

檔案館社群後設資料發展與數位典藏應用之分析

陳淑君

數位典藏國家型科技計畫後設資料工作組 計畫共同主持人

陳雅惠

數位典藏國家型科技計畫後設資料工作組 研究助理

摘要

本報告旨在對檔案館的後設資料（metadata）標準在台灣地區的應用進行相關探究，共分為五個部分：首先，就國際間檔案館後設資料標準的發展與趨勢進行分析；其次，以檔案編碼描述格式標準（EAD）為例進行標準的深入剖析；接著，以「數位典藏國家型科技計畫」為個案研究，觀察台灣地區檔案館社群的後設資料實施現況；之後，根據文獻與個案分析，本研究提出研究發現；最後，提出結論與建議。

首先，本報告以檔案理論與後設資料之間的關係進行探討，以檔案三理論「尊重全宗原則」、「來源原則」，以及「尊重原始順序」為基礎，討論相對的後設資料元素需求，藉以掌握檔案描述與編排的重要特徵與內容。同時，本報告探討現行檔案館相關後設資料國際標準，包括：MARC AMC（MARC Format for Archival and Manuscripts Control）、DC（Dublin Core）、EAD（Encoded Archival Description）、ISAD（General International Standard Archival Description）等，並釐清這些標準的類別與關係。綜合之，檔案館社群將標準區分為三大類型，包括：資料結構標準（data structure）、資料內容標準（data content），以及資料值標準（data value）。資料結構方面，目前以 EAD 的發展與應用為主流；資料內容方面，各國目前採行自行發展的細部描述規則，如：美國的 APPM（Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries）與 DACS（Describing Archives: A Content Standard）、加拿大 RAD（Rules for Archival Description）、英國 MAD（Manual of Archival Description）等；在國際間則共同遵守高層次的 ISAD(G)（General International Standard Archival Description）。

第二，本研究以檔案編碼描述格式標準（EAD）為例進行標準的深入剖析，包括該標準的發展歷史、發展單位、發展方法、標準的目的，以及標準的結構。EAD

是一個以編碼型式為基礎，具巢狀結構的標準，內容包括三大結構 146 項基本元素，簡述如下：

- EAD 標目<eadheader>：紀錄查檢工具本身的書目性與描述性資料
- 前面事項<frontmatter>：說明<檔案描述 (archdesc)>之前的前言文字，主要敘述查檢工具的產生、出版與使用，而非描述資料的資訊。如：封面、前言、獻辭，及查檢工具的使用說明等。
- 檔案描述<archdesc>：包含一組可重複使用於各層次的檔案資料相關描述資訊，包括檔案實體描述、檔案內容或範圍與背景描述，以及有助資料使用的管理與補充資訊。

第三，以「數位典藏國家型科技計畫」為個案研究，本報告分析臺灣地區檔案館社群的後設資料實施現況。研究的對象是「數位典藏國家型科技計畫－後設資料工作組」在 2001 至 2004 年期間所支援的六個機構共計十二項主題計畫（請詳圖一）。研究資料包括：後設資料需求規格書、內涵分析報告書、後設資料需求表單，以及參考詢問題記錄單等文件。資料分析的結果，分述如下：

1. 主題計畫應用後設資料常見問題分析：根據收集的 53 個問題，分析歸納為三大類型，包括：
 - (1) 標準層面
 - ◆ 通則性與方向性問題
 - ◆ 三類型標準的問題：資料結構、資料內容、資料值
 - (2) 系統層面
 - ◆ 功能需求問題
 - ◆ 目錄功能問題
 - ◆ 轉檔問題
 - ◆ 軟體問題
 - (3) 數位化層面
 - ◆ 基本數位化問題
 - ◆ 命名規則問題
2. 主題計畫資料類型：共包含七類，分別是：
 - (1) 公文檔案 (10 個計畫)
 - (2) 公報 (1 個計畫)
 - (3) 議事錄 (1 個計畫)
 - (4) 圖表 (3 個計畫)
 - (5) 照片 (3 個計畫)
 - (6) 書信 (1 個計畫)
 - (7) 文物 (1 個計畫)

