

Metadata 與數位典藏之研討

The Research on Metadata and Digital Preservation

陳和琴 **Ho-chin Chen**

淡江大學資訊與圖書館學系副教授

Associate Professor, Department of Information and Library Science

Tamkang University

E-mail : chin@mail.tku.edu.tw

【摘要】

metadata 在數位化資料的保存或典藏扮演著極其重要的角色。本文試從數位典藏的策略及 metadata 對數位典藏的重要性，進一步探討 metadata 對資訊資源的典藏管理所應該包括的資料要素，及標準的建立等相關論題。

【Abstract】

It is now widely accepted that metadata is a key part of digital preservation strategies. The purpose of this paper is to examine issues related to the preservation metadata. The paper begins by describing the uses of metadata and clarifying key concepts of digital preservation, then moves into the development of existing preservation metadata projects and the specification of preservation metadata elements.

關 鍵 字：數位典藏；詮釋資料；數位化資料保存

Keywords : digital preservation ; metadata ; preservation metadata

壹、前言

metadata 成為資訊時代的熱門論題已有多年。不同人士發展其不同的 Metadata 計畫 (metadata initiatives)，其中雄心勃勃的，例如 OCLC 大力推動的 Dublin Core Metadata Initiative (DCMI)。圖書館界討論 metadata 多半把焦點放在資源描述及發現的議題，其實 metadata 的建立在典藏數位資訊物件上也扮演重要的角色。關係資訊資源的典藏管理，metadata 究竟應該包括哪些資料要素？是否也應該建立標準？如何建立？國外是否有相關的計畫方案可供參考？本文擬就這些方面作進一步探討。

貳、metadata 的類型

metadata 最通俗的解釋是 data about data, information about information。中文譯名或為元資料、超資料、詮釋資料。metadata 的用途或功能，各界看法不一。Michael Day 認為 metadata 的用途包括三方面：資源發現 (resource discovery)、語意互通 (semantic interoperability)、及資訊資源管理 (information resource management)。他將典藏 (preservation) 列入資訊資源管理的範圍之內 (註1)，並且認為 metadata 的概念應該擴及提供數位典藏的訊息。(註2) Wendler 也將 metadata 的功能分成三類：記述的 (descriptive)、行政的 (administrative)、結構的 (structural)。(註3) Gilliland-Swetland 則將

metadata 的類型分成行政的 (administrative)、記述的 (descriptive)、典藏 (preservation)、技術的 (technical) 及使用 (Use) 等五種。其中 metadata 的典藏功能主要在於記載資訊資源的實體狀況、及實體與數位資訊資源的典藏策略。(註4) Gilliland-Swetland 將 metadata 的典藏功能分隔出來，Michael Day 將典藏功能歸屬於資訊資源管理一類，而 Wendler 則將之隸屬在“行政性 metadata” (administrative metadata) 一類。雖然三者類屬各異，但都明顯可見的是：metadata 的典藏功能用來協助資訊資源的管理，有助於系統呈現特別的數位物件給使用者，確保數位資訊資源提供後代子孫使用。Metadata 的典藏功能與數位典藏有著密切的關連。

參、數位典藏概述

數位典藏通常所指的是、數位資料以數位形式典藏的過程。數位資料有些是原生的，本來就是數位資料媒體 (born digital media)；有些則是因為典藏而改變成數位化格式的物件。數位典藏主要目的在於確保所有數位資料的可用性、持久性及智慧整合性。更精確的解釋，數位典藏是數位物件長期的儲存、維護及檢索取得 (accessibility)。(註5) 由於早期的數位化方案多為典藏之目的，所以或有“數位化過程等於數位典藏”的誤解，其實數位典藏更加強調的是長期的維護及可用性。“長期”往往指的

