

數位保存營運實務理論之初探

張文熙

檔案管理局

whchang@archives.gov.tw

摘要

數位內容保存作業包括保存(Preservation)及維護(Conservation)兩大技術類型。本文以保存者眼光，針對數位保存營運實務所需之技術理論，包含數位保存作業範圍、數位保存內容之定義、封裝保存架構及樣版分析等方法進行初步探討。

關鍵詞：數位保存、電子檔案。

1. 前言

數位內容保存營運的重要問題之一，在於如何因應貯存媒體的脆弱性，避免貯存內容無法繼續使用。一般以備份機制，配合貯存媒體更新、轉置或模擬來因應，但鮮少對數位保存內容，日後是否能夠長期運作加以討論。本文則試以文獻檔案紀錄保存者觀點(Archives Preserver's View)針對數位內容保存營運實務切入，對相關技術需要加以討論。

2. 數位內容保存的目的

數位內容保存之目的在於避免有價值數位紀錄永久消失。而內容價值之判定，目前大多以專家主觀判斷其對某領域知識的重要性來取決。但就保存者而言，保存內容不應僅對個別專家有意義，保存物與其他保存物間必須存在完整情境互動的結構，才具備值得保存史料紀錄(Records)特質。也就是保存內容必先經挑選、評鑑符合文獻紀錄特質者，方得以 archives 稱之。由於，評鑑保存內容價值之標準，與保存媒體及形式無關，數位保存內容是否符合史料紀錄性，則可以檔案學界判定一般文獻紀錄性良窳之基本法則進行，稱為法理分析(Diplomatics)。藉由分析結果建立起完整保存物件間之知識結構(Contextual Structure)，並訂出適當管理方式。此間所謂知識結構，並非指內容意涵或主題分析結果，而著重在與其他保存文獻間關係之描述。因此，保存者(Archivist 或 Records keeper) 選擇處理數位內容保存技術，是以能否在數位環境下保存文獻間情境結構優先考量。

對文獻保存者言，保存問題的重點，在於可否保存內容的真實性(Authenticity)及可靠性(Reliability)，數位資料亦同。所謂真實性指保存內容必須完整(intact)及未被損壞(uncorrupted)。實務上，保存內容(preservation)部分被外力損壞時，可能仍具備真實性；以博物館中藏品為例，部分破損者，如損壞部分無損於對真實性特質認定者，無損

於其應被保存之價值。而內容可靠性(Reliability)，則需檢視保存內容是否具唯一性及識別性。強調必先去內容重複者，確定保存物為第一手資料來源(primary resources)，並可在內容中提供具識別的資料。故處理數位內容時，有增加辨識防偽等機制之必要。

雖然保存內容之價值，並不因保存形式而有所差異；但不同保存形式，可能涉及以不同的技術來延長保存年限。進一步探討數位保存技術則分為兩大面向：Preservation 及 Conservation。Preservation 所指為維持保存物現狀，並防止劣化，僅強調內容具真實性及可用性。大多數現階段數位保存議題均局限於此，如改善貯存材料保存能力及貯存環境以延長保存期，或防止已經部分損壞者繼續損壞等。常見研究均以 Electronic Records Preservation; 或 Long-term Preservation 為名。而 Conservation 則以擴及恢復數位保存物內容、環境及作業完整性。包含修復內容及模擬舊系統環境。數位內容保存可能因為外力或轉置更新，造成資料損壞或遺失，必須修復內容以維持保存物之價值，包含底層數位資料修復、缺漏值(Missing Value)的處理，進而對隱藏內容進行挖掘（如猜出密碼內容）等均屬之，但有能力進行的保存單位不多。

