

# 數位資訊資源長久保存之探討

陳雪華、洪維屏

## 摘要

資訊科技帶來無比的便利，不但體積小、儲存量大，同時易於複製、傳播和再利用，越來越多資訊、甚至是知識、文化，原生且單純只以數位形式存在。然而資訊保存問題並未因資訊資源數位化得以完全解決，反而可能因而永久遺失，成為「資訊黑暗時代」。本文探討數位資訊資源長久保存議題，從長久保存的重要性談起，說明保存數位資訊資源的新挑戰，與學者專家們所提出的保存方法並重新思考數位化的意義和價值。

## 壹、數位資訊資源長久保存的重要性

隨著科技日新月異的進步，與網路技術的發展，人類歷史的發展來到了資訊時代。由於資訊科技帶來無比的便利，不但體積小、儲存量大，同時易於複製、傳播和再利用，越來越多資訊、甚至是知識、文化，原生且單純只以數位形式存在。據加州柏克萊大學於 1990 年代晚期所做的一項預測顯示：包括印刷、電影捲片、磁性或光學等形式的產品，全世界每年的總儲存量約有 15 億 GB，相當於每人（不論男女老少）的儲存量約為 250MB。（註 1）超過 93% 的新資訊是直接以數位形式創造，或可以說是「數位原生」。（註 2）在發現資訊科技所帶來的益處之後，人們越加依賴，並誤以為資訊只要以數位形式儲存，便能一勞永逸，徹底解決資訊保存問題。然而，科技發展的速度越來越快，永遠都有新的格式、新的技術、新的軟硬體產生，而逐漸取代舊型的地位；因此，以舊的格式儲存，或是存在於舊式媒體上的資訊，極有可能再被讀取而使用。外國學者稱之為「資訊黑暗時代（information dark age）」，意指由於數位媒體的易碎、易變性，以及技術的老舊過時問題，導致儲存的資訊即使在毫髮無傷的情況下，仍無法加以讀取與解釋其中內容，提供資訊需求者取得和使用。

這對於肩負保存文化與人類知識、智慧的圖書館、博物館、以及檔案館工作人員而言，無疑是一項嚴重的挑戰。近年來，資訊資源的數位化在圖書館、博物館、以及檔案館等領域中，成為重要的發展趨勢之一，其主要的動機有二：一為保存，二為增加資訊資源的可得性。數位化保存是一項基礎工程，同時從長遠的

角度思考，應將數位化保存提升至數位資訊資源長久保存的層次，肯定其重要性，並給予相當程度的重視和嚴謹的研究，以避免類似資訊黑暗時代的情況發生，而造成永久的遺憾。

## 貳、數位資訊資源長久保存所面臨的挑戰

數位資訊資源長久保存所面臨的問題與挑戰，與傳統資源格式保存的不同。為了達到長久儲存、保存及取用的目標，傳統格式的保存、典藏所需要注意的，可能主要在於典藏環境的設定、資源的使用與管理等方面；然而，數位資訊資源的長久保存則還需要注意到關於媒體衰退、技術老舊過時、以及數位資訊資源過於依賴軟體等問題。以下就各項問題加以說明。

### 一、數位媒體易於損壞變質的特性

紙本資源已被證實的確可以保存很久，以手稿為例，在特定的條件與環境之下，手稿可以持續保存幾百年。（註 3）相對的，數位資源的載體是較易衰退、損壞或遺失的，即使在理想的環境條件下，它們的壽命還是較傳統形式的資源壽命為短。依推測，磁片的壽命約為 5 至 10 年，光碟的壽命則可能為 10 到 100 年。（註 4）由於這些數位媒體會快速衰退，與具有易損壞變質的特性，為避免資料內容遺失，需要做重新保存的決定和行動的時間單位為 1 年，而非 10 年。（註 5）

### 二、技術的老舊過時

比媒體衰退更為潛藏和更具挑戰性的問題為技術的老舊過時。在電腦世界裡，硬體、軟體、和儲存技術等的創新，持續以快節奏的步調進行。記錄和儲存的裝置、處理方法、和軟體等定期以約 3 到 5 年為週期，被新的產品和方法所取代。（註 6）在新技術不斷發展的同時，另一個值得關心的問題是相容性問題。依據不同的需求和資源類型，產生多種不同的資料格式，然而，新版本格式的出現，是否能與舊的格式相容，這點卻無法完全保證。（註 7）

