就不正確。

要找到多媒體檔案的位址，除主機的位址外還包括路徑及檔名，而主機位址目前已使用 DNS 方式來做間接定址，但路徑和檔名仍還是採直接定址方式。電腦界目前正在推廣全部間接定址方式，目前較流行的為一致性資源命名 (Uniform Resource Names, 簡稱 URN) 和數位物件識別碼 (Digital Object Identifier, 簡稱 DOI)，為使我們多媒體檔案名稱未來能與國際數位化檔案命名方式接軌，有幾點要注意：

1. 資料提供者若為分散式，須依檔名辨識此資料是由那一個單位所提供。
2. 此命名方式可支援同一物件的多種格式檔案及其使用目的。
3. 依命名方式在整個系統中，每一數位資源皆有唯一的檔名。
4. 命名的字集符合下列規則：
 - (1) 不使用中文字。
 - (2) 檔案名稱英文字大小寫不作區分。
 - (3) 儘量不使用 %、/、?、#、* 字元。

數位化檔案命名是一項複雜的議題，目前國際上各種命名方式歸納起來，可以「命名方式 + 註冊機關代號 + 註冊資源代號」來呈現：

1. 命名方式如以 URN 方式則為 urn，DOI 則為 doi。

[4] 國家圖書館館訊 90 年第 3 期—數位化檔案命名原則 1-5。

2. 註冊機關代號如以 URN informal 方式則由申請機關向註冊中心 (IANA) 申請分發為 urn-d (d 為數字), DOI 方式向註冊中心 (FDI 或 CrossRef) 申請分發一代碼。
3. 註冊資源代號則由註冊單位內部自編, 無一定格式但要內部為唯一代號。如 URN 則需要提出內部編碼方式給 IANA 協會審查, 而 DOI 只要資源識別碼註冊時不與現有重複即可。
4. + 為區分碼, 如 URN 為 ":", DOI 為 "/" 等。

由上分析, 不管加入那一個網路資源組織, 其註冊資源代號都是要由註冊機關來自訂, 所以我們要設計多媒體檔案命名方式也能提供未來註冊資源代號使用, 故在未來不管那一種網路資源組織盛行, 都能快速簡單的轉換成該組織命名方式, 使其符合系統擴充性及未來性。

六、多媒體檔案連結

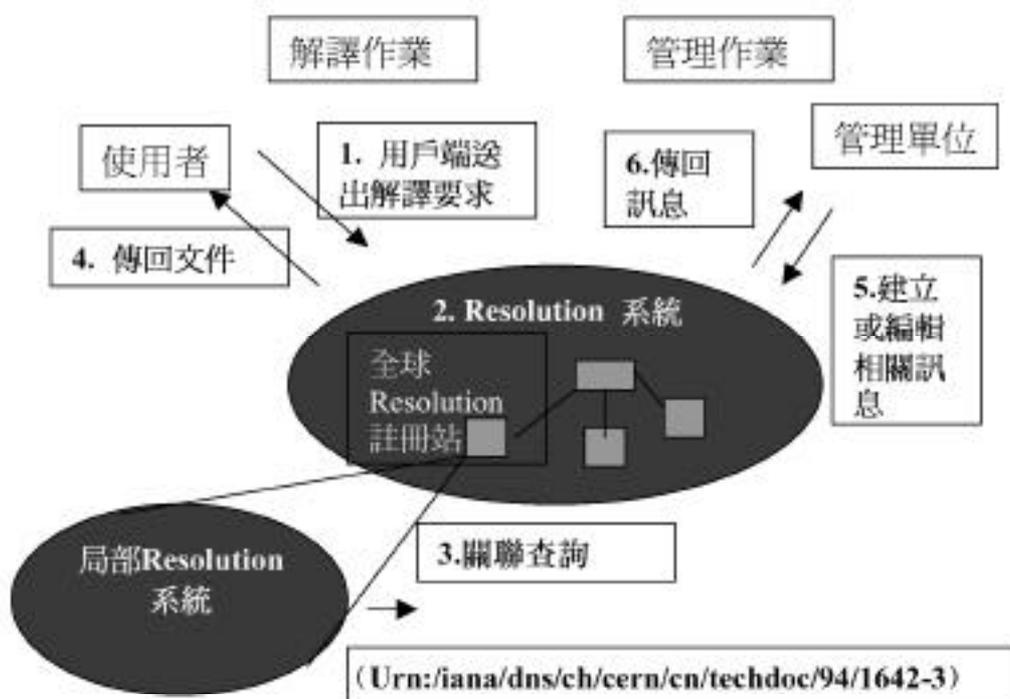
多媒體檔案的命名規則擬定完成, 再來就是規劃設計詮釋資料 metadata 與多媒體檔案做連結, 目前一般有兩種做法, 第一種做法是在建立詮釋資料時將多媒體檔案的 URL 位址建入, 第二種做法則在建立詮釋資料時可先保留欄位內容, 等詮釋資料和多媒體檔案都進入系統時, 再由畫面挑選其相對應之多媒體檔案。兩種做法的差別是第一種要預先規劃多媒體檔案位置、檔名命名, 規劃擬定後就不能輕易更動; 第二種做法為建立詮釋資料時, 不需要知道多媒體檔案位置和檔名, 等到資料都進入系統時再建立其資料連結, 第一種做法只要計畫做得好, 在實際執行時就非常方便, 但若有意外或架構改變時就要花很大成本去更動, 第二種做法就很有彈性可適應各種不同環境, 但缺點需要在系統上選擇對應的多媒體檔名, 需要花很多時間與人力。