3. 數位內容保存理論

由於數位內容保存環境以及對思考方向的差異，產生不同之保存理論。保存者共通之保存目的在維持保存物內容之真實、完整及可用性；故保存者必須檢驗保存物之可識別性(identity)、周延性(wholeness)及健全性(Soundness)，保存數位內容也必須滿足此一前提；但對原生性數位內容而言，大大提高作業複雜度，並影響數位保存技術之選用及發展模式。主要分為兩派：一為封裝派(Encapsulation)，另一者以樣版(Templates)派。封裝是以資料保存作為思考主軸，以後設資料(Meta data)作為保存真實性之依據，並將所有相關資料共同包裹在一起，作為紀錄保存處理之基本單元。

以樣版派者，則認為數位內容生存環境複雜，不可能存在單一作法足以適用所有數位內容之類型，保存應依個案特性處理。僅能從各類型中選取代表性範例，應用法理分析方法，分析出應用系統中具保存價值之內容，再行施以保存技術為之，形成該類數位內容保存作業參考樣板。換言之，數位保存結果僅從資訊系統中選取重要部分來保存。但現代資訊系統存在多種作業類型，故保存重點也有

所不同：如資料庫查詢時，不同條件查結果不同，稱為動態變化性實體(Dynamic entities)，因此不保存所謂 view 內容；從其他系統輸入結果會導致內容或格式改變者，稱互動性實體(Interactive entities)，則重視保存互動作業模式。如多媒體藝術、音樂、劇場等，則以保存使用者體驗感官效果為主，稱體驗性實體(Experiential entities)；對於科學性資料，如實驗數值重視重複實驗正確性，而非保存物外觀的特質等。相關數位保存觀念可參考加拿大 InterPARES 計劃[1]，其對數位長期保存之定義，針對資訊系統所產生之數位型記錄(electronic records)，分為藝術、科學及電子化政府活動三類，惟均須以具真實性之格式(authentic form)保存者，同時執行保存作業前必須透過法理分析(diplomatics)[2]來確認保存內容之精確(accurate)及可靠(reliable)。內容分析是執行軟體更新、轉置及模擬作業之等保存營運實務之必要前期作業。其原則是使數位內容保存、維護(maintained)及評鑑，能克服貯存媒體脆弱性(media fragility)及快速過時之特性，以滿足紀錄產生者短期及長期應用之需要。

4. 數位內容保存技術標準

現階段數位保存內容仍以文件資料為主，數位文件內容保存架構設計所涉之技術標準，必須配合數位資料之生命周期理論，方能涵蓋完整管理流程。數位內容生命周期理論，分為文件生命周期理論(Documents Life Cycles)及紀錄生命理論(Records Life Cycles)兩類，兩者差異在於介入管理的時機。由於紀錄生成是從 document 至 record 到 records (archives)連續階段，故不同生命周期理論影響管理處理單位的設計(item vs file)。簡言之，紀錄(Records)生成(Creation)之時，並不同於製作文件之時，因為僅有少數文件會被選入保存機關加以保存。文件經過鑑選清理(Disposition)後，不被銷毀(Destruction)或移轉(Transfer)者，方為典藏範圍(records)。另外，保存文件中工作流程(Workflow)相關紀錄與否，足以影響後續保存作業技術標準規畫。

以我國政府公文為例，簽署流程與內容一樣重要；但對西方而言，文件形成紀錄離開業務處理機關後移至保存單位時，簽署流程僅為判斷紀錄真實性參考項目之一，不足以影響價值。以美國為例其保存文獻紀錄的條件，是以內容具唯一性之非現行活動者(an inactive electronic record)為對象。由於文件產生過程可能經過多次修改或會簽，故其選擇所要保存的是符合紀錄真實性條件(as an authentic record)之最近版本(its last instantiation)。雖然國際上數位保存技術標準，不外上述原則進行規畫[3][4][5]：如 ISO Archiving Standards[6]，以及美國之開放檔案資訊系統參考模式(Reference Model for an Open Archival Information System (OAIS))[7]等，但因為保存範圍不同，實質影響數位內容真實性判斷的條件，造成數位簽章應用範圍及版本控制的層次等作法不同。



圖 1 美國之開放檔案資訊系統參考模式
(<http://public.ccsds.org/publications/archive/650x0b1.pdf>)

