

第一章 緒論

第一節 研究背景與研究動機

資訊科技的日新月異與網際網路之蓬勃發展，徹底改變了我們所處的資訊環境並影響資訊取得的方式。長期以來，各類典藏機構，如圖書館、檔案館、博物館習慣以「單件」(item)為資源描述的單位，由於描述資源的單位最主要是以「單件」為主，以編目層級(cataloging level)而論，屬於單件層次描述(item level description)。傳統之單件層次詮釋資料(item level metadata)雖然可以對單一資源做詳細描述，然而對一個身處數量龐大、資源分散與不瞭解典藏特性之使用者而言，詳細的單件資源描述仍是不夠。蘇軾詩云：「橫看成嶺側成峰，遠近高低各不同。不識廬山真面目，只緣身在此山中。」可巧妙比喻使用者身陷龐大資源與不聞典藏特性時之處境。因此，典藏機構如何有效組織為數龐大的資源，並讓使用者在使用時能夠看到資訊的「全貌」與找到所需的資訊實為當務之急。

此外，藉由數位科技的普及，網路上存在著無數同性質的資料，但如何以有效地方式，讓使用者跨越不同系統，透過網路，快速找到相關連的資料，也是一大挑戰，數位典藏系統的互通性問題是跨越的關鍵。

為了解決上述『單件層次詮釋資料在某些情境下無法發揮揭示典藏資源的存在與可利用性』、『幫助使用者瀏覽為數龐大的異質性與分散式數位資源』以及『典藏機構之間資源互通性』等問題，全集描述(collection description)的問題受到重視。

全集描述對圖書館界來說並不陌生。傳統上，圖書館在做全集描述時大多是以非正式、無特定結構的文本形式來描述典藏的資源，例如以紙本方式出版館藏簡介或在圖書館的網站中介紹館藏資源與特色；然而由於這些描述並無標準化格式，不同的圖書館採用不同的方式來描述館藏特色，因此很難將這些描述加以組織與定位出該描述之位置，大大降低此描述的效用。全集層次詮釋資料格式就是以典藏資源的集合體(collection，本研究以全集稱之)做為描述的單位，不同於以典藏資源中的「單件」為描述的對象。更進一步來說可以視全集層次描述為一個結構化、開放的、標準化與機器可讀之詮釋資料格式(Macgregor, 2003)。

透過全集描述建立起全集層次描述，除了可以讓全集管理者更方便管理全

集之外，對一般的使用者來說，可以從廣大的資訊森林中找出自己有興趣的全集、評估搜尋到的全集是否符合所需與定位出組成該全集之單件所在位置。古云：「見樹又見林」，意思是對全局掌控的重要性；透過全集層次描述讓使用者在資訊查詢時，不會有只見樹不見林的困擾。

全集層次詮釋資料之標準仍然處於發展的階段，尚未有普及的國際標準格式。有些詮釋資料，雖未以「全集」為名，但卻有「全集」描述之實，如用在歷史檔案的「檔案描述編碼格式」(EAD)。國內外以「檔案描述編碼格式」(EAD)來描述檔案的案例很多，但將單件轉為全集的卻很少。

事實上，現有的典藏描述多以單件為單位，若要以「全集」方式呈現給使用者，最有效的方法就是自動由單件轉為全集。

因此本研究希望以「國民黨黨史館」現行用於單件檔案描述的「國民黨歷史檔案詮釋資料」以及「檔案描述編碼格式」(EAD)為例，透過可延伸樣式表語言轉換 (eXtensible Stylesheet Language Transformation, 簡稱 XSL Transformation 或 XSLT) 進行詮釋資料格式結構轉換實作，建立起國民黨歷史檔案之全集描述，探討由目前「國民黨歷史檔案詮釋資料」底層的單件層次詮釋資料轉換成「檔案描述編碼格式」(EAD)全集層次詮釋資料，發掘自動轉換過程中 (由單件層次描述自動轉換為全集層次描述，由平面轉換為階層) 可能會產生的問題與解決之道，最後並探討如何達到全球化之全集描述與檢索。