另外，從技術的汰舊換新速度問題延伸探討，可以說是資訊科技典範轉移問題，此乃數位資訊資源不易長久保存的主因。所謂典範轉移是指某學科理論及作

法產生結構性之重大改變，例如資料庫的設計從網路式資料庫、階層式資料庫轉移到關聯式資料庫或物件導向資料庫，這種轉移往往需要重新設計整個資料庫，以便轉移到新的典範。在新觀念、新技術一點一滴的突破下，迫使典藏單位或一般使用者必須經常進行更新媒體、轉置系統等動作，而耗費大量人力、時間、與經費。（註 8）由此可見，技術不斷推陳出新與資訊科學的典範轉移，所帶來的影響更為廣泛而深遠。

### 三、數位資訊資源無法獨立存在，必需依賴軟體

數位資訊資源藉由軟體的存取、顯示，因而得以確定其「存在」。而最能讀取、解釋及顯示數位資訊資源的，自然是原始程式。然而這些程式的可用性又和儲存媒體及媒體讀取程式有關。所以，從典藏的角度來看，無法得知該如何用過時的原始程式，來讀取未來的媒體，或用未來的軟體讀取過時的媒體。（註 9）

### 參、數位資訊資源的存方法

數位資訊資源長久保存問題，在國外已受到普遍的關心，亦提出許多可能的解決辦法，卻無一可一勞永逸。在面對數位媒體壽命短，而技術更替速度快的情況，也許最直覺的解決辦法會是將這些數位檔案再以紙本形式印製，或以其他傳統媒體儲存，但是不能否認的是，這種方法失去了數位典藏的意義，無法應用數位化的特性和功能。

此外，儲存多個複本也不失為一種解決方法。由 Sun 實驗室和 Stanford University 合作進行的 Lots of Copies Keeps Stuff Safe ( LOCKSS ) 計畫，便是使圖書館能夠保存管理其訂購的原件，並且提供足夠的複本，讓全世界任何經過授權的使用者，都能有使用權使用其所需的文件。除此之外，當線上的複本被移除或有任何損壞時，LOCKSS 系統會隨時注意並更新取代之。（註 10）

由文獻探討可知，一般常被使用或討論的保存方法，包括更新、轉置、模擬、使用標準、以及技術保存。以下將就各種保存方法加以說明。

#### 一、更新 ( refreshing )

更新是由複製(copying)的想法而來。誠如前述，目前尚未有永恆和持久耐用的媒體來儲存數位資訊資源，導致必須在適當的時間，進行資訊的測試與更新。

更新資料到新的媒體上，便能重新開始資料的生命週期。以 CD 光碟為例，假設其保存期限為 20 年，則可以每 10 年便更新一次以保存資料。

新媒體的選擇不斷出現，而每一代新的版本出現時，會有更多的儲存量和更少的花費。在此趨勢之下，很可能資料都不再更新於同一種媒體之上，而改以另一種新的媒體儲存。實際的做法是複製到當下已普遍被接受的媒體格式上。（註 11）以嚴格的角度來看，將資料更新到新的媒體或版本上的做法，可以說是「轉置」，而非單純的更新。

更新的觀念簡單易懂，在執行上也較其他方法單純，但此方法並不能確保資訊在未來能夠被檢索與處理。

## 二、轉置（migration）

轉置是比更新的內涵為寬廣而複雜的概念。依據 RLG/CAP 數位典藏報告書對於轉置的定義，轉置是指一套有組織的而被設計為達成數位資訊資源從一套軟硬體轉移至另一套，或是將其轉移到新一代電腦技術的方法。其目的在於：1. 保存數位物件的完整性；2. 維持讓使用者可以在技術不斷轉變之下，仍可以持續進行數位資訊資源的檢索、展示、與使用。（註 12）

由於新技術的發展和採用，以及媒體衰退等問題，故採用轉置的方法以應對。這看似容易，然而在規劃和執行上，卻可以發現其中的困難：無法預測新的技術何時會出現、成為市場主流；新的技術和媒體將為何、包括何種功能；是否與現有系統相容；以及該如何避免轉置所造成資料遺失的傷害。而不論進行更新或轉置，勢必會再次投入人力、時間和經費，為達到有效率且有效能的更新與轉置資料，事前細詳的規劃不能少，莽撞地汰舊換新並無濟於事。研究未來科技發展的趨勢，瞭解保存典藏和使用者的需求並加以調和，都有助於轉置的進行。而多從事各種可行性研究和成本效益分析，對於克服前述的困難仍有幫助（註 13）

