



明清內閣大庫檔案metadata研擬程序

中央研究院 歷史語言研究所

洪一梅 研究助理

93.12.20

前言

- 自2001年起，史語所全面重新規劃數位化作業，針對本所藏內閣大庫檔案，評估科技的進展及國際檔案管理趨勢，重新確認發展方向，主要目的在於尋求更多的人力、物力，加速檔案的整理工作、減緩檔案的毀損、促進檔案的開放，提昇檔案的利用價值，並同時盡到對國家財產保存與維護的使命與責任
- 本文主要著重在metadata研擬的解析，詳細描述了metadata研擬的程序，及過程中考量的及所受的內外影響因素，嘗試提供予檔案典藏單位進行metadata分析的參考

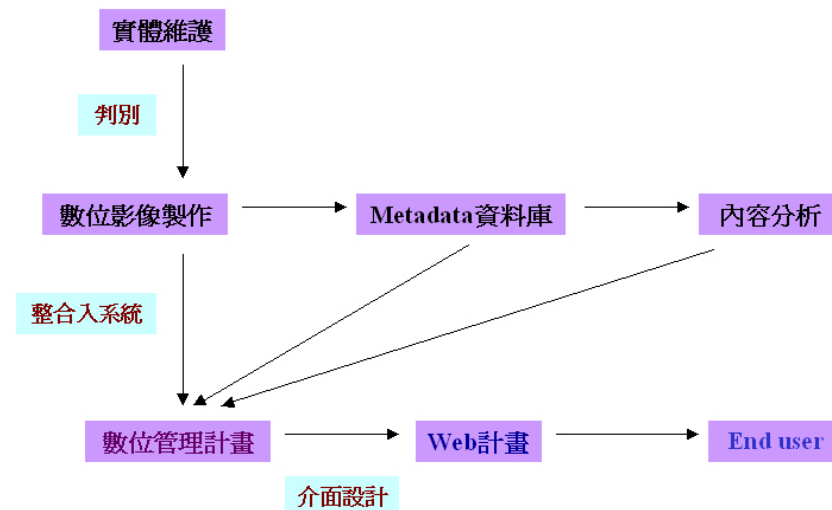
目標與目的

■ 計畫目標

- 整合式的數位管理機制，兼含保存、維護與開放利用

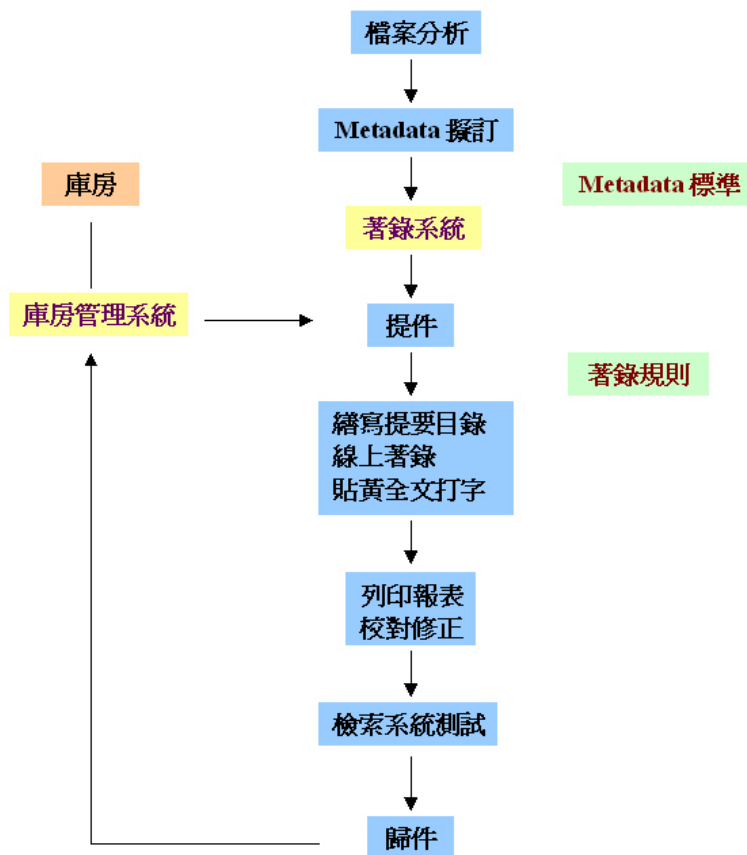
■ 計畫目的

- 數位檔建置計畫
- 數位管理計畫
- web計畫



數位檔建置計畫 - metadata

■ 系統 ■ 流程 ■ 參考標準



[metadata 分析程序]