另外還有一種變形方式就是將部分規則在系統軟體建立, 在詮釋資料只存放所必須欄位, 如此架構需更動時, 只要修改系統的資料存取架構即可, 而不需要更改全部的詮釋資料, 好處是除了有上述第一種建檔方便外, 更可使架構性變動的修改簡單, 但其缺點仍是移植性困難, 不能將多媒體檔案給別系統使用, 即不能將資料提供其他系統使用。

目前數位典藏技術發展一日千里, 建置單位已漸由單一單位朝向數個單位聯合發展, 使數位化資料的儲存分散性與可移植性越來越重要, 而國際上為解決多

媒體檔案的存取問題，使數位資源不因架構、技術及位址的改變能使需求者都能尋找到，採用 URN、DOI 等命名方式並配合解譯系統（如 handle system）來做多媒體檔案的連結，此做法使得多媒體檔案有一全球唯一識別碼，不會因其他因素改變而使多媒體檔連結失敗，只要詮釋資料存放其識別碼都可找到其資料，能解決資料之分散性與可移植性問題。

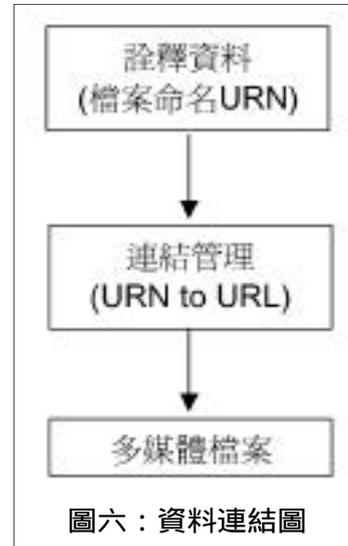
以國際 URN 作法為例，首先有一個命名規範如 RFC 2141 RUN Syntax，來統一全球各多媒體檔案的命名方式，其次設立一個數位資源註冊機關如 IANA (Internet Assigned Numbers Authority)，提供命名方式的管理及相關服務，再來訂定使用者、解譯機關及提供多媒體檔案機關的溝通方式規範，如 RFC 2169 A Trivial Convention for using HTTP in URN Resolution、RFC 2276 Architectural Principles of Uniform Resource Name Resolution、RFC 2483 URI Resolution Services Necessary for URN Resolution 等文件，圖五為一個 Resolution 運作的系統圖。



圖五：Resolution 系統圖

依目前國內狀況，使用國際的解譯系統基本上還有幾個困難，一個是將多媒體檔案登入此解譯系統時需要負擔額外費用，如同在網際網路登記網站的 domain name 時需繳交維護費給 domain name 維護組織，二是在使用者閱讀軟體需再額外下載解譯軟體，使其判別其多媒體檔案命名，能尋找到多媒體檔案位址，最後是國內並無單位向國際命名組織申請中華民國的代碼，且無基礎建設以建立局部 Resolution 系統以供使用者解譯用。

為能解決國內數位典藏發展問題並能與國際接軌，筆者提出可在多媒體檔案和詮釋資料 (metadata)