現存各機關之數位內容檔案，絕大多數並不符合前述條件，如將現有數位內容均改為合乎保存標準之條件，實務經費及人力上都難以克服，顯見落實數位內容保存標準並非易事。

5. 數位內容保存系統的功能需求

就保存者言，數位保存系統功能要自資訊系統建立何為保存紀錄之概念(a concept of record)開始，並能取出足以應保存之紀錄(record)，而非保留資訊系統內之文件資料(document)為目的。保存系統不但能維持紀錄之正確性(accuracy)、可信度(reliability)，以及真實性(authenticity)，並貫穿數位內容生命周期之保存模式鍊(a chain of preservation model)，成為保存系統基礎功能需求。並可由數位內容生命周期管制範圍，反應出保存系統功能。

北美地區管理文件紀錄生命周期的應用軟體，稱為 RDIM (Record, Document, Information Management System)系統；即紀錄、文件及訊息的整合管理系統。其以紀錄群集(Records)為管理的主體，而非文件，與我國文件管理概念不同。系統在數位文件生成周期結束達到清理年限時，由系統判定要保存或銷毀，被留下者方為值得數位保存之典藏紀錄。以美國國家檔案文件署(NARA)所保存數位內容與我國電子公文檔案兩相比較，其定義之基本差異，造成保存系統設計之功能明顯程度上的差異。

由於西方保存系統有紀錄之概念，使得保存系統功能以群集(Aggregations)著手，不但繼承傳統文件紀錄以群集作為保存結構(Archival Bond) [8]之基本概念：傳統文件因存量過大，無法逐件清點處理，則以箱或系列作為管理及傳統檢索工具之主要描述單位，行有餘力時描述箱與箱間文件之關係描述。這種概念影響後設資料、管理系統及制度之設計，包含點收、移轉、編目、應用、保存單元功能設計，均以群體為主。但問題在應用於原生性數位內容時，往往是一個資料庫表格(table)，和傳統文件不同，但群集應用概念卻比較符合文獻研究者及保存者應用習慣，使得實際數位資料系統處理和保管產生困難。由於保存者注重數位內容是否符合紀

錄生成條件，即保存結構形成後，可完成其下紀錄(records)關連性情境脈絡(Archival Context)[9]，成為保存數位化內容的八大項目及保存功能重要標的：

1. 貯存媒體 (medium)
2. 實體格式 (physical form)
3. 知識結構格式(intellectual form)
4. 對象(persons)
5. 動作(action)
6. 情境資訊(context)
7. 集結成檔的結構體(archival bond)
8. 正文主體內容(content)

因以群體來思考數位保存內容，和資訊系統內部處理方式不同，也和我國向來以單一文件物件來進行相關管理設計習慣不同，因此形成數位文件內容貯存結構的設計之困難。雖然，文件群體必由許多單一的文件組成，作業上也要先行檢視單件數位保存紀錄，但其核心思維是必須可以辨識數位內容群體之管理性及紀實性的情境資訊(administrative and documentary contexts)，而非單一文件的描述資訊。這些影響數位內容保存封裝檔的設計之細節，進而影響檔案決策系統的設計概念，衍生出數位內容必須保存不同開發版本，確保情境資訊之完整，導致數位內容保存架構必須支援版本控管的能力，以符合文件紀錄形成基本條件。

美國國家檔案文件署，基此建立出以知識為本之長期數位內容保存結構(Knowledge-Based Persistent Archives)，以保存紀錄內容知識結構之五大情境資料，作為文件紀錄保存知識(archives knowledge)的主軸。