是很長的一段時間，數位媒體會因技術的改變而受到影響。

有兩個主要的難題是數位典藏想要克服的：一為數位媒體的易碎性或強度的限制性（fragility）。據了解，磁性媒體（magnetic media）的壽命短則 10 年，長則 30 年，視其儲存及處理的過程而定。製作廠商 IBM 曾表示，光學性媒體 WORM disc cartridge 的壽命如果在正確的使用環境之下，可以長達 500 年。不過，National Institute of Standards and Technology 則表示 WORM 如果在正常室溫及 90% 相對濕度下，預估其壽命僅長達 57 年，比較寬鬆的算法，也不過是 121 年。這一百多年的壽命或許已經不算短，但是和長達 300 年的微縮資料、及好幾百年的羊皮紙比起來，顯然短少許多。

（註 6）此外，硬體及軟體易於過時作廢、則是數位典藏的第二個難題。

數位典藏比紙本典藏複雜許多，其處理過程由簡單而複雜，可包括 refreshing→migration→emulation→technology preservation。（註 7）

Refreshing 的中文解釋是提神，或使之恢復原狀。就概念而言，Refreshing 基本上就是複製，例如磁片的複製。

Migration 的中文解釋是轉型或移植。Migration 比 refreshing 的概念較廣，是一套有組織的工作，其設計在於達成將數位媒體從一硬、軟體的組態設定（configuration）轉換成另一組態設定，或從某一代電腦技術轉換成新一代電腦技術。Migration 的目的在於典

藏數位物件的完整性；並且面對技術多變之際，讓使用者保留檢索、展現及使用的能力。如何讓數位物件在這急速變化的數位環境中，從一電腦平台安全地“投胎轉世”到另一電腦平台，所牽涉到的是標準化的問題。

Emulation 的中文解釋是實體模擬、仿倣。Emulation 所指的是在軟硬體中仿效或模擬的過程。在某種情況下，待儲存的數位物件需有一為特定操作系統的模擬器，以及為特定應用軟體的模擬器以便處理。emulation 依賴於原始格式原始資料的典藏，並非典藏原軟體及硬體，軟體工程師撰寫模擬程式仿效過時硬體平台的行為及仿效相關操作系統。（註 8）

Technology preservation 的中文解釋為“技術典藏”。典藏資訊使硬體有可能仿倣或模擬，自然也有可能典藏硬體本身。或許這是昂貴的選擇，甚至是不可行的。許多科學博物館想要技術典藏，不過不難想像這有多難，就好像某一應用軟體依賴於磁性媒體的特殊型式，偏偏這個媒體卻停止生產了。

數位物件如何典藏之議題在目前尚停留在研究、實驗的階段，在各種典藏策略當中，migration 及 emulation 受到特別的關注。

肆、Metadata 與數位典藏

隨著資訊科技的進步與發展，我們會越來越依賴數位資訊資源，包括原生或後天的數位資料（物件）本身

及其替代品。然而至今，還沒有一種正式的機制可以確保數位資源得以典藏，作為長期或永久使用。也沒有一個國際性的法定繳存處所，用來承擔這項任務。但圖書館職責所在，必須研究數位典藏的方法及技術，使它具有永久保存、使用的功能。因而數位典藏的議題必須得到普遍的重視。不管數位典藏的策略是 migration、或是 emulation，為了典藏數位物件，都有必要蒐集大量有關數位物件的詳細資料，將之編碼為 metadata。這 metadata 將有助於使典藏的數位物件的檔案得以保存，並且提供後代子孫查檢的可能性。

Rothenberg 指出數位記錄只有在建立此記錄的軟體時才可以查取及了解。增加使用圖檔、超本文、多媒體使數位記錄越來越依賴特定的軟體以為解釋。資料檔若沒有特定軟體解釋其結構及意義，幾近毫無用處。解決之道，就是連同建立此數位記錄的軟、硬體都保存下來。包括記錄、軟體、硬體等整套都納入一個“特殊包裝”中保存以避免損壞。不過問題是閱讀數位記錄的人要曉得如何打開此“特殊包裝”。因此，為了便利未來的使用，Rothenberg 建議另附上“註解的 metadata”在每一“特殊包裝”的表面，解釋如何為納入“特殊包裝”中的過時記錄解碼，及提供有關這些記錄的資訊。這些表面的 metadata 也可包含資源發現的相關資訊，保存時以標準格式，如此才可轉換成新格