圖六：資料連結圖

中間設立一個連結管理層，如圖六資料連結圖所示一樣，在詮釋資料內建入多媒體檔案的識別碼 (如 URN)，當使用者經由搜尋引擎找到這筆資料的詮釋資料時，系統可經由詮釋資料中的多媒體檔案識別碼去訊問連結管理層相對應 URL (位址)，得到其多媒體檔案的 URL，再依其 URL 讀取其多媒體檔案內容。因 URL 可跨單機、平臺及單位故可達成資料的分散式管理。當資料輸出時，只要輸出詮釋資料和多媒體檔案，不須輸出連結管理資料，資料輸入時再依系統現況決定其主機與目錄位址，動態建立其資料連結管理，解決其多媒體檔案的移植性問題。當系統的多媒體檔案要加入國際數位資源管理組織時，可直接將連結管理層的資料上載至國際管理組織的數位資源登錄系統 (事先需將原資料的識別碼加上國際管理組織給的單位代碼)，便能將其多媒體檔案與國際組織接軌。而此做法的缺點則是應用系統設計較複雜。

連結管理層所需具備的功能為：

1. 多媒體檔案識別碼在系統中為唯一，並可自動經由資料庫的 metadata 中的指定欄位建立。
2. 多媒體檔案實體儲存位址 (URL) 可由多媒體檔案識別碼加上其他固定字元整批自動建立資料。
3. 建立資料庫的 metadata 與多媒體數位檔案連結。
4. 可整批或單筆增、刪、修改多媒體檔案實體儲存位址 (URL) 和多媒體檔案識別碼。

5. 可在輸入多媒體檔案識別碼時設定部份參數同時自動建立多媒體檔案實體儲存位址 (URL), 並可修改預設值。

多媒體檔案連結的作法有很多, 在系統建置時選擇使用那一種作法則有下列要點要考慮:

1. 實體典藏與多媒體檔案關係。
2. 使用者的使用方法及所想從多媒體檔案獲取的資訊。
3. 與其他數位典藏系統的關係。
4. 數位典藏系統的未來規劃。

七、總 結

多媒體技術近年來已如日中天, 數位化資訊的多媒體化時代已經來臨了, 電腦表現從文字進步到圖片、聲音、影片, 而在數位典藏系統中多媒體檔案佔有重大地位, 故我們在做多媒體檔案管理時, 就需要時常注意現代科技的發展, 才不會規劃出一個不合宜的系統, 現今多媒體技術趨勢:

(一) 多媒體檔案種類更多元化

現今多媒體檔案種類包含文字、圖片、聲音、影片等種類, 而立體實物表現和電子書的技術和規格則在不斷的成熟。

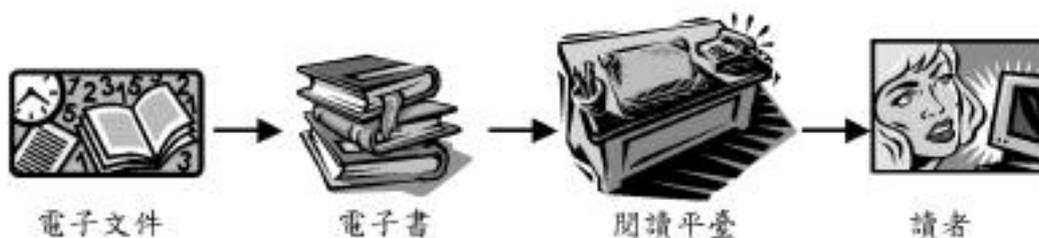
立體實物表現最早是有 7D 理論出現, 提出空間一點只要儲存 7 項資訊就可表現這點的特性, 可以從不同地點不同角度看觀此物件, 但因資料太龐大複雜, 故有 6D、5D 和 4D 理論產生, 每減少一維資料儲存則增加其觀看的限制, 像 4D 就只能產生環場效果, 此做法為 model base。另外有 image base 方式為從各種角度拍攝實體影像, 而我們需要多細密觀看實體就拍攝多少張影像。Model base 和 image base 是可相互支援, 將實體物件數位化時, 其較簡單部分可採 model base 方式, 而複雜部分可使用 image base 方式執行, 如此可降低建立物件的複雜度和所需資料。