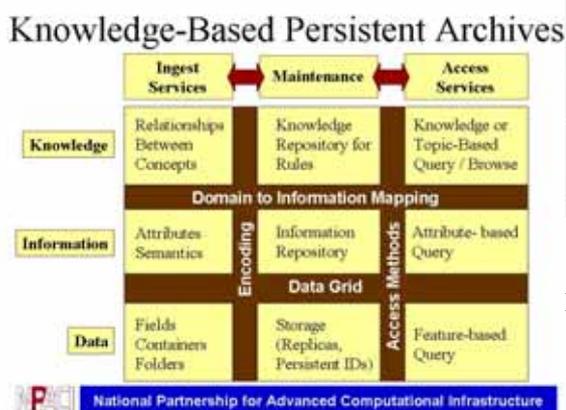


圖 2 美國檔案文件總署規劃之以知識為本之長期數位內容保存結構 (by Ken Thibodeau, “NARA’s Electronic Records Archives Program”)

其所指之知識承續傳統文件保存概念，是為呈現文件紀錄群集間關係，而非單一內容剖析，並定義數位內容情境資料，包含法律及行政(Juridical-administrative)、來源(Provenancial)、程序(Procedural)、文件結構敘述(Documentary)、技術(Technological)、使用環境、呈現效果等相關資料。據此建構自動保存數位化文件知識結構及系統，稱為文件紀錄知識結構(Intellectual Form)。保存內容必須提供闡釋性訊息，以增加內容可靠度，如說明該文件紀錄之組成元素，日期(date)、描述對象(addressee)、敬

語(salutation)及編排方式(arrangement)等、以及處理過程中，所外加用以支援保存或銷毀決策之加註(Annotations)訊息：諸如所記錄事件中，其後續處理或結果，均成為知識價值之所在，而精確保存及勾稽訊息就成為保存系統重要之功能需求。

美國檔案文件總署基於紀錄為主體之概念，規劃之自我描述的電子檔案樣板 (Self-describing Electronic Records Template Example)，套用法理分析原則，其將整個資訊系統視為紀錄(records)，並將傳統文件紀錄之內文闡釋方式，對應到數位內容管理上，進行保存系統設計。如在傳統文件中其開端(The protocol)說明管理情境，如相關之時間、地點、人物及事件等。紀錄事件之原由 (reasons)、緣起(preamble)、發生背景(background)、雜記(narration)、內涵(content)、清理標準(disposition)及其他有關的條件(related conditions and clauses)；在紀錄末端(The eschatocol)記錄驗證文句(corroboration)、簽名(signature)、證書(attestation)、簽署的頭銜、職權等概念，均能與純文字紀錄內容(Plain Text)對應結合，系統就可達成自我描述之功能，充分反應數位內容保存系統整合史料紀錄特性管理之需要。

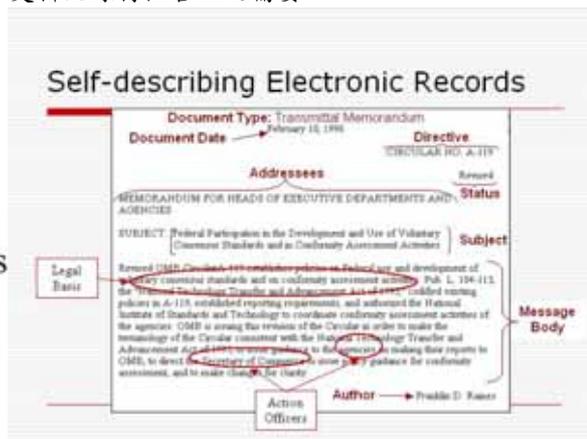


圖 3 美國檔案文件總署規劃之自我描述的電子檔案樣板 (by Ken Thibodeau, “NARA’s Electronic Records Archives Program”)

我國檔案管理局對電子化公文保存，傾向以文件生成開始進行，其理念與澳洲之電子文件連續性生命週期理論相近，立論基於電子文件紀錄離開資訊系統都不能使用，故應自文件生成加以控管，並且以封裝檔作為數位保存之基礎，將保存資料與系統作顯著的切割，以簡化保存作業。但由於未強調保存內容的史料紀錄的群集特性，而側重政府資訊處理特性，在目前典藏來源以公文為主時，尚可因應。惟就未來數位典藏的資料來源、型態、範圍擴大時，群集性的紀錄保存模式似乎較為合於整體文化保存及研究應用的概念，值得再加以討論。