3. 主題計畫檔案描述層次分析：不同主題計畫在控制層次的劃分上通常依其藏品的屬性與特質而產生不同的層次劃分原則，並影響其採用的層次數量；以下以「臺灣文獻館：臺灣總督府專賣局」的藏品層次關聯圖以為示例：

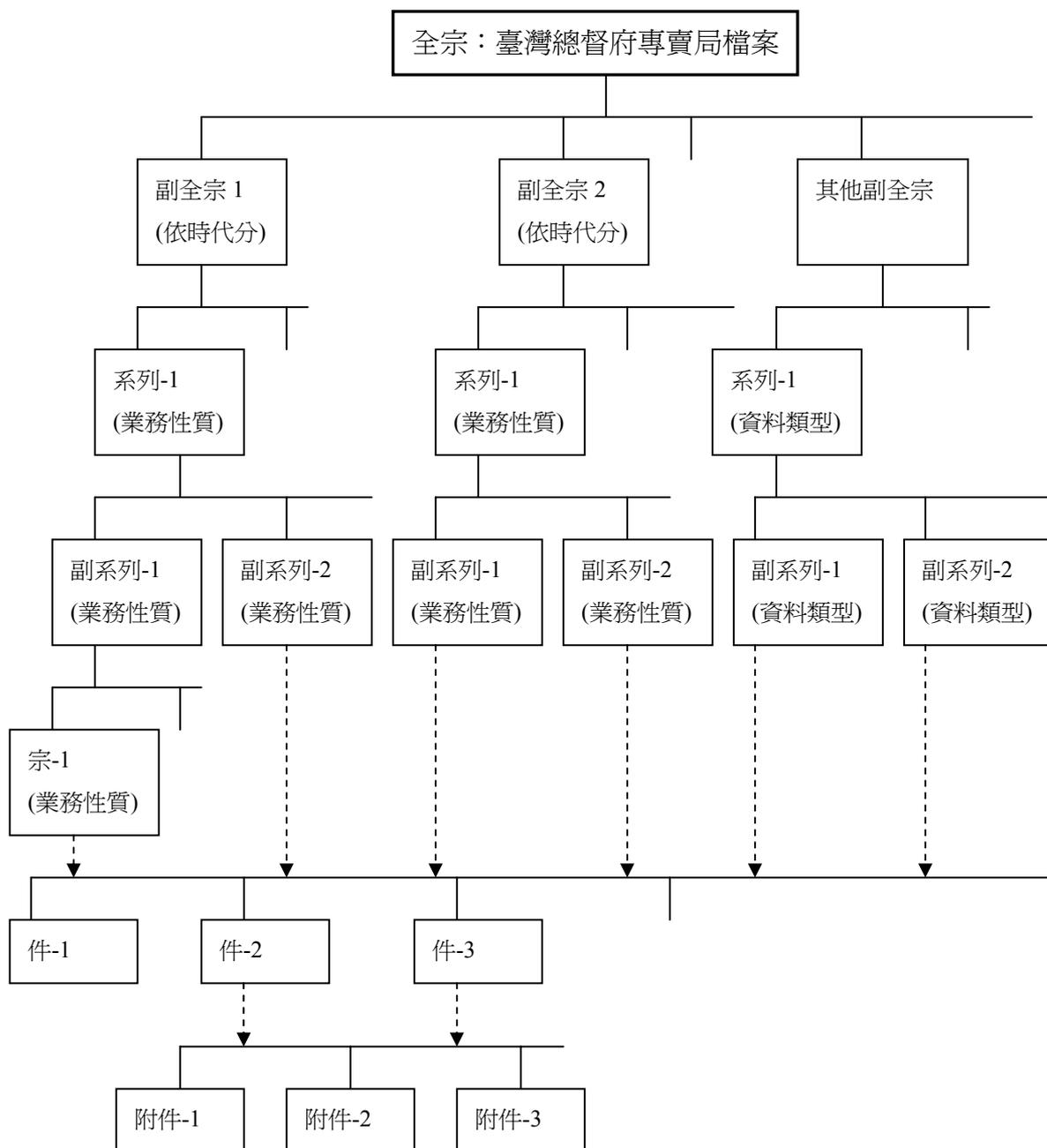


圖 1 臺灣文獻館：臺灣總督府專賣局

為清楚綜觀主題計畫在控制層次的應用現況，本研究以 ISAD(G)所提出的「Model of the levels of arrangement of a fonds」為基礎，將目前實際已應用到的不同層次走向以顏色區分標示出來，各個應用層次走向及採用的主題計畫數量圖示如下：

