

能快速處理大量文物影像之數位化儲存備份機制

-以國立故宮博物院為例

劉家銘

國立故宮博物院

資訊中心

886-2-28812021#162

jml@npm.gov.tw

游孝國

國立故宮博物院

資訊中心

886-2-28812021#163

reco@npm.gov.tw

謝俊科

國立故宮博物院

資訊中心

886-2-28812021#150

Kenneth@npm.gov.tw

林國平

國立故宮博物院

資訊中心

886-2-28812021#323

jameslin@npm.gov.tw

摘要

本文主要是從系統整合與實作的角度來探討建立快速的數位化流程後，後端儲存備份機制所面臨的問題與瓶頸，以及如何改善，並將前端文物影像數位化流程與後端儲存備份機制有效的整合與串連，建置一能快速處理大量文物影像之數位化儲存架構，以適合快速且大量文物數位化需求之環境。

關鍵字

文物影像數位化流程、儲存備份機制

1. 前言

國立故宮博物院(以下簡稱故宮)進行文物數位化工作已超過四年，主要工作內容之一即是將故宮收藏的豐富文物拍照並數位化，作為永久典藏之用。且數位化後之影像檔會分別存放於硬碟、光碟片與磁帶上，以分散風險，保障辛苦數位化之成果。

初期故宮數位化之目標僅針對國寶與重要古物等級以上之文物，數量不多，故可以採用傳統拍照，再利用正片或負片掃描之方式進行數位化動作。然而，故宮所收藏之六十多萬件文物皆有其珍貴價值，因此於 93 年度調整數位化目標，希望可將所有藏品全面數位化。但是，要將故宮所收藏的六十多萬件文物全部數位化並保存下來，是非常困難的。

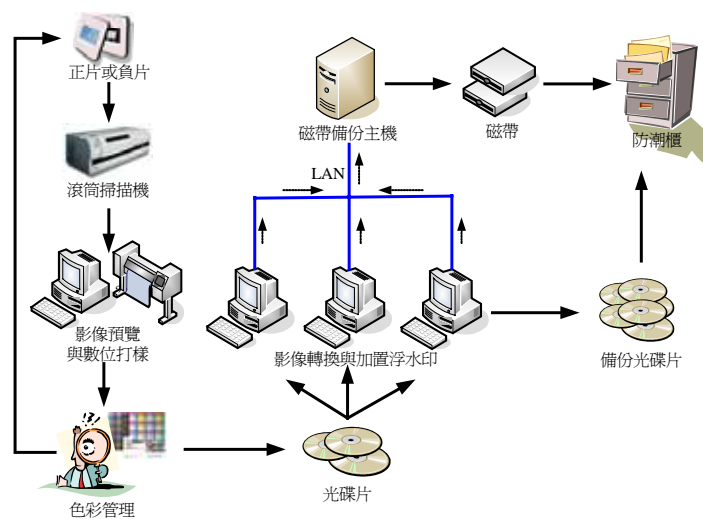
為達到全面數位化目標，故宮開始加速進行文物數位化工作，除了將原本每年掃描約 2000 張正負片提高到每年掃描約 8000 張左右，也陸續建置了三套數位攝影系統，因此同時有四套數位化流程在運作，以如此方式加速且加量的進行文物數位化工作。

因為同時有四套數位化流程在運作，所以產生的數位影像檔不但多，且檔案容量也非常大，而後端的儲存與備份機制就顯得相當重要，如果後端的儲存備份機制效率不彰或無法快速處理又多又大之數位檔

案，將造成加速文物數位化的瓶頸。故如何建置一高效率的儲存備份機制，以銜接前端快速且大量的數位化流程，為一值得探討的議題。

2. 故宮初期之文物影像數位化儲存備份機制

以下針對故宮初期之文物影像數位化流程與儲存備份機制作一簡介，其流程架構圖如圖一所示。



圖一 故宮初期之文物影像數位化儲存備份機制

故宮數位典藏計畫初期目標為院藏國寶或重要古物等級以上之文物，故預期數位化的量不多，且大多國寶或重要古物等級以上之文物已拍攝過照片，留下不少正片與負片，所以初期影像數位化採掃描之方式，如有未拍攝或正片、負片品質不佳之文物，才提供補拍。

