

# 北平「世界日報」：民初歷史性新聞報紙數位化之研究

## Digitization of the Early Republic Newspapers: A Case Study of Peiping's "Shih Chieh Jih Pao"

余顯強

Shien-Chiang Yu

世新大學資訊傳播學系副教授

Associate Professor, Department of Information and Communications

Shih-Hsin University

E-mail : ysc@cc.shu.edu.tw

### 【摘要 Abstract】

民初新聞報紙具有相當豐富的歷史研究價值，是瞭解當時政治、社會、經濟、生活型態等情形的絕佳工具，但是歷史性報紙體積龐大、紙張脆弱，不僅不易保存，也不易檢索，因此數位化便是使歷史性報紙重獲新生命的唯一途徑。

本文以世新大學北平「世界日報」數位化計畫為例，擬藉由實體原件品質、Metadata制訂與著錄、搜尋結果呈現等特質，探討民初新聞報紙數位化與現代報紙數位化不同之處。並針對民初新聞報紙數位化潛在的議題加以分析與建議。

The Early Republic newspapers are materials replete with valuable literature for historical study, since they may serve as useful resources for understanding the backgrounds of politics, society, economy, and lifestyle of the Early Republic. However, the Early Republic historical newspapers are usually large in page size and fragile in paper quality, which causes difficulties both in preservation and in information retrieval. To overcome the difficulties, digitization is a significant medium to refresh the values of historical newspapers.

This paper takes as an example the digitization project of Peiping's "Shih Chieh Jih Pao" processed by Shih-Hsin University. Through an analysis of the quality of original physical materials, definition and formulation of metadata, and display modes of search results, this study attempts to reveal the different conditions of digitization between historical newspapers and modern newspapers. This paper also offers suggestions about potential issues and problems of the digitization of historical newspapers.

### 關鍵詞 Keyword

民初新聞報紙 歷史性報紙 數位典藏 北平世界日報

Newspaper in the early republic ; Historical newspaper ; Digital archives ; Peiping Shih Chieh Jih Pao

## 壹、前言

報紙為每日發刊之連續性出版傳播媒體，其所記載的新聞訊息兼具內容豐富、論點深入等優點，因此歷經幾個世紀的發展與考驗，不僅在電子媒體與網路媒體的挑戰中，仍佔有傳播媒體之重要地位，此外報紙更扮演了為電子媒體設定議題的規範性角色。(鄭植榮，2004年，3月)

報紙除了提供新聞訊息的功能之外，對歷史文化的保存也有相當貢獻，其特點不僅是第一手文獻、內容豐富、包含全社會的檔案，且具備了有序性與可近性的特點，甚至有許多內容為報紙獨有。(張琪玉，1999年，頁7)(李學軍、鄧小川，2001年，頁69)歷史性報紙中不僅可以看出對當時新聞事件的報導，也可以從社論或新聞事件的發展之中，窺知當時的政治、經濟及文化局勢，傳遞當時社會普遍的價值觀；而報紙中附帶的副刊、廣告、新聞圖片，更富有歷史價值，能夠顯示出當時人民休閒、民生之概況。

雖然報紙資料內容包含豐富多元的資訊，但報紙出版的考量以一次性(Ephemeral)為主(Deegan, King, Men, & Steinvel, 2001, July)，著重在快速印刷以立即傳遞當天新聞資訊，因此採用的印刷技術與紙張品質必然不利於保存。經過長久時間的流轉，眾多讀者的翻閱使用，光線照射、保存環境的溫濕度不佳等因素(林淑芬、宋美珍，2002年6月，頁198)造成報紙紙張的酸化、脆化，而不利於長久保存。

由於網路科技的發展，現代報紙媒體已有兼營網路電子報的趨勢。以全世界來看，西元1994年開始有報紙媒體在網際網路上建立網站，發佈網路電子報。到了1994年底，共有78家報紙發行網路電子報。西元1997年，網路電子報已發展到1900多家。1998年10月初，全世界的網路電子報已增長至4295家，1998年底增至4925家(姚雪痕，2003年，5月)，

近幾年更有驚人的成長。這樣的趨勢奠基於現代報紙編輯流程的電子化，平面報紙編輯完成之後，新聞資訊很容易便能轉進數位資料庫，在網路上進行第二次的利用。因此現代報紙的數位化過程並無太多困難度。相對而言，早期發行的報紙，因長時間累積不僅數量極為龐大，並且由於儲存保管上的困難，可能造成很嚴重的損害情形，若有缺期、破損情形，亦不容易補齊實體。