電子書規範目前可分兩大類, 一類是將書頁當成圖片處理, 並建立其書的架構將其每頁串聯起來, 使人使用時像書一樣翻頁, 如國家圖書館博士全文影像系統使用即是; 另一類則將其數位資料轉換成全文檢索格式, 目前如 PDF、XML、OEB 等方式, 可提供內文檢索外還可提供非線性相關閱讀, 我們以 OEBPS 的架構來向大家簡介電子書的整個運作環境。

Open eBook Publication Structure，簡稱 OEBPS，是以 XML 為基礎，對電子圖書的內容、架構及表現提供一個規範。其主要目標為：

1. 提供資料內容提供者（例如出版商或其他要將資料展現者）和工具提供者一個共同性使用方法，使各種不同電子書系統平臺間能正確表現電子書。
2. 朝向建立電子書內容格式標準。
3. 讓電子書供應商能依一共通格式，使其使用者能在多種電子書閱讀系統上使用。

OEBPS 架構如圖七，主要可分為三大部分：電子文件（OEB Document）、電子書出版（OEB Publication）、閱讀平臺（Reading System）



圖七：OEBPS 架構

1. 電子文件（OEB Document）：電子書的基本單元，以 XML 語法為基礎並且符合本規範，分為基本型（Basic）和擴充型（Extended）兩種電子文件型態。
2. 電子書出版（OEB Publication）：除包含電子文件集外，並有其他不同型態檔案，包括結構化文字和圖片，提供電子書出版各種資訊。
3. 閱讀平臺（Reading System）：將電子書能直接或間接提供讀者使用的系統，包含其所需要的硬體及軟體。閱讀平臺依其實際使用而有不同的架構，但閱讀平臺一定要能處理電子書，並且具有壓縮、檢索、所有權管理等系統功能。

(二) 多媒體檔案資訊物件化

如何能在浩瀚的多媒體檔案中快速有效的檢索到所需資訊，是多媒體檢索技術發展的目標。目前對多媒體檔案提供檢索標準有 MPEG7，MPEG7 正式名稱 "Multimedia Content Description Interface"，目的提供對多媒體檔案資料內容的描

述標準，使多媒體資訊能如文字檢索一般，可依內容來檢索相關內容。

技術發展除標準建立外，還有兩個重要主流技術在研發，一個是從原視訊檔中如何擷取出有效資訊，建立起多媒體檢索檔案，如視訊攝景分段、畫面人物臉部影像擷取技術等；另一技術為檢索資料比對及方式，如前面視訊檔中之臉部圖像與檢索的臉部圖像比對、如何描述檢索臉部圖像資訊方法等。

(三) 多媒體檔案傳輸更便捷

多媒體檔案之資料量較傳統一般檔案大很多，在網路的傳輸上就顯得反應時間較慢，目前技術發展朝向能即時提供資訊，使用者不需等待長久下載時間，主要有 JPEG2000 和 MPEG4 規格訂定，JPEG2000 主要在制定可整合不同種類、特性的影像壓縮格式，可提供無失真壓縮、解像漸進式傳輸及抗傳輸誤差。而 MPEG4 則為提供低傳輸率的網路視訊使用，主要方式從網路上傳來資料存放在暫存區，等暫存區存放一定資料量後就開始播放，而不需等待整個檔案下載完成，可線上直接播放節省使用者等待時間，但會因網路頻寬負載量太大而導致停格現象產生。

參考書目

1. 圖書館法圖書資訊相關技術規範—資料數位化標準草案。
2. 陳建中、江政欽。 簡介影像視訊技術趨勢與多媒體內容描述介面 MPEG7 。《電腦通訊》，第 86 期。
3. 張家瑜。 儲存架構雙雄NAS&SAN 。《網路通訊》，第 126 期。
4. 賴文正。 如何做好資料備份管理 。《網路通訊》，第 126 期。
5. 黃振麟。 企業DAS 儲存系統建構白皮書 。《資訊與電腦》，第 250 期。
6. OEBPS。 網址：<http://www.openebook.org/oebps/oebps1.0.1/download/>。

Multimedia File Management

Li-Yuan Chen

System administrator, National Central Library

【 Abstract 】

Multi-media techniques are one of the most important aspects of information technology. An actual practice of which is the digital archive system in a library. This paper discusses mainly the management of multi-media files when constructing and operating a digital archive system. The outline contains the producing process, naming mechanism, data storage of the multi-media file, linking of the digital archive system, etc. Finally, this paper will make an introduction of the current development of the multi-media techniques.

Keywords: multi-media files, digital archive system, data storage, naming mechanism, system linking