- 界定使用者及系統目的
- 參考文獻蒐集與研究
- 檔案特質分析
- 系統關聯分析
- 模組欄位組織與關聯
- 人力資源配置
- metadata研擬

界定使用者及系統目的

- 建立典藏資源整理的基礎架構與標準規範
- 預期使用對象為研究人員，研究人員的需求在於「嚴謹」、「正確」、「快速獲得」
- 「整合式檔案管理自動化系統」的發展
- 綜合上述，metadata的分析不僅止於內容資訊的分析，還包含了管理資訊的分析及使用資訊的分析；而分析的程度則是以深入分析、研究性分析為原則

【參考文獻蒐集與研究】

- metadata內涵
- EAD
- MARC21
- ISAD(G)(1999年第二版)
- ISAAR(CPF)(2004年第二版)

參考文獻蒐集與研究

■ metadata內涵

- 管理的metadata(Administrative Metadata)
- 結構的metadata(Structural Metadata)
- 描述的metadata(Descriptive Metadata)

■ 影響

- 決定採用包含三種資訊種類的metadata內涵。
- 權威控制採外部系統控制方式，另建權威資料庫。

參考文獻蒐集與研究

■ EAD

○ 結構

- Header <eadheader> and Front Matter <frontmatter>
- Archival Description <archdesc>
- Description of Subordinate Components <dsc>
- Adjunct Descriptive Data <add> and Other Descriptive Data <odd>
- Controlled Access Headings <controlaccess>
- Pointer, Reference, and Linking Elements
- Table Elements

○ 編碼格式

○ 著錄規則與標準

參考文獻蒐集與研究

■ EAD

○ 影響

- 將內閣大庫視為一個全宗，以現有館藏區分為「明檔」及「清檔」，而這二個層次不在目錄中描述，放置於finding aid的Header <eadheader> and Front Matter <frontmatter>。
- Metadata需包含內容的及層次的描述資訊，並應提供查尋標目，且可連結相關的索引資訊或其它檢索工具，並可指引或直接對映連結數位物件。
- 針對各表單欄位需進行重組。
- 需有可對映的資料值著錄規則。
- 採用XML編碼語言，應先編製DTD。
- 建立與Dublin Core的對映。
- 應隨時注意技術與環境的發展和變化

參考文獻蒐集與研究

■ MARC21

- MARC為長期以來圖書資訊界所使用的機讀格式，甚至在美國許多的檔案典藏單位亦使用它或使用MARC AMC來進行檔案資料的編碼格式，因此整合國際MARC的MARC21成為圖書館使用最多的標準。而在EAD中也進行了二種格式的對映，以利圖書館檔案館之間的資源交換及分享。且EAD格式中亦允許直接註記對映之欄位代碼。
- MARC21針對權威資料有非常詳細的編碼標準。

■ 影響

- 遵循EAD格式進行與MARC21的欄位對映。
- 參考MARC21對人名權威資訊的編碼標準，研擬人名權威資料庫之組織架構，並建立二者間的標準對映。

參考文獻蒐集與研究

- ISAD(G)(1999年第二版)
 - 多元層次描述
 - 欄位分析：區分七個資訊領域
 - Identity Statement Area ; Context Area ; Content and Structure Area ; Conditions of Access and Use Area ; Allied Materials Area ; Note Area ; Description Control Area

參考文獻蒐集與研究

■ ISAD(G)(1999年第二版)

○ 影響

- 考慮內閣大庫檔案的層次問題。界定內閣大庫為「管理全宗(Collection)」；明檔及清檔為「全宗(Fond)」；之下除「冊檔」有「案(File)」的層次外，其餘皆以「件(Item)」層次進行描述。
- 每一層次依其檔案單元之特性，選擇適當的描述欄位，含括七大領域的資訊。不同層次間欄位有可能重覆，但因描述層次不同，資料值應為不同；應避免欄位與資料值皆同者在不同層次間重覆描述，除要做為層次連結的主欄位(primary element)之外。
- 每一個欄位應有定義，及其在系統中的目的與功能，以及著錄該欄位的規則。
- 應注意描述資料中所使用的國標準代碼，如語文、國別等。

參考文獻蒐集與研究

■ ISAAR(CPF)(2004年第二版)

○ 結構

- Identity Area
- Description Area
- Relationships Area
- Control Area

○ 影響

- 檔案權威記錄因應檔案的特質需更具詳盡性及可區辨性。
- 檔案權威記錄所描述的標目即為檔案檢索工具中的檢索點，提供控制功能。
- 檔案權威記錄的欄位設計應包含四種資訊領域。