故宮正片與負片掃描採外包方式，將品質良好之正片與負片交予廠商，利用高階滾筒掃描器掃描，取得高解析度之數位影像檔，並利用此數位影像檔作數位打樣，之後將數位影像檔燒錄成光碟與數位打樣稿

一起送回故宮，由故宮檢查影像之色彩與文物實際顏色是否符合，如果不符合，則退回廠商重新掃描；如果符合，則進行後續處理與儲存備份動作。

取得色彩正確之數位影像光碟片後，後端儲存備份機制利用多台電腦同時讀取光碟片之影像檔，進行確認與檢查；接下來將影像檔轉成各種應用所需之格式，或加置數位浮水印等後處理事宜，並加以分類，再燒錄成光碟片，存放在防潮櫃內作為長期保存。

影像後處理與光碟備份完成後，透過內部區域網路將檔案傳送至磁帶備份主機，進行磁帶備份，並將備份好之磁帶存放於防潮櫃保存，最後將最高階典藏級影像檔存放於主機硬碟。

故宮初期數位化的進度是每年 2000 張正負片左右，每張正片或負片會產生約 90MB 左右之影像檔，因此初期的儲存備份機制約每星期可以處理的檔案大小為 3.75GB。

3. 大量文物數位化之瓶頸

故宮以上述方式進行文物數位化至 93 年後，調整文物數位化目標，期望儘快將所有收藏之文物全面數位化。而面對多達六十多萬件之藏品，為加快數位化之速度，除了增加正負片掃描的數量外，故宮也陸續建置了三套數位攝影系統於文物庫房內，直接對文物進行數位攝影，取得數位影像檔，這樣的流程，避免了傳統攝影需要提件拍攝、正片沖洗以及掃描等繁雜流程。且因為這三套數位攝影系統加上原本之正片掃描數位化流程，大大提昇了前端文物數位化流程之速度。

而這三套數位攝影系統每星期約有六個工作天，共可拍攝 340 張影像檔，每張影像檔又需轉成各種所需之格式，因此一張影像檔約會產生 260MB 之檔案大小，因此每星期將產生 88.4GB 之檔案，再加上正負片掃描數位化進度提升至每年約 8000 張，每星期也約有 15G 之數位化檔案產出，因此每星期共約有 103.4G，數位化檔案量是傳統掃描流程的 27 倍之多。

但是前端文物影像數位化流程雖然加快了 27 倍之多，但後端的儲存備份機制卻仍然只能應付每星期 3.75GB 的檔案量，因此面臨了非常大的瓶頸，造成瓶頸的原因有以下幾項：

1. 儲存空間不足：因數位化速度加快，且愈來愈多，主機內建之儲存空間無法有效擴充，大量數位化之資料面臨無空間儲存之問題。
2. 網路速度慢且佔用內部網路頻寬：因為需要將後處理完成之影像檔透過區域網路傳送到其它伺服器或磁帶備份主機作磁帶備份，而故宮內部網路為

100Mbps 之乙太網路，面對每天產生的大量數位影像檔，除了網路傳輸速度非常慢以外，在傳送影像檔的同時，網路頻寬資源也幾乎被傳輸影像檔所佔用，造成故宮內部網路服務的癱瘓。

3. 安全疑慮：因儲存備份機制之系統與內部網路連接，為半開放式之環境，有遭受駭客或病毒入侵之可能，導致珍貴影像檔外流或損失之風險。
4. 無法自動備份光碟：將影像檔燒錄成光碟是故宮主要的備份方法之一，且基於分散風險的原因，會備份兩片光碟以上，並存放於不同地點做異地備份。但由於數位化流程加速的關係，所需備份的光碟數量急遽增加，再加上燒錄光碟片必須隨時有工作人員在旁，不但非常浪費人力，也無法利用晚上空檔時段進行光碟燒錄，也浪費了時間與資源。