式作為數位典藏循環週期 (cycle) 的一部分。此種方法非常技術性，如果協議書及標準能夠建立，有一中央機構能夠維護一套標準化的軟、硬體描述檔，或許不需要每一單位都重複進行此項工作。(註9)

David Bearman 也指出如果記錄轉型到新的軟體環境，其內容、結構及本文資訊必須與軟體之功能相連結。成功的轉型策略有賴於所建立的 metadata，由於記載著數位物件的轉型歷史，未來的使用者可以據此重組或了解該數位物件當時建立的技術環境。(註10)

伍、為數位物件而建立的典藏 metadata 之方案計畫

管理數位館藏需要一有效於決策的 metadata。為了長期保存數位資料，近年來美國、英國、澳大利亞紛紛致力於尋求理想的 metadata 參考模式，發展用以支援數位典藏的 metadata 方案。例如 National Library of Australia (以下簡稱 NLA) 於 1999 年所發表的“Preservation Metadata Set 草案”即是參考許多模式的研究結果。(註11) 這些模式包括有範圍較廣泛的、如開放式檔案資訊系統參考模式 (Reference Model for an Open Archival Information System (OAIS) Draft Recommendation for Space Data System Standards) (註12)，也有特定方案的資料模式，例如 NEDLIB project (註13) 及 NLA 的 PANDORA project。(註14) 此外，還有特定方

案的metadata規格，例如Library of Congress-CNRI Experiment Project（註15）；The Making of America II Project（註16）；the CEDARS project（註17）；the National Archives of Australia的Record-keeping Metadata Standard（註18）等等。NLA的“Preservation Metadata Set 草案”以 Research Libraries Group（RLG）PRESERV Working Group on Preservation Issues of Metadata（註19）為發展的起點。

以下就上述各國主要的努力試作一簡略介紹。

一、開放式檔案資訊系統參考模式（OAIS Reference Model）

OAIS是國際標準組織（International Standards Organization，簡稱ISO）所屬NASA的Consultative Committee for Space Data Systems（CCSDS）於1999年五月所發展的一套標準模式。（註20）OAIS模式可應用於任何檔案，特別是對有責任想要使資訊長期提供的組織機構最有應用價值。應用OAIS模式，該機構能得到國際標準的種種好處。藉著使用共同的參考模式、共同的用詞及共同的概念結構，大家更容易分享想法及交換經驗。不只在圖書館之間，也跨越各類型機構，例如檔案界及圖書館界。許多方案計畫使用這套參考模式迎合其需要，像NEDLIB方案（註21）、Cedars方案（註22）就是其中有名的例子。

OAIS資訊模式將Information Object

分成四類的資訊：

- 內容資訊（Content Information, CI）
 - 典藏描述資訊（Preservation Description Information, PDI）
 - 包裝資訊（Packaging Information, PI）
 - 描述資訊（Descriptive Information, DI）
- 其中的典藏描述資訊，OAIS模式

又將之區分成四種形態的資訊：

1. 參考資訊（Reference Information）：列舉及描述指定於內容資訊的區別號碼，例如ISBN, URN等。
2. 出處資訊（Provenance Information）：記載內容資訊的歷史，例如其起源、典藏作業及效果等等。
3. 內涵資訊（Context Information）：記載內容資訊與其環境之間的關係，亦即為何建立，對其他內容資訊之間的關係等。
4. 固定性資訊（Fixity Information）：記載確保內容資訊不改變的認證機制，例如數位簽名等。

二、美國 Research Library Group 之 Metadata 典藏議題工作小組

早於1997年五月 Research Library Group（簡稱RLG）即成立Metadata典藏議題工作小組（Working Group on Preservation Issues of Metadata）。1998年五月工作報告中推薦16個典藏用途之Metadata資料要素：（註23）