參考文獻蒐集與研究

■ 綜合影響

- Metadata含管理的、結構的、描述的資訊內涵
- 採用國際標準EAD的簡化
- 採用XML編碼語言，並製DTD
- 進行相關標準的對映
- 需有可遵循的著錄規則
- 檔案權威記錄因應檔案特質具獨立性、詳盡性及可區辨性

檔案特質分析

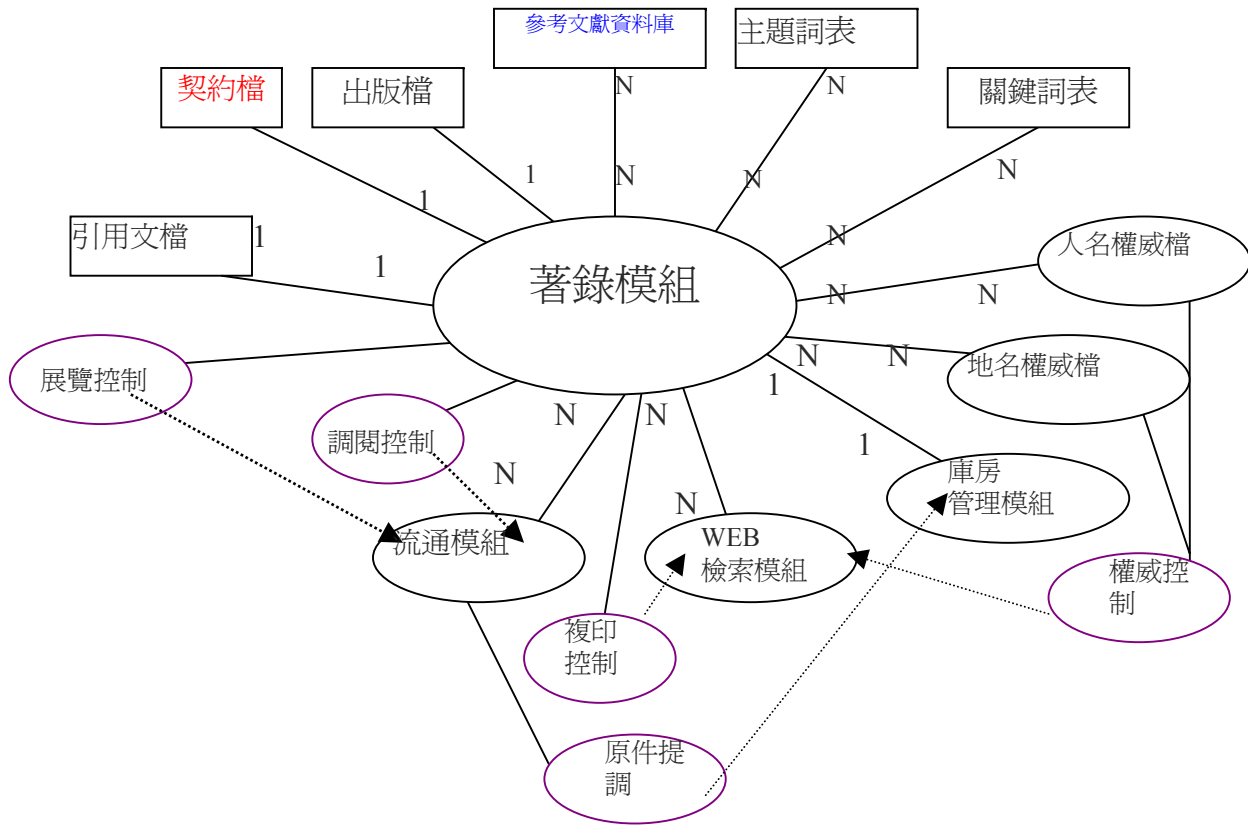
- 為極重要之研究史料
 - 思考描述內容及metadata的廣度與深度
- 為明清內閣大庫之檔案，多為殘件，並已遺失檔案原始順序
 - 避免人工分類破壞檔案原始順序及價值，採「件」為核心描述層次
- 文件類別與格式複雜
 - 欄位名稱及定義應更謹慎思考，避免誤導
- 整理程序複雜，耗時且費工
 - 文件化數位工作流程，標準化數位工作規範

