

後設資料發展趨勢與數位典藏的應用

陳亞寧

中央研究院計算中心 系統分析師兼圖書組與 Metadata 組組長

數位典藏國家型科技計畫 後設資料工作組 計畫主持人

摘要

本報告共分為五個部份，包括全球發展現況、全球發展趨勢之分析、數位典藏國家型科技計畫（National Digital Archives Program, NDAP）後設資料應用現況及其分析、研究發現及結語與建議等。首先，在全球發展現況方面，本報告選取四篇有關後設資料發展的論文作為文獻分析的基礎，這四篇論文如下：

- Dempsey, L, & Heery, R. (1998). Metadata: A current view of practice and issues. *Journal of Documentation*, 54(2), 145-172.
- 陳雪華。(民 87)。網路資源組織與 Metadata 之發展。《圖書館學刊》，12，頁 19-37。
- Gilliland-Swetland, A. (2000). Introduction to metadata: Setting the stage.
- Turner, J.M., Moal, V., & Desnoyers, J. (2004). MetaMap: Acronym index(Ver. 1.2) .

另外，本報告選取近年來與後設資料發展與應用極為相關方面的文獻作為補充，求取後設資料在全球發展現況更為全面化的了解，這些文獻涵蓋了下列主題：

- 後設資料著錄規則、控制詞彙與 RDF (Resource Description Framework, RDF)。
- 長期保存 (long-term preservation and archiving) 與著作權權限管理 (rights management)。
- 數位學習 (eLearning) 與電子化政府 (e-Government)。
- 館藏資源描述 (Collection Level Description, CLD)。
- 書目記錄需求功能 (Functional Requirements for Bibliographic Records, FRBR) 模式。
- XML (eXtensible Markup Language) DTD 與 Schema。
- 開放式檔案後設資料擷取協定 (Open Archives Initiative – Protocol for Metadata Harvesting, OAI – PMH)。
- 後設資料註冊中心 (Metadata Registries, MDRs)。
- 應用特徵檔 (Application Profile)。

第二，經由上述探討，本報告分析、歸納出目前後設資料的全球發展趨勢共有

11 項的重點方向，如下：

- 功能 (functionality)：如資源探索 (resource discovery)、描述 (description)、管理 (administrative)、長期保存與著作權權限管理。
- 領域 (domain)：數位學習、電子化政府與科學等領域應用與發展正在興起。
- 社群 (community)：以圖書館、博物館、檔案館與網路上虛擬自治社群 (如 W3C) 為主。
- 描述深度 (level)：向上包括館藏資源 (collection)，向下則以全文為主。
- 應用 (application)：為了適應各計畫與個案的應用，適度採取客製化 (customization) 方法來利用既有的後設資料標準，逐漸形成主流；這種方式名為應用特徵檔。
- 概念模式 (conceptual model)：利用概念模式來增益現有後設資料標準的應用層面與深度，包括了圖書館界的 FRBR、博物館界的 CIDOC CRM、檔案館界的 Records Continuum Model、數位資訊長期保存方面的 OAIS Reference Model，以及 W3C 的 RDF Model 等。
- 語意 (semantics)：後設資料除了注重在資源的探索與描述外，如何應用著錄規則、控制詞彙與 RDF，達成資料品質的一致性、深度化語意描述與標引，乃至於如何在網際網路 (Internet) 上宣告所屬領域空間 (name space)，促成數位資源網路化、互通化 (interoperable)、共享化 (sharing) 與結構化 (structured)，則是近年來後設資料發展重點逐漸從描述與探索朝向語意化邁進。
- 資源交換 (interchange)：後設資料如經由 XML 的包裝與規範化，達成資源後設資料結構的明確定義與互通有無。
- 協定 (protocol)：鑑於各式後設資料標準描述深度的詳簡不一，在 OAI 的推動下，OAI-PMH 協定逐漸成為數位資源的共通查詢協定，以利網路上的檢索與傳佈。
- 框架 (framework)：在 Warwick Framework 概念推出後，W3C 也受到影響後並制訂出 RDF，作為融納與共享各式後設資料標準的基礎。
- 註冊與登記 (registration)：如何有效利用既有各式後設資料標準，以及汲取各計畫與個案的實務經驗，經由 MDRs 的建置與註冊，除了促成上述目的外，也可以推廣各標準與個案的應用經驗。

第三，本報告以 NDAP 為研究對象，進行各計畫在後設資料方面的分析與探討。主要重點如下：

- 採用標準：包括藝術博物館類 (如：CDWA、CIDOC Ethno Core Data Standards、VRA Core Categories 等)、檔案館類 (如：EAD)、圖書館類 (如：MARC21、FRBR-based Metadata)、生物類 (如：HISPID、SPECIES 2000、

- Darwin Core)、空間資訊類(如:CSDGM)、數位學習(如:LOM)、電子商務(如:ECML、XrML)、語言類(如:OLAC)與共通類(如:DC)。
- 計畫屬性:以社群(community)、資料類型(data type)、學科(discipline)與功能(function)四大主軸為主要方向。
 - 標準應用:採平行策略,除了採用某一特定屬性的標準外,亦同時採用DC,以達成資源探索與互通。
 - 應用層面:以數位資訊生命週期而言,集中在組織方面。
 - 應用屬性:以物、人、空間與地名為主。

第四,如果將全球與NDAP作一比較,NDAP主要特色在於多元化,主要研究發現有五項,如下:

- 社群:除了集中在圖書館、博物館與檔案館外,也包括了標本館。
- 學科:以藝術、人文、社會為主外,也涵蓋了生物多樣性。
- 功能:也包括了電子商務。
- 標準應用:並不以單一標準為主,常有同時應用兩種標準的實務案例與計畫。
- 應用領域:雖有數位學習方面的計畫應用,但是未涉及電子化政府方面。

最後,本報告從後設資料的研究發展與實務應用等觀點切入,建議未來NDAP在後設資料的應用發展方面,提出下列建議:

- 應用領域(domain):如何與將應用與數位學習予以銜接,並擴展至電子化政府。
- 類型(attribute):考慮是否發展時間與事件類的共通性後設資料。
- 功能(function):長久保存日漸重要,尤其珍貴典藏品可否經由一次數位化作業後,如何妥善保存這些數位資源,以利長久使用,避免第二次原件物品的數位化作業。
- 產生方式(creation):可否經由電腦資訊科技的導入,半自動或全自動產生相關的後設資料,以節省成本的投入。
- 描述深度(level):除針對單一物件的描述,可考慮同時採取廣且深的方式,擇計畫進行館藏描述與全文標誌(markup)。
- 語意(semantics):著錄規範與索引典的導入及其應用,將有助於資訊品質一致性的要求,以及未來聯合目錄的建置。
- 資訊交換(interchange):除了現行的XML DTD外,可留意發展專屬於NDAP的XML Schema,以利於華文數位資源規範與標準的發展。
- 多種標準的整合與應用:應用兩種標準不是特例,未來應用兩種以上的後設資料標準逐漸增多,屆時相關的資料轉換機制與服務也須一併列入考量。