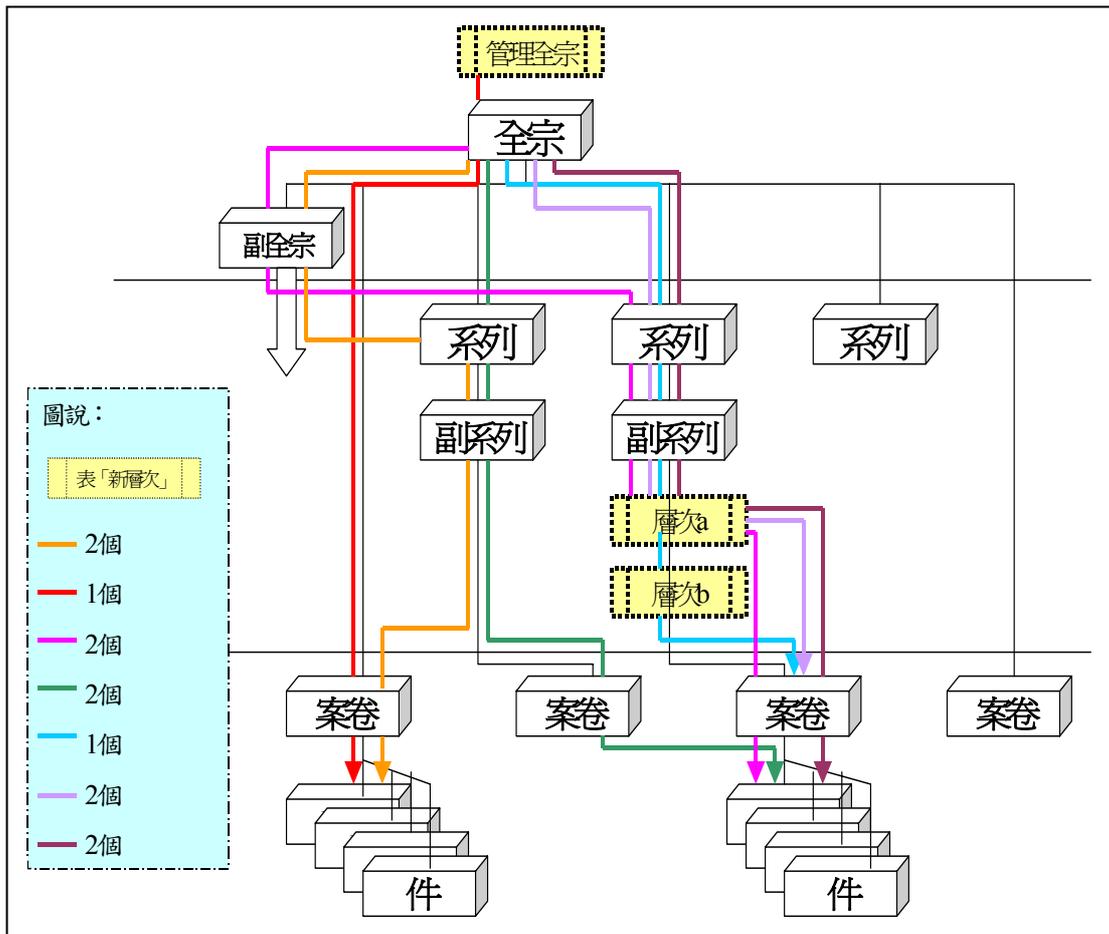


圖 2 主題計畫採用控制層次分析

進一步統計後發現，目前主題計畫在檔案的控制層次採用上，主要介於 4~7 層之間，並且以採用 6 個控制層次者最為普遍。詳細統計情形如下：

- 採用 4 個層次：1 個
- 採用 5 個層次：4 個
- 採用 6 個層次：5 個
- 採用 7 個層次：2 個

其中，採用最少的 4 個層次者，其層次分布為--管理全宗/全宗/宗/件，而採用到 7 個層次者，其層次分布為--全宗(2 層)/系列(3 層)/卷/件。

4. 主題計畫後設資料標準應用分析：在資料結構標準方面，本研究個案的 12 個主題計畫中，各計畫皆採用 EAD 為後設資料標準。統計各計畫的元素清單數量在 45~117 個元素之間，而其中有 38~121 個元素應用到 EAD 中之相關屬性。
5. 資料內容標準方面（檔案描述規則），部分計畫以 ISAD(G)為基礎，並配合個別計畫或組織機構需求，訂定合適的描述規則，其他多數計畫係為自訂著錄

規範。

6. 資料值標準方面（控制詞彙），本研究個案的 12 個主題計畫在其所採用的元素清單中，約有 10 至 30 個元素使用控制詞彙，主要控制詞彙的類型包括：詞控制清單(簡單式代碼表)、主題(結構式代碼表)、權威控制三種。請詳下表所示：

序號	單位	總計畫名稱	子計畫項目	簡單式代碼表	結構式代碼表	權威控制
1	國史館	典藏國家檔案與總統文物數位化中程計畫	國民政府檔案	V	V	人名/地名
2			蔣中正總統文物	V	V	人名/地名
3			資源委員會檔案	V	V	人名/地名
4	國史館臺灣文獻館	典藏日據與光復初期史料數位化計畫	臺灣總督府檔案	V	V	人名/機關/地名
5			臺灣總督府專賣局檔案	V	V	人名/機關/地名
6			臺灣省行政長官公署檔案	V	V	人名/機關