## 三、模擬（emulation）

數位資訊資源難以長久保存的原因之一，在於數位資訊資源越來越依賴軟體，唯有在有軟體可以讀取和顯示其中內容時，才是有意義而「存在」在的。現在已經發現有些資料格式無法存取解讀，儲存媒體沒有適當的硬體設備讀取其中

內容等問題，在科技不斷發展的情況下，未來勢必亦面對相同的問題。而 Jeff Rothenberg 提出的解決方案是將檔案（如資料、文件等）和其應用軟體，包括支援該軟體的系統和硬體環境描述都一起儲存，以便能夠執行整套程式。未來，硬體環境描述將用來建構一個模擬器(emulator)，模擬原始硬體環境，於是可以在任何電腦上執行，以讀取原始檔案(original records)。而這整套資訊，包括檔案、軟體和硬體描述，將被模擬以防損毀。（註 14）

這個方法即所謂的「模擬」，指在軟硬體中仿效或模擬的過程，能在未來、未知的系統上模擬舊系統，使得數位物件的原始程式能在未來運作。模擬依賴於原始格式、原始資料的保存典藏，並非保存原軟體及硬體，乃軟體工程師撰寫模擬程式仿效過時硬體平台的行為與仿效相關操作系統。其中牽涉到三種技術的發展：（註 15）

1. 為特定的模擬器發展通用技術，使之能在未來的電腦上運作，且可抓取所有必要的屬性以重建目前及未來數位文獻行為。
2. 發展儲存詮釋資料的技術，此詮釋資料是用來找尋、存取與重建數位文獻，且是以人類可讀的格式呈現。
3. 發展壓縮文獻與其附隨的詮釋資料、程式、模擬器規範等之壓縮技術。

這在技術上相當有趣，但在實際執行上帶來了經常費用的支出以使其運作的議題。當決定要採用模擬策略時，假若相關參與者在意識上與標準上達成共識，此方案應被集中管理、進行，或是集中維護一套標準的軟硬體檔案。有了這些標準，每份文件或數位物件可能不僅需要在詮釋資料中包含一個「模擬編號(emulation number)」，以對應到特定的軟硬體檔案。（註 16）

不同的保存方法適用於不同種類的資料。根據 McCray 與其同事所進行的研究，他們認為模擬與轉置的不同，在於模擬所保存的，不只數位資訊資源的功能，還包括其「外觀」和「感覺」。由於模擬使電腦可以模擬出原本檔案的樣貌，增加生動活潑的價值，所以在應用上，遊戲、虛擬化或用電腦模擬的生物科技實驗等，都是很好的例子。而轉置的方法較適用於資料使用度高，以及保存內容為原始資料的情況。美國國會圖書館的資訊技術服務組自 1960 年代開始，負責圖書館目錄的維護，至今經過了三代軟體格式、三種磁帶格式和六種硬碟轉換。像這

樣以保存原始資料為主的情況，便適合採用轉置的方式，以確保數位資訊資源長期得以取得、使用。（註 17）

#### 四、使用標準

最重要的保存策略之一是利用適當的開放性標準和盡可能避免專有的資料格式。即使標準改變和演變，開放性的標準還是較專有的標準穩定而可靠，專有的標準可能會受限於某一廠商。

為確保數位資源能夠被長久保存與取用，發展結構化的方式描述與記錄用來管理典藏物件所需的資訊，此乃所謂的典藏詮釋資料（preservation metadata）。不同於 MARC、Dublin Core 等用於數位資訊資源發掘與識別的描述性詮釋資料，典藏詮釋資料被歸類於管理性詮釋資料的類別之下，或被稱為技術性的詮釋資料，主要在協助資訊的管理以及數位內容的取用。為達到數位資訊資源的長久保存目的，典藏詮釋資料應包括關於物件檔案與結構的技術細節、使用方法等資訊，還有執行該物件所有活動的歷史，管理者的身份識別，與保留關於哪些人有責及有權執行物件典藏動作的資訊。（註 18）國外對於典藏詮釋資料的研究眾多，其中最具代表性的長久典藏模式是「開放性典藏資訊系統（Open Archival Information System, OAIS）」。<sup>18</sup>此一架構由 NASA's Consultative Committee for Space Data Systems 所發展，並被認可為國際性的標準。雖然原先是為太空資料所設計，但 OAIS 被廣泛的接受與使用於更廣的資料類型。該模式定義了數位典藏庫的功能與需求，依據標準提供系統間詮釋資料與互通性的架構，讓不同類型的詮釋資料得以交換與再使用。（註 19）當我們從數位化的開端便以標準化的方式進行，原有的相容性、互通性等問題，都能一一解決，且遵循國際標準，便能與國際接軌，做更廣泛而普及的使用，發揮數位典藏最大的效益。