除了保存文化資產之外，數位化還能夠賦予歷史性報紙新生的價值。透過資訊系統的輔助，不僅歷史性報紙的新聞內容能夠容易地被查找，更能達成許多超乎報紙原有功能的應用。例如聯合報的「聯合知識庫」計畫，將聯合報系50年來的歷史性報紙數位化，並開放給一般民眾使用，特別是除了學術研究者、其他媒體從業人員、圖書館加入會員之外，尚有許多警察局等警政單位亦上網利用聯合知識庫，從歷史性報紙中找尋早期類似作案手法案情或未破案的線索，協助刑案的偵辦。(何銘傑，2001年5月，頁11)不只是現代報紙數位化可以方便讀者使用資訊，歷史性報紙的數位化也能夠帶來相當大的貢獻。

## 貳、北平「世界日報」數位化計畫簡述

北平「世界日報」(以下簡稱「世界日報」)數位化計畫包括兩個計畫，分別是隸屬於「數位典藏國家型科技計劃－內容分項發展計畫」的「北平世界日報內容數位化開發計畫」與「技術研發分項計畫」的「後設資料動態管理系統研發計畫」。為紀念世新大學創辦人成舍我先生，在1998年開始著手將成先生創辦的北平「世界日報」、「上海立報」和「香港立報」進行報紙數位化典藏計畫。希望能將民國初年的歷史性報紙新聞資料，透過數位化的方式再現，另外也希望能透過數位化的工作，延長世界日報的保存年限，增進其使用的便利性與時效性，並發揚成舍我先生辦報的理念。

「北平世界日報內容數位化開發計畫」於1998年進行作業測試，1999-2000年進行初步全文數位化測試。於2002年被列入數位典藏「內容發展分項計畫」，隸屬新聞主題小組(數位典藏國家型科技計畫：新聞主題小組簡介)，並於2004年申請「後設資料動態管理系統研發計畫」研擬設計與建置民初新聞後設資料(Metadata)儲存、檢索與管理系統。(余顯強，2004年，7月)

以下分別針對「世界日報」的史料價值、計畫實施方式、成果與應用三個部分，對「世界日報」數位化相關計畫做一概括性的介紹。

### 一、「世界日報」的史料價值

「世界日報」於1925年2月10日創刊，1937年對日抗戰開始，於是被迫停刊。直到1945年8月抗戰勝利才復刊，復刊三年半後，1949年2月「世界日報」被共產黨查封，成舍我先生也跟隨國民政府撤退來台為止，總共發行16年，其間歷經軍閥割據、國民革命軍北伐、對日抗戰、國共戰爭，涵蓋軍政、訓政與憲政三個時期。(馬之驢，1986年，10月，頁166)

成舍我先生辦報堅持「不受津貼」、「不畏強暴」，「不懼軍閥」的暴行恐嚇，甚至曾經因此得罪軍閥，差點惹來殺身之禍。如此立場堅定、言論公正的辦報理念深獲北京人民認同，發行數字甚至高達3萬份，超越當時北京城其他各大報的紀錄。(馬之驢，1986年，10月，頁189)「世界日報」新聞內

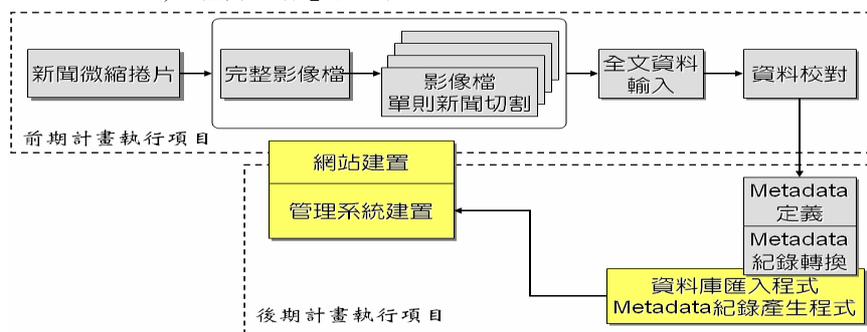
容除了具有公正客觀反應民初當時真切的社會狀況外，也有幾項鮮明的特色足以為新聞史學研究者注意：

- (一)由於當時教育界正熱烈討論教育制度的改革，加上老師罷教、學生罷課等學潮運動活躍，因此「世界日報」特別重視教育新聞，還特別闢立「教育界」專欄。(馬之驢，1986年，10月，頁178)
- (二)特別規劃「婦女界」專欄，旨在積極推動婦女運動，此為「世界日報」的特有版面。(金秉英，1998年，8月，頁248)
- (三)副刊名為「明珠」，主編張恨水先生自撰「金粉世家」、「啼笑姻緣」等，均獲得廣大迴響，是「世界日報」特色之一。