7	中研院近史所	近代外交經濟重要檔案數位典藏計畫		V		機關/職銜/人名/地名
8	中研院史語所	珍藏歷史文物數位典藏計畫	史語所藏內閣大庫檔案	V	V	人名/地名

註：「權威控制」一欄中，標示粗體者代表已經完成權威控制系統開發。

第四，本報告提出兩項研究發現，如下所述：

1. 檔案館社群對於後設資料標準已具共識，需包括「資料結構」、「資料內容」、「資料值」等將三大類型的標準同時搭配，才會形成完整的後設資料標準。由本報告的個案研究也可以觀察到，研究對象在應用 EAD 的過程中，常面臨只有資料架構卻不知如何著錄，換言之，「資料內容」標準的導入是本研究個案目前面臨的需求也是挑戰。在「資料結構」標準 EAD 的應用方面，本研究個案多集中於 EAD 元素的直接應用，對於 EAD 屬性方面的熟悉度與應用較少，顯示未來可以更具深度、更詳盡地研究 EAD 標準與推廣。
2. 本研究個案在描述層次方面，主要描述元素與層次皆分佈於「卷」與「件」層，與國際多數的應用多於「系列」層有所差異。此現象顯示以數位典藏為任務的檔案館，較偏重以更細緻的處理單元，為每一個系列下每卷的每件檔案，進行微觀式檔案描述，相較之下，在全宗與系列層次中的巨觀式檔案描

述，則顯示較為缺乏。

3. 本研究個案除了檔案描述需求，多數也顯示有功能方面需求。換言之，檔案館在進行後設資料的規劃與建置時，多期望能同時滿足檔案描述、館藏管理與運作、資料交換與資料呈現等不同面向的需求。因此，對於檔案館相關後設資料標準的類型與定位需更進一步界定，甚至納入其他社群的相關標準，以符合檔案館社群進行數位工作的完整需求。

最後，本報告提出結論與建議，包括如下重點：

1. 規劃與研發檔案後設資料「最佳實務規範」(Best Practice)。
2. 未來的研發重點
 - (1) 完整 EAD 的系統與軟體發展
 - (2) 檔案描述規則的共識與標準的遵循
 - (3) 不同類型控制詞彙的整合
 - (4) 檔案描述層次的模式