為了解決上述之問題，故宮利用以下三項改善方式，將原有儲存備份機制作了小部份的調整，成為故宮目前現行之儲存備份機制，如圖二所示。

1. 建置高速虛擬區域網路(VLAN)：為讓前端數位攝影流程與傳統數位化流程之影像檔與能快速傳送至後端儲存主機，故宮建置一 1Gbps 之高速虛擬區域網路(VLAN)，獨立於內部區域網路外，不但可達到快速傳輸大量影像檔之目的，又可避免佔用內部網路頻寬。此外，因為獨立於內部區域網路之外，避免了直接與區域網路連線，降低了遭受駭客與病毒之風險，進而提高安全性。
2. 建置儲存區域網路(SAN)：為能存放大量的影像檔，成為最終備份區，加上同時會有四套數位化流程之影像檔會透過網路傳送進來，因此需要高速的儲存空間，以應付大量的資料同時儲存的問題，所以建置了具有高擴充性與高速特性的儲存區域網路，以存放不斷產生的大量影像檔。
3. 採用光碟櫃備份光碟：利用光碟櫃與其附加軟體可以設定排程與可同時置放數百片以上光碟片之方便性，避免燒錄光碟時需隨時有人員在旁，且可利用下班時段進行備份工作，以節省人力並充分利用晚上資源，提升光碟備份效率。

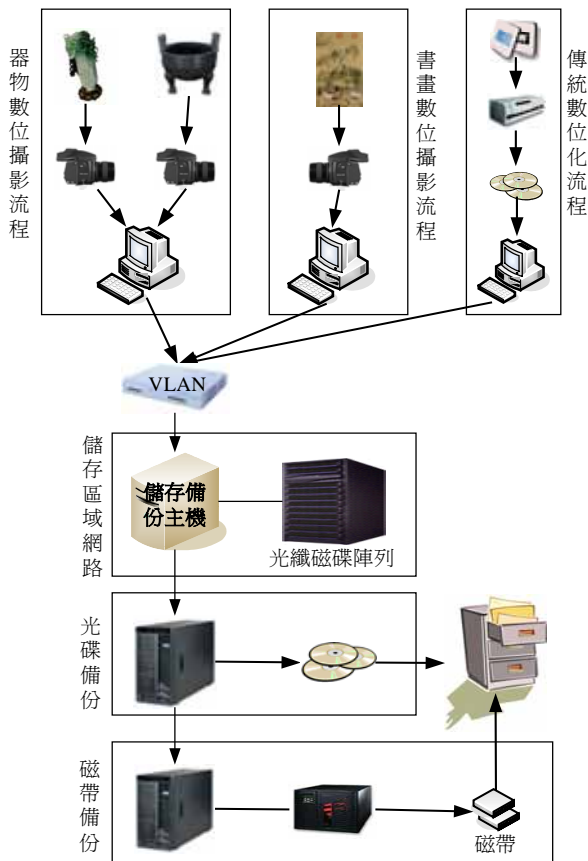
在現行的儲存備份機制中，前端數位化流程所產生之數位影像檔，透過高速虛擬區域網路傳送到儲存備份主機上，將最高階之影像檔儲存於磁碟陣列中，再將這些大量影像檔全部傳送到光碟備份主機，利用光碟櫃自動備份光碟，節省大量的人力，然後再全部傳送到磁帶備份主機做磁帶備份。