採用標準應是最理想的模式，然而說起來容易，實際實行上有其困難。由於標準牽涉層面多，其中任何一項都相當專業且複雜，不易處理；而使用者需求、組織目標、發展環境的不同，有關標準的共識並難以達成。最後，即使標準產生，亦受資訊科技轉移影響，仍需要轉移或更新，其相關的軟硬體仍需同步提升修正。（註 20）

#### 五、技術保存（technology preservation）

除了將所需要的硬體資訊保存，以進行模擬之外，保存硬體本身亦為可行方案。這種方法應維護完整的過時設備，將任何老舊的軟硬體組態做複製。其中涉及原始應用程式、操作系統軟體以及硬體平台的典藏，強調原始運作環境的行為應被典藏，才能呈現數位物件的面貌。

這可能是一項昂貴的選擇，以長遠而言也是不實際的做法。為了保證每個時期的資料都能被存取、解釋，而保存了每個時期每種格式的硬體設備；但資訊科技的發展迅速，不斷累積各式硬體，於是乎，代表各種技術的硬體博物館儼然形成。當然其中的維護成本和所需的空間也相對地龐大。技術保存方法是否值得採行，值得大家深思。（註 21）

每種保存方法有其優點，亦有其缺點；更重要的是，在處理不同種類的資訊資源，或在不同的環境條件、需求下，為保證在未來能依舊取得使用，所採用的保存方法應有所不同。依據英國由 Tony Hendley 所主持，於 1995 年開始進行的一項計畫，將數位資源加以區分為十大類，包括資料組（data sets）、結構化文件（structured texts）、工作文件（office documents）、設計資料（design data）、圖表（presentation graphics）、影像（visual images）、演講與錄音資料（speech & sound recordings）、錄影資料（video recordings）、地理/地圖資料（geographic/mapping data）、互動式多媒體產品（interactive multimedia publications）等。再依 Daniel Greenstein 所提出的架構，將數位化工作分成七大模組，應用於該研究所界定的十大類數位資源之中，並考量資訊資源的種類、結構、所使用的應用程式，以及管理或使用的方式等情況，就轉置、模擬和技術保存三種方法相比較，發現該研究所建議的保存方法，仍是以資訊的轉置居多，而技術保存與模擬兩者在應用的時程上，多為短期而暫時的因應之策。（註 22）可能是因為技術保存與模擬這兩種保存方法，所牽涉的技術較為複雜，而在執行上的困難度較高，故在選擇保存方法時，除非有特殊的要求，如對於外觀和使用時的感覺的保存之外，否則以其複雜性，往往帶來許多限制而無法實行。

檔案局所的委託研究計畫--「電子媒體類檔案管理制度及保存技術之研究」，該報告彙整出九種不同的數位資訊資源保存策略，包括更新、標準化、詮釋資料、轉置、模擬、封裝、系統保存重複一套系統、及印成紙張或其他可瀏覽媒體等。該報告將此九種方法加以分類，歸納成三級不同的層次，如下表所示：

(註 23)

類別	第一類	第二類	第三類
層次	基礎層	核心層	輔助層
保存方法	<input type="checkbox"/> 更新 <input type="checkbox"/> 標準化 <input type="checkbox"/> 詮釋資料	<input type="checkbox"/> 轉置 <input type="checkbox"/> 模擬 <input type="checkbox"/> 封裝	<input type="checkbox"/> 系統保存 <input type="checkbox"/> 重複一套系統 <input type="checkbox"/> 印成紙張或其他可瀏覽媒體

所謂「基礎層」是指隨著時間的展延，必須經常、定期的實施，是最基礎的工作；「核心層」是保存技術中最為核心與重要的部分；「輔助層」則是指在核心層技術保存數位資訊資源時，若有困難或其他特別因素的狀況下，所需使用的保存方法。(註 24)