1. 日期（Date）：建檔日期
2. 記錄者（Transcriber）：負責記錄metadata的個人或團體名稱。
3. 製作者（Producer）：負責檔案實體

- 製作的團體或個人。
4. 留存裝置 (Capture Device)：數位照相機或掃描機的廠牌、型號。
 5. 留存細節 (Capture Details)：掃描軟體等等。
 6. 改變史 (Change History)：檔案修正記錄。
 7. 確認機制 (Validation Key)：用以確認電子傳送檔案之機制。
 8. 編碼 (Encryption)：資料傳送前編碼之技巧。
 9. 水印 (Watermark)：指出檔案是否改變以便建立數位指印 (digital fingerprint)。
 10. 解析度 (Resolution)：藉像素 (pixels) 的大小尺寸或每英寸的像素決定解析度，例如 4096 x 6144 pixels; 600 dpi; 320 dpi 等等。
 11. 壓縮 (Compression)：指出檔案是否曾有轉換。
 12. 來源 (Source)：來源的實體特徵等等。
 13. 色彩 (Color)：像素深度。
 14. 色彩管理 (Color Management)：用以增進色彩一致性的系統
 15. Color Bar/Grayscale Bar: 指出有無，如果有的話是何類型。例如 Kodak Q13 或 Q14 Color Separation Guide and Gray Scale。
 16. 控制標的 (Control Targets)：掃描檔的標的之相關資訊。

在 RLG 的 16 個資料要素中，除一兩個亦可以作為資源發現之檢索點外，其餘主要作為典藏管理。

三、澳大利亞的 PANDORA project 及典藏 metadata 的資料要素 (註 24)

在澳大利亞方面，PANDORA Project 全名為 Preserving and Accessing Networked Documentary Resources of Australia (PANDORA)，成立於 1996 年 6 月，以澳大利亞國家圖書館 (National Library of Australia，簡稱 NLA) 可使用的數位檔案為基礎，為專門使用於典藏及長期查檢的澳大利亞網路出版品。PANDORA 方案的目標包括如下：

1. 為典藏及提供查檢澳大利亞線上出版品而發展其政策及過程。
2. 為經過選擇的澳大利亞線上出版品建立檔案，包括電子期刊、組織機構之網站、政府出版品等。
3. 為長期典藏這些出版品的全國性措施提供建議。
4. 成立索引及摘要服務機構以應需要，並且為索引及摘要分配持久性辨識號碼。

NLA 於 1999 年成立工作小組，致力於發展一套專為典藏而設的 metadata 資料要素。NLA 認為“數位典藏 metadata”的使用主要包括下面要點：(註 25)

1. 儲存有關典藏決策及行動 (或策略) 的技術性資訊
2. 記載典藏所採之策略 (例如 migration 或 emulation)
3. 記錄典藏策略的效果 (effects)
4. 長期確保數位資源的認證
5. 附註館藏管理及權利管理 (rights management) 的相關資訊

換言之，NLA 認為典藏用途的 metadata 能用來做為儲存支援數位物件典藏的技術性資訊。除此之外，它也能用來記載紀錄轉型及模擬的策略，有助於確保認證、及記載權利管理及館藏管理資料，也需要與資源發現性 metadata（例如 Dublin Core）之間取得連繫。

1999 年 10 月 15 日 NLA 建議數位物件典藏功能之 metadata 包括下列資料要素：（註 26）

1. 永久性識別號（persistent identifier）
2. 創建日期（date of creation）
3. 結構類型（structural type）：例如 Image, Sound, Video, Text, Database 等等
4. 複雜物件之技術性架構（technical infrastructure of complex object）
5. 檔案描述（file description）
6. 系統需求（known system requirements）
7. 安裝需求（installation requirements）
8. 儲存資訊（storage information）
9. 查檢限制（access inhibitors）
10. 尋找及查尋輔助方法（finding and searching aids）、便利查檢之方式（access facilitators）
11. 許可之典藏行動陳述（preservation action permission）
12. 確認（validation）
13. 關係（relationships）
14. 功能缺失（quirks）
15. 存 檔 決 策（archiving decision（work））：work 指作品的概念。
16. 決 策 之 理 由（decision reason