檔案特質分析

■ 影響

- 以「件」為描述層次，若為「冊檔」則增列「案」層次；力求詳細深入與完整
- Tag名稱多採史學術語(如批語、貼黃)；但因格式複雜，為求描述之定義一致，亦有部份採用通用術語(如責任者)
- 以開發整合式自動化系統為目標，應考量整體工作流程，從實體維護、內容整理到開放利用應納入管理及流程的思維

系統關聯分析



模組欄位組織與關聯

- 著錄模組
 - 著錄模組是整合式自動化系統中對檔案資訊描述的主要 schema，因此包含大部份的結構性 metadata 及描述性 metadata
 - 內涵：Administrative Metadata – Right and Access Metadata；Structural Metadata；Descriptive Metadata
- 庫房管理模組
 - 庫房管理模組是最前端與實體管理相結合的 schema，包含有對檔案資訊的簡易描述及部份結構性 metadata 與部份 Preservation Metadata
 - 內涵：Administrative Metadata – Preservation Metadata；Descriptive Metadata – Data that identifies an information package；Structural Metadata

模組欄位組織與關聯

- 權威資料庫
 - 因權威控制採外部系統，所以會與描述性metadata中的權威控制、相關作品、姓名、主題配置、不同實體間的關係、主題標目、分類號、檢索點等具關聯，而其主要資訊內容為針對欲控制之檢索標目進行分析與描述
 - 內涵：Descriptive Metadata – Intellectual organization data(關聯)；
Descriptive Metadata – Intellectual access data(關聯)；
Information package about access point(檢索標目)
- 檢索調閱模組(Finding Aid)
 - 檢索調閱模組與檔案的各種性質metadata皆有關聯
 - 內涵：
 - Administrative Metadata – Meta-metadata
 - Administrative Metadata – Right and Access Metadata(關聯)
 - Descriptive Metadata – Data that identifies an information package(目標項)
 - Descriptive Metadata – Intellectual organization data(關聯)
 - Descriptive Metadata – Intellectual access data(關聯)
 - Information package about access point(檢索標目)(控制項)

人力資源配置

■ 著錄模組VS.庫房管理模組

- 該部份metadata由本所資深助理進行描述，每位至少皆有五年以上的明清檔案讀檔經驗，並由研究員進行審核。研擬有著錄規則，以為遵循

■ 權威資料庫

- 該部份由新任助理在接受過訓練後進行描述，目前為止至少皆有二年以上史書傳記資料研讀與整理的經驗，並由研究員進行審核。研擬有著錄規則，以為遵循

■ 關鍵詞VS.主題詞

- 該部份由所有助理進行，過程中由研究員帶領，每月進行關鍵詞圈選及原則歸納。日後以所圈選之關鍵詞進行系統分析，以整合出適用之主題詞。目前仍在測試階段

[metadata 研擬]

- 欄位分析
 - 案；件
- 欄位簡易定義
 - 每一欄位給予要著錄的資料值定義，以區辨不同欄位所代表的資訊內容及作用。
- 欄位組織(至系統中)
 - 依據欄位分析表及工作流程，將欄位組織至不同的系統模組中，並進行系統及功能分析。
- 實作
 - 實作過程中，因工作流程的調整或是著錄之難易與問題發現，進行欄位、欄位定義或工作流程、工作程度或系統功能的調整。

[metadata 研擬]

■ 著錄規則擬訂

- 因由多人同步進行檔案描述，過程中以討論方式，將每個人對metadata中每一的認知進行協調與統合，並考量系統設計及檔案特性，付予制度化的著錄規則，重新以條文式撰寫欄位定義及著錄規則，以為遵循。

■ 欄位修訂

- 欄位名稱的改變；分欄或併欄；取消欄位或新增增位

■ 持續修訂

- 全方位審視，包含著錄規則

機構間的整合

- 對Metadata的影響
 - 增加對「finding aid」的描述—header(權責描述)
 - 突破TAG的字面限制，尋找定義的整合與對映
 - 修訂著錄規則
 - 外部一致性；內部彈性
 - 促進各典藏單位對相同領域之典藏更密切的整合
 - 跨典藏機構間考慮檔案層次的界定

結論

研擬程序	內容重點
計畫目標、目的及預期對象 檔案特質分析 參考文獻	Motive Resource Reference
系統分析	Extent Function Purpose
欄位分析	Structure Context Content
實作與研擬	Try and Experience Principle and Standard
修訂與整合	Review Evaluation
Metadata	Product

報告完畢，敬請指教

本文感謝中央研究院歷史語言研究所
劉錚雲教授、何漢威教授、陳國棟教授
及明清檔案工作室同仁的協助