然而，這機制只能勉強應付三套數位化流程(器物數位攝影兩套與傳統數位化流程)每星期所產生之大量

數位檔案，如果要應付四套數位化流程同時上線，則無法負荷，數位化進度就會馬上受到影響。分析了其原因，故宮現行儲存備份機制仍有以下幾項問題：

1. 網路頻寬不足與易產生碰撞：因為前端四套數位化流程所產生之大量影像檔皆傳送到儲存備份主機，如果同時傳送的話，會因為頻寬不足或易產生網路封包之碰撞等因素，導致傳送時間過久，並且造成主機的負載過重。
2. 移動大量數位影像檔：為了節省人力，精簡備份步驟，故約每星期批次備份一次數位化檔案，每次處理的檔案量約 100GB，而為了作磁帶與光碟備份，需要移動 100GB 的檔案兩次，每次皆需耗時數小時甚至十多小時以上，相當的浪費時間。
3. 儲存空間不敷使用：雖建置了儲存區域網路，但因數位影像檔成長速度太快，且儲存空間成本高昂，因此無法不斷擴充儲存空間，導致儲存空間不敷使用。

因為上述三項問題，導致整體的數位化速度仍無法完全提升，因此改善現有的儲存備份機制，使得可以快速且大量的處理影像檔是必須且刻不容緩的。

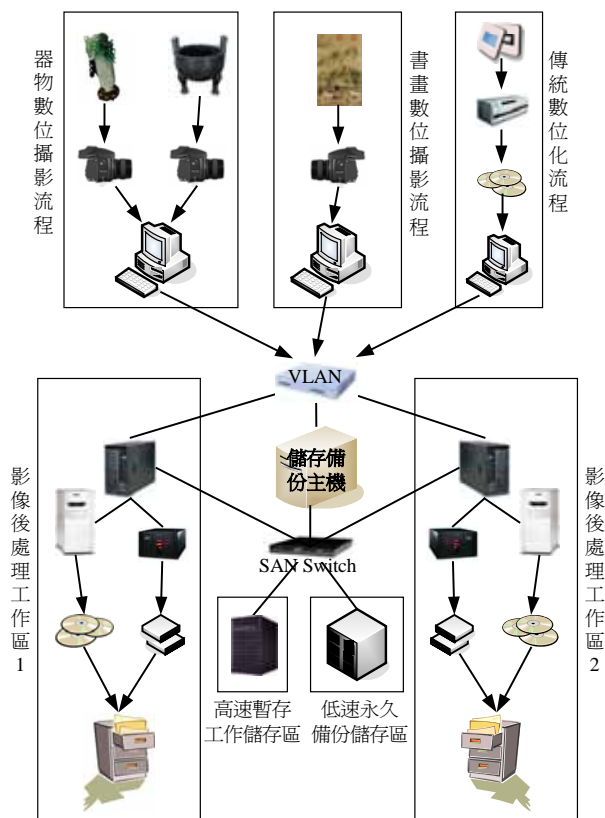


圖二 故宮現行之儲存備份機制

4. 可快速且大量處理之文物影像數位化儲存備份機制

為改善後端儲存備份機制，使能夠快速應付前端數位化流程產生之大量資料，改善後之儲存備份機制如圖三，改善方法如下：

1. 建置暫存工作儲存區：暫存工作儲存區主要目的是提供前端數位化流程產生之大量影像檔暫時存放，以加置浮水印、轉換影像格式等後處理工作，以及可同時燒錄光碟與備份磁帶等多項備份工作。為能同時符合這些工作之需求，將以光纖介面、具有高速傳輸特性之儲存區域網路作為暫存工作儲存區，解決工作空間不足、快速傳輸與處理大量影像檔等之問題。
2. 建置永久備份儲存區：永久備份儲存區是當影像檔後處理與備份完畢後，用來存放最高階典藏等級影像檔之用。因其主要目的為備份儲存，無即時作業之需求，故採用較低速但高容量之硬碟，如 SATA 規格之硬碟。但備份好的大量影像檔要傳送到此區域儲存，仍需要快速的儲存網路，故永久備份儲存區同樣採用光纖網路為基礎之儲存區域網路，且具備高擴充性與高容量等之特性，可方便未來繼續擴充使用。
3. 結合磁帶備份與光碟備份：為降低移動大量影像檔所浪費之時間，將磁帶備份與光碟備份結合於同一台機器上，減少移動大量影像檔之次數。而為了能同時備份光碟與磁帶，需搭配上上述提及之高速工作儲存空間，以防止備份失敗。