（work））

17. 負責存檔之機構（institution responsible for archiving decision（work））
18. 存檔決策（archiving decision（manifestation））：manifestation 指作品的實體或虛體的表徵。
19. 決策之理由（decision reason（manifestation））
20. 負責存檔之機構（institution responsible for archiving decision（manifestation））
21. 使用目的之類型（intention type）：例如 Preservation master 或 Access copy
22. 負有典藏責任之機構（institution with preservation responsibility）
23. 過程（process）
 - 23.1 過程之描述（description of process）
 - 23.2 負責過程之機構名稱（name of the agency responsible for the process）
 - 23.3 使用於過程之重要硬體（critical hardware used in the process）
 - 23.4 使用於過程之重要軟體（critical software used in the process）

四、英國的 Cedars Project（註 27）

Cedars Project 是英國電子圖書館（eLib programme）的一個方案。Cedars 方案的全名為 CURL Exemplars in Digital ARrchiveS，正式開始於 1998 年四月一日。Cedars 方案在大學研究圖書館聯盟 Consortium of University Research Libraries（CURL）管理之下，由三所大學 Leeds, Oxford 及 Cambridge 執行。Cedars 方案的目的是在於為“數位館藏

發展政策”發展一“策略性架構”，並改進方法以適用於長期典藏，包括 metadata 的建立。Cedars 方案以 OAIS 參考模式為 metadata 架構發展的起點。

五、NEDLIB 方案（註 28）

網路化歐洲寄存圖書館（Networked European Deposit Library，簡稱 NEDLIB）開始成立於 1998 年元月，由 European Commission's Telematics for Libraries Programme 出資，由 National Library of the Netherlands（Koninklijke Bibliotheek）為首共有 12 個夥伴包括寄存圖書館、檔案館及資訊科技發展者等。有三大出版商支持此一方案、並提供電子出版品以為展示之用。寄存圖書館需要為電子出版品建立寄存系統（deposit system for electronic publications，簡稱 DSEP）。1998 年十二月，DSEP 的概念設計被推薦採用 OAIS 參考模式，類似英國的 CEDARS 方案及澳大利亞的 PANDORA 方案。

六、OCLC/RLG 數位典藏 Metadata 工作小組

2000 年三月，RLG 與 OCLC 對外宣佈共同成立數位典藏 metadata 工作小組（OCLC/RLG Working Group on Metadata for Digital Preservation）。（註 29）數位典藏之 metadata 以下簡稱為“典藏 metadata”。工作小組的主要目標在於發展“典藏 metadata”的全面性架構，使之可以應用於廣大範圍的數位典藏活動，其工作項目列之於

下：（註 30）

1. 發展全面性 metadata 架構。
2. 定義“典藏 metadata”主要的資料要素以支持此架構。
3. 定義及評鑑替代性執行方式。
4. 發展測試平台（testbed）及導行研究（pilot）之應用。
5. 推薦使用“典藏 metadata”的最佳實務及一般方式。

2001 年元月 31 日，在 OCLC/RLG Working Group on Preservation Metadata 的白皮書中指出（註 31），工作小組以 CCSDS 所發展的開放式檔案資訊系統參考模式作為發展“典藏 metadata”架構的起點。工作小組針對“典藏 metadata”的使用，比較了四個不同機構的方案計畫，包括 Cedars Project, National Library of Australia (NLA), Networked European Deposit Library (NEDLIB) 及 Harvard University 的 Digital Repository Services (DRS)。比較後的結論是“典藏 metadata”要凝聚共識、建立共同的標準是可行的。

陸、結語

從上述各國對典藏功能 Metadata 方案的努力，加上國際上已有許多機構呼籲合作建立共同標準的必要性，我們除了應該密切注意其未來動向之外，更應該思考“典藏 metadata”的相關問題，包括國內該類問題究應由誰來主持處理及標準的建立等等。