圖 3 我國檔案管理數位內容檔案封裝結構
(資料來源:機關檔案管理資訊化作業要點)

6. 結論

保存者處理數位內容保存問題，是以結合傳統檔案保存學理著手，便於應用為原則，來設計數位內容保存架構，期使不因軟硬體環境及媒體稍加改變就必須重新製作，以減少保存維護負擔。但絕大多數資訊系統設計時，均未以保存者觀點考量其產出內容長期保存之需求。目前數位內容保存作業多僅依賴系統備份機制運作，並未針對軟硬體技術之變革規劃長期因應對策，故數位內容保存的困難將隨時間而突顯。長遠來看，如在資訊系統設計之時，就考量如何取出具價值保存內容(records)，事先加以處理，應可簡化保存作業程序。以資料庫應用系統為例，分散貯存在不同表格中的資料，缺乏具辨識性之情境資料說明內容，對保存者言，不易保存維護。為簡化保存作業，就嘗試將作業環境需求與保存內容加以區隔，使情境資料與內容結合成處理單元，形成封裝檔，並在資料中定義保存作業需求，藉以減輕資訊科技改朝換代的衝擊。然而，以封裝檔為作業單元之系統，並非完美無暇，除明顯地增加開發及應用之複雜度，提高成本外，並不能根本解決已然存在之資訊系統保存問題。使用封裝格式也無法全然避免軟硬體過時(obsolete)，及貯存媒體更新、轉置、模擬之必要，形成推展的主要阻力。

由於封裝貯存有其障礙，故有以建立參考模式取代制訂強制性封裝標準的方法出現，此則以 InterPARES 計畫為代表。其由保存現存之資訊系統內容切入，選擇政府及業界已運作良好之重要資訊系統，區分為電子化政府(e-government)、電子藝術

(e-art)、電子科學活動(e-science)三大區塊，分別進行個案分析，建出應保存內容(records)之模型，再依個別需求選擇適用技術。其主張數位保存實務，與保存內容性質有直接關係，無法以單一方法全面解決保存議題，使得現存資訊系統紀錄內容保存也能在合乎保存者目的規劃下進行。

我國檔案管理局目前首重政府公文資訊的保存，封裝處理之政策似可統一政府公文類資訊保存標準，但對非公文形態之其他政府資訊保存，則仍有待思考。當以文化保存作為前提時，目前封裝之保存標準勢將出現應用的困難，InterPARES 計畫的分析作法就值得我國參考。總體而言，完整的數位保存維運所需考量問題甚多，但若能以保存者之觀點切入保存架構規劃，似乎才能使保存作業得以落實數位文化內容長期保存的目的，未來發展有待諸位先進深入研究解決。

參考文獻

- [1] Luciana Duranti, <http://www.interpares.org>
- [2] Luciana Duranti, *Diplomatics: new uses for an old science*, Scarecrow Press, Lanham, Md., 1998
- [3] Design criteria standard for electronic records management software applications (DOD 5015.2-STD), <http://www.dtic.mil/whs/directives/corres/pdf2/p50152s.pdf>
- [4] Model Requirements for Electronic Records Management Systems (MoReq), <http://www.cornwell.co.uk/moreq.pdf>
- [5] International Council on Archives (ICA), *Guide for Managing Electronic Records from an Archival Perspective*, 1997, http://www.ica.org/biblio/guide_eng.html
- [6] ISO Archiving Standards – Overview, <http://nost.gsfc.nasa.gov/isoas/>
- [7] ISO Archiving Standards - Reference Model papres, http://nssdc.gsfc.nasa.gov/nost/isoas/ref_model.html
- [8] InterPARES, http://www.interpares.org/book/interpares_book_q_gloss.pdf
- [9] InterPARES, http://www.interpares.org/book/interpares_book_q_gloss.pdf