第二節 研究目的

本研究之目的在探討全集描述之發展趨勢。具體而言，本研究之研究目的如下：

- 一、探討全集描述發展及應用趨勢。
- 二、探討檔案特性與全集描述之需求。
- 三、探討由單件層次詮釋資料轉換為全集層次詮釋資料、平面資料自動轉換為階層性資料時可能發生之問題。

第三節 研究問題

本研究主要在探討全集描述發展趨勢及其相關應用技術，以「國民黨歷史

檔案詮釋資料」與「檔案描述編碼格式」(EAD) 為例，透過可延伸樣式表語言轉換(XSLT)進行詮釋資料格式結構轉換實作，探討自動全集轉換相關問題。具體而言本研究之研究問題為：

一、全集層次詮釋資料方面

1. 全集層次詮釋資料目前之發展趨勢為何？
2. 全集層次詮釋資料若要應用於跨領域檢索與達到全球性描述與檢索時所面臨之問題？

二、「檔案描述編碼格式」(EAD) 方面

1. 檔案的特質為何？全集的特性為何？全集的需求面為何？

三、單件轉全集問題方面

1. 單件轉全集之必要性？
2. 單件轉全集問題為何？資料如何對映？轉換上會遇到的問題？

四、從單件層次轉換成全集層次技術方面

1. 就 XML 文件而言，轉換的工具為何？
2. 從單件層次自動轉換成全集層次時會面臨的問題與解決方法？

五、全集描述呈現介面

1. 檔案全集合理的呈現方式為何？需提供的功能為何？

六、「檔案描述編碼格式」(EAD) 與現行全集描述詮釋資料之關係

1. 「檔案描述編碼格式」(EAD) 與現行全集描述詮釋資料之關係為何？

第四節 研究範圍與限制

本研究於人力及時間等因素之考量下，有以下之範圍與限制：

- 一、在理論探討方面，將探討各種全集描述的發展現況；但在實作方面，主要以檔案全集描述為主，尤其著重在「檔案描述編碼格式」(EAD) 全集模式的轉換，不涉及其他全集描述格式的处理。

- 二、本研究主要研究對象以「國民黨黨史館」中所典藏之歷史檔案為主。目前由「國民黨黨史館」典藏之歷史檔案為數龐大，達三百多萬件之多，囿於人力及時間因素，本研究將針對其典藏中的部份歷史檔案為主要研究實作對象，包含「國防最高委員會檔案」、「國防最高委員會會議記錄」、「中央改造委員會檔案」、「中央改造委員會紀錄」、「上海環龍路檔案」、「中央改造委員會各種工作簡報」、「第七屆中央常務委員會會議記錄」與「吳稚暉先生檔案」。
- 三、本研究所蒐集閱讀之文獻資料以中英文文獻為主，其他語文不在研究範圍之列。

第五節 名詞解釋

一、全集 (collection)

「collection」中文可譯為聚集、收藏或文集等等意涵。「collection」應用在圖書資訊學，常用來指圖書館當中的「館藏」；本研究以歷史檔案為主要研究實作對象，因此以「全集」稱之。

「全集」最簡單的定義為「任何單一項目（或稱資源、物件）的集合體」。構成「全集」之單件類型可以是實體單件或數位化單件。一個「全集」必須由多少個單件所組成並無嚴格的限制。在理論上，一個「全集」當中僅由一個單件所構成之情況是可能發生的。(Johnston & Robinson, 2002)「全集」更為嚴謹的定義為「依照某個準則 (criterion) 或特性所構成的資訊資源或數位化集合體」。

不同類型的典藏機構環境下例如圖書館、檔案館、博物館與美術館，由於典藏的資源特性與管理文化上之差異，除了對全集之概念有不同的認定與定義之外，也產生不同的全集型態。例如圖書館中常見的主題全集，就是以書的內容主題當作是聚合全集之準則。

本研究主要研究對象為國民黨歷史檔案，因此是以檔案界的角度來定義「全集」的概念，「全集」在本研究所指的是同一個起源（組織機構）的歷史檔案之集合體。

二、全集描述 (collection description)

「全集描述」就是全集的檢索工具 (finding aid)，即對資源集合體進行描述，提供使用者查詢全集的檢索工具。描述一個全集可能提供以下資訊：

(Johnston & Robinson, 2002)