註釋

- 註 1 : Michael Day. "Metadata developments in the library community."
<<http://www.ukoln.ac.uk/metadata/presentations/flat-ifta/tsld001.htm>>
- 註 2 : Michael Day. "Extending metadata for digital preservation."
<<http://www.ariadne.ac.uk/issue9/metadata>>
- 註 3 : "Preservation metadata for digital objects: a review of the state of the art": a white paper by the OCLC/RLG Working Group on Preservation Metadata, January 31, 2001.
<http://www.oclc.org/digitalpreservation/presmeta_wp.pdf>
- 註 4 : Anne Gilliland-Swetland . "Setting the stage"
<http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html>
- 註 5 : Kelly Russell. "Digital preservation: ensuring access to digital materials into the future."
<<http://www.leeds.ac.uk/cedars/Chapter.htm>>
- 註 6 : 同前註
- 註 7 : Stewart Granger. "Metadata and digital preservation: a plea for crossinterest collaboration."
<<http://ds.dial.pipex.com/stewartg/metpres.html>>
- 註 8 : Michael Day. "Issues and approaches to preservation metadata."
<<http://www.rlg.org/preserv/joint/day.html>>
- 註 9 : 同前註
- 註 10 : 同註 8
- 註 11 : National Library of Australia. "Preservation Metadata for Digital Collections Exposure Draft"<
<http://www.nla.gov.au/preserve/pmeta.html>>
- 註 12 : Consultative Committee for Space Data Systems (CCSDS) , CCSDS 650.0-R-1, May 1999. "Reference Model for an Open Archival Information System (OAIS) Draft Recommendation for Space Data System Standards."<<http://www.ccsds.org/RP9905/RP9905.html>>
- 註 13 : Koninklijke Bibliotheek. "NEDLIB Networked European Deposit Library"
<<http://www.konbib.nl/nedlib/>>
- 註 14 : National Library of Australia. "PANDORA Project: Preserving and Accessing Networked Documentary Resources of Australia" <<http://www.nla.gov.au/pandora/>>
- 註 15 : Carl Fleischhauer. "Library of Congress-CNRI Experiment Project Proposed Metadata Set."
<<http://lcweb2.loc.gov/ammem/award/docs/nisometa/NISOintr.html>>
- 註 16 : "The Making of America II Testbed Project White Paper." Version 2.0 (September 15, 1998.) <<http://sunsite.berkeley.edu/MOA2/wp-v2.html>>

-
- 註 17 : “Michael Day. “Metadata for Preservation Cedars Project Document AIW01. CEDARS, ”
<<http://www.ukoln.ac.uk/metadata/cedars/AIW01.html>>
- 註 18 : National Archives of Australia. “Recordkeeping Metadata Standard for Commonwealth Agencies.” Version 1.0 <<http://www.naa.gov.au/govserv/techpub/rkms/intro.htm>>
- 註 19 : “RLG Working Group on Preservation Issues of Metadata.” Final Report. May 1998.<<http://www.rlg.org./preserv/presmeta.html>>
- 註 20 : “Reference Model for an Open Archival Information System. Red book.”
<<http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf>>
- 註 21 : “NEDLIB: Applying the OAIS Reference Model to the Deposit System for Electronic Publications (DSEP) ”
<<http://www.kb.nl/coop/nedlib/results/OAISreviewbyNEDLIB.html>>
- 註 22 : Cedars Project <<http://www.leeds.ac.uk/cedars/>>
- 註 23 : 同註 19
- 註 24 : PANDORA Project<<http://pandora.nla.gov.au/documents.html>>
- 註 25 : Brian Lavoie. “OCLC/RLG Working Group on Metadata for Digital Preservation.”
<<http://www.rlg.org/events/pres-2000/lavoie/tsld001.htm>>
- 註 26 : 同註 11
- 註 27 : 同註 22
- 註 28 : 同註 13
- 註 29 : 同註 24
- 註 30 : “About the Metadata Work Group.”
<<http://www.oclc.org/digitalpreservation/wgmetadata.htm>>
- 註 31 : 同註 3