總而言之，各種數位資訊資源保存的方法，不應被視為單一、各自獨立的策略，因應不同的情況與需求，加以配合應用，發揮各項技術的功能，整合出最適當、最佳的方案，以求完整地長久保存數位資訊資源。

#### 肆、資訊資源數位化的省思

近年來，各項數位化工作積極的展開，不僅是為了保存知識和記憶，同時也是為了增加資訊資源的可得性。但已有研究和實際經驗顯示，數位化雖然帶來便利與無限的可能性，畢竟不是萬靈丹，仍需要搭配完善的策略和管理方法，以達到終極目標。一方面回歸到原點，重新思考數位化的意義與價值，另一方面則應該積極面對數位化所帶來的挑戰，致力發展優良的數位典藏。從宏觀的角度來看，資訊資源數位化的工作為最基礎的部分，若能將根基打好，則對於後續的使用和應用，都能有所助益。建置優良的數位典藏，無疑是所有數位化計畫的首要目標。而所謂「優良」的意義，從早期環境的實驗性質高，偏重概念的驗證，到整體環境成熟階段，優良的意義提升到使用、可得性和適當性的層次；現在資訊科技和網路蓬勃發展，物件、詮釋資料和典藏不只為計畫所產生的成果，同時也是大眾可以重複使用、重新包裝、和發展服務的對象，所謂優良的指標相同地必須將重點置於促成互通性、重複使用性、持續、可證明 (verification)、以及文件化 (documentation) 等部分，同時也必須符合版權和智慧財產權相關法律之要

求。（註 25）

資訊黑暗時代是一項警訊，而非禁止大眾以數位形式儲存與典藏資訊資源，甚至造成不必要的恐慌。每個時代皆有其不同的發展、特性和需求，在資訊資源的使用和保存上亦是如此。既然資訊時代已然來臨，生活與資訊科技逐漸密切結合；也許科技並不完美，但相信在集合眾人的智慧與努力，必能找到所謂的「最佳方案」，長久保存文化遺產、人類的智慧結晶。

## 註釋

- 註 1 Amy Friedlander, "Digital Preservation Looks Forward," Information Outlook 6:9 (Sep. 2002) Retrieved from Academic Search Premier database (3 April 2003).
- 註 2 Bernard Smith, "Preserving Tomorrow's Memory: Preserving Digital Content for Future Generations," Information Services & Use 22 (2002): 133.
- 註 3 Marshall Breeding, "Preserving Digital Information," Information Today 19:5 (May 2002), p. 48-50.
- 註 4 陳昭珍,「電子資源的長久保存」, <<http://www.gaya.org.tw/journal/m25-26/25-main3.htm>>
- 註 5 Margaret Hedstrom, "Digital preservation: a time bomb for Digital Libraries." <<http://www.uky.edu/iernan/DL/hedstrom.html>>
- 註 6 同上註。
- 註 7 同註 3。
- 註 8 同註 4。
- 註 9 同註 4。
- 註 10 LOCKSS <<http://lockss.stanford.edu>>
- 註 11 同註 3。
- 註 12 陳和琴,「Metadata 與數位典藏之研討」, 大學圖書館 5 卷 2 期(民國 90 年 9 月), 頁 2-11。
- 註 13 周欣鶯,「數位館藏的維護與保存」, <<http://public1.ptl.edu.tw/publish/suyan/51/72.htm>>
- 註 14 Stewart Granger, "Metadata and Digital Preservation: a plea for cross-interest collaboration." <<http://dspace.dial.pipex.com/stewartg/metpres.html>>
- 註 15 同註 3。
- 註 16 同註 14。
- 註 17 同註 1。
- 註 18 Deborah Woodyard, "Metadata and preservation," Information Services & Use 22 (2002), p. 121-125.
- 註 19 同註 4。
- 註 20 同註 13。
- 註 21 同註 14。
- 註 22 Tony Hendley, "Comparison of Methods & Costs of Digital Preservation." <<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>>
- 註 23 歐陽崇榮,「電子媒體類檔案管理制度及保存技術之研究」(台北市:檔案管理局,民國 91 年 11 月), 頁 57, 國家檔案局委託研究報告。
- 註 24 同上註, 頁 74-75。
- 註 25 "Guidelines for Developing Good Digital Collection Projects," Online Libraries & Microcomputers 20:3 (Mar2002) Retrieved from Academic Search Premier database (3 April 2004).

□