1. 視資源集合體為一個整體的資訊。
2. 組成全集的個別單件資訊。
3. 單件之群集 (grouping) 所構成的子集 (subset) 資訊。

基於上述，Heaney 建議將全集描述 (或稱檢索工具) 細分成四種更小的類型，分別為「整體性檢索工具」、「解析性檢索工具」、「階層性檢索工具」與「索引性檢索工具」等四種類型，分述如下：(Heaney, 2000)

1. 整體性檢索工具 (unitary finding aid)：整體性檢索工具將全集視為一個整體進行描述，僅描述全集的整體資訊，不深入描述構成該全集的單一項目資訊。
2. 解析性檢索工具 (analytic finding aid)：解析性檢索工具提供單一項目及其內容的資訊。例如圖書館之目錄即屬於解析性檢索工具。
3. 階層性檢索工具 (hierarchic finding aid)：階層性檢索工具提供關於整體與項目之間資訊，包含有關項目對整體之來龍去脈。例如檔案檢索工具就是屬於階層性檢索工具。
4. 索引性檢索工具 (indexing finding aid)：索引性檢索工具記錄來自全集中項目的資訊，但不記載項目的內容。例如網路搜尋引擎會索引文件中的字，但不會記錄辨識文件中的內容為何。索引性檢索工具之目的是在加快檢索時的速度。

本研究實作部分由單件層次 (item-level) 轉換成全集層次 (collection-level) 即是希望能夠達成全集描述中的整體性檢索工具，僅提供檔案較高層次資訊，方便使用者在瀏覽為數龐大的歷史檔案資料時能夠快速發掘所需的歷史檔案。

三、全集層次詮釋資料 (collection level metadata) 與全集層次描述 (collection level description)

「全集層次詮釋資料」與「全集層次描述」兩名詞不同，然而所代表之意義是相同的。對「全集」進行描述的詮釋資料可稱之為「全集層次詮釋資料」。更進一步來說可以視「全集層次描述」為一個結構化 (structured)、開放的 (open)、標準化 (standardized) 與機器可讀格式 (machine-readable form) 之詮釋資料，它提供單一項目集合體一個較高層次的描述 (high-level description) (Macgregor, 2003)。

四、歷史檔案

通常是對一定歷史時期所產生檔案的習慣稱謂，歷史檔案一般都具有年代久遠、載體特殊、所記載內容史料價值較高等特點，在保管和利用上具有特別的意義。在實際工作中所說的歷史檔案有時是只相對於現行機關形成的檔案而言，過去的機關形成的和已經移交檔案館的檔案，習慣上亦稱為歷史檔案。此外，舊政權檔案和革命歷史檔案或保管了一定的年限，國家政策允許向社會公眾開放的檔案亦可稱之歷史檔案。(來新夏，民 80)

五、全宗 (fonds)

一個機關、組織或個人在社會活動中形成的檔案的有機整體。是檔案的基本分類和管理單位。全宗既是一定數量檔案的實體概念，也是整理檔案原則的重要體現和方法，在檔案學中有時稱為全宗原則和全宗理論。有的國家不用全宗概念，如美國的文件組合、英國的檔案組合等，其含義與全宗大同小異。

全宗的成分：一個全宗的檔案是歷史地形成的，它是由特定的來源單位在履行職能任務過程中積累下來的原始記錄組成的不可分割的有機體系，其成分包括該機關各種門類的檔案，各種記錄方式、各種載體和文種形式的內部文件、收文和發文。全宗是檔案館對檔案進行整理分類、編目和日常管理，以及國家對檔案進行統計監督的基本單位。在檔案的整理、流轉和保管過程中，同一全宗的檔案不得分散，不同全宗的檔案不得混雜，這被稱作檔案管理中全宗不可分散性原則。

全宗的形式：主要分為常規全宗和特殊形式的全宗兩種類型，常規全宗即一般情況下的獨立全宗，有的國家也稱其為標準全宗；在難以區分或不必要區分獨立全宗的情況下，則採取聯合全宗、彙集全宗和檔案彙集等辦法，作為全宗的特殊形式。(中國大百科智慧藏)