4. 建置兩個影像後處理工作區：影像後處理工作區之主要目的為轉換各種影像檔格式、加置浮水印及備份光碟與磁帶等，因光碟與磁帶之備份速度相對較慢，即使前端數位化後之檔案可以很快的傳送到後端，但備份光碟與磁帶時，仍成為整體備份效率的瓶頸，因此為解決此瓶頸，建置了兩套備份主機、光碟櫃與磁帶機等，形成兩個影像後處理工作區，以加快光碟與磁帶之備份速度。且因多了一套光碟與磁帶備份設備，分散了大量影像檔資料流，不但達到負載平衡(load balance)之效用，對整體儲存備份機制而言，更兼具容錯(failover)之優點，提高了整體系統的可用性與信賴度。



圖三 能快速處理大量文物影像之數位化儲存備份機制

圖三中，前端的數位化流程，包括數位攝影、傳統數位化流程等，將數位化後的檔案儲存於該工作區域的工作站上，進行初步的影像轉檔等動作，接著再將所有影像檔透過 VLAN 傳送到高速暫存工作儲存區內。

傳送到高速的暫存工作儲存區後，可由影像後處理工作區中的工作站針對這些大量的影像同時做轉檔、附加浮水印與備份至光碟或磁帶中，因此所需的儲存空間須為高速的儲存媒體，以即時傳輸這些龐大的資料量，所以我們採用光纖網路為基礎搭配光纖磁碟陣列的儲存區域網路。

利用高速的暫存工作儲存區做完影像後處理與備份工作後，再將最高階的影像檔儲存到最終的儲存區備份，因為只需備份最高階的影像檔，因此不需要移動所有的檔案至永久備份儲存區，減少檔案移動所需時間，且最高階的影像檔沒有線上即時使用的需求，故永久備份儲存區不需要高速的儲存媒體，只需要儲存在一般較低速的磁碟陣列裡，以節省經費。

而此備份機制即能處理每星期約 100G 的數位檔案，處理的量是故宮最初儲存備份機制的 27 倍之多，因此改善後的文物影像數位化儲存備份機制，搭配前

端快速的數位化流程，可使整體文物影像數位化的效率大幅提升。

5. 結論及未來發展

此文物影像數位化儲存備份機制，主要是利用高速且獨立之網路環境，讓大量的文物影像可以快速的傳送到後端儲存設備中，藉由高速的光纖磁碟陣列與兩組影像後處理工作區，來處理影像轉換、加置浮水印與備份等工作。並利用光碟機與磁帶機同時備份成光碟與磁帶，避免多次的移動大量影像檔，以大幅縮短資料傳輸、處理與備份的時間，實現可快速處理大量的文物數位化影像，大幅的加快整體數位化效率。

此文物影像數位化儲存備份機制雖大大的加快了儲存備份的速度，但仍有一瓶頸無法突破，即是影像檔的檢查工作。

因為不論是數位攝影或傳統正片掃描的方式所取得的數位影像檔，當傳送到後端儲存設備後，皆須以人工的方式檢視所有影像檔，以確認所有影像檔皆正常，才可進一步做後處理與備份的動作。且當影像檔都備份成光碟片後，仍需要再一次的人工檢視，檢查所燒錄的光碟片內容都正常，才能確保備份的影像檔無誤。

以上這些動作皆需要耗費相當多的人力與時間，這不但是原本數位化流程的瓶頸之一，也是本儲存備份機制的瓶頸，因此未來將繼續尋求此問題的解決之道，避免太多的人力負擔，使得整體數位化流程能更為快速、有效率。