「世界日報」在動盪的政治環境中，仍堅持一貫公正客觀的理想，報導許多其他報社不敢揭露的新聞事實，頗具歷史研究價值。而該報特別重視教育新聞、婦女新聞及副刊，深具時代意義。

### 二、計畫實施方式

由於必須回溯收集完整發行的世界日報，礙於人力與時程的急迫性，計畫執行的程序主要分成如圖一所示5個工作階段，包括：由國內外各圖書館蒐集微縮捲片資料，然後轉製成紙本的新聞內文，利用人工進行新聞資料輸入，並且經過資料校對，最後設置網站呈現數位化成果。



圖一：典藏數位化工作流程圖

由於民初用字與今日習慣的差異、排版與印刷技術的限制，再加上保存本身的損害，以及報紙內容缺字、漏字、難字、不常用字等問題，經過廠商使用光學字元辨識(Optical character recognition, OCR) (Allen & Schalow, n.d.)正確辨識率低於50%(世新大學北平世界日報內容數位化開發計畫工作流程簡介)，因此並不適用OCR自動辨識文字，而必需採用人工打字輸入。

計畫初期，由於微縮資料閱讀不易，因此將一版的微縮捲片分割成9份，以A4紙張列印出來，然後再將九張A4紙拼成一個完整的版面。當時工作不僅費時(完成5個版面的時間約1小時)，而且印刷品質也不盡理想。後來放棄這種耗時費力的方式，直接將微縮捲片資料掃描成整版的數位圖檔，一個圖檔就是一個版面的新聞，再將整版新聞圖檔切割成單則新聞圖檔，並且以流水號為圖檔命名。打字員直接依據單則新聞數位圖檔進行文字輸入，省略紙本的製作時間，而且報紙影像的清晰度也隨之提升。

由於人力與時程的限制，計畫初期並無先分析資料著錄的描述範圍，而未定義符合「世界日報」的Metadata。因此其計畫後期則將上述編製的全文資料，配合轉換成符合自訂的Metadata與新聞小組聯合目錄所定義之Metadata規範的資料。

### 三、成果與應用

「北平世界日報內容數位化開發計畫」目前僅處理「世界日報」的新聞版面的內容，包括：國內要聞、世界要聞、各省新聞、經濟界、教育界、世界瑣聞、地方新聞；其他版面因經費及人力問題暫不收錄。由於民初報紙廣告、副刊、周刊的研究價值和趣味性很高，因此未來將陸續處理這些版面，以呈現「世界日報」的完整風貌。

為達到資源存取與利用的目的，「後設資料動態管理系統研發計畫」負責導入相關的技術與應

用，並規劃試用後設資料與開發適當的管理系統。初期已完成規劃的工作包括：符合歷史性新聞報紙特性Metadata的定義、動態資料庫管理架構、商務平台與經營模式的設計等。

但實際執行上仍舊缺乏適當的軟體開發人力。因此，現僅完成將全文資料匯入資料庫並依據參數定義轉換成指定Metadata之XML格式資料。目前可供檢索的資料為已完成建檔之1926至1929年的新聞內容，共計48547筆全文資料，包括文字模式與符合聯合目錄新聞主題後設資料的XML格式。

### 參、民初歷史性報紙執行之狀況

以「北平世界日報內容數位化開發計畫」為例，分別就「實體原件品質」、「Metadata制訂」與「Metadata著錄」三方面，探討民初歷史性報紙的特殊性對數位化過程造成的影響。

#### 一、實體原件品質

本計畫是將「世界日報」微縮捲片直接掃描成TIFF格式的數位圖檔，然後再以人工輸入成文字檔。數位影像的處理所遭遇的狀況包括歷史性報紙固有的問題和微縮捲片轉換數位圖檔，茲分述如下：

##### (一)歷史性報紙固有的問題

除了紙本原件印刷本身清晰度之外，報紙的邏輯結構(Logical structure)和實體輸出樣式(Physical layout)相當多樣且複雜，並不適合OCR處理報紙的文字自動辨識。此外「世界日報」還具有許多特殊的版面與內容結構問題。首先，歷史性報紙排版方式與今日不同。舉例來說，現代新聞報紙編輯電腦化，傾向於塊狀排版，也就是將每則新聞加上標題，排列成方形，然後像堆積木一般，在整個報紙版面上排滿新聞及圖片，如此一來版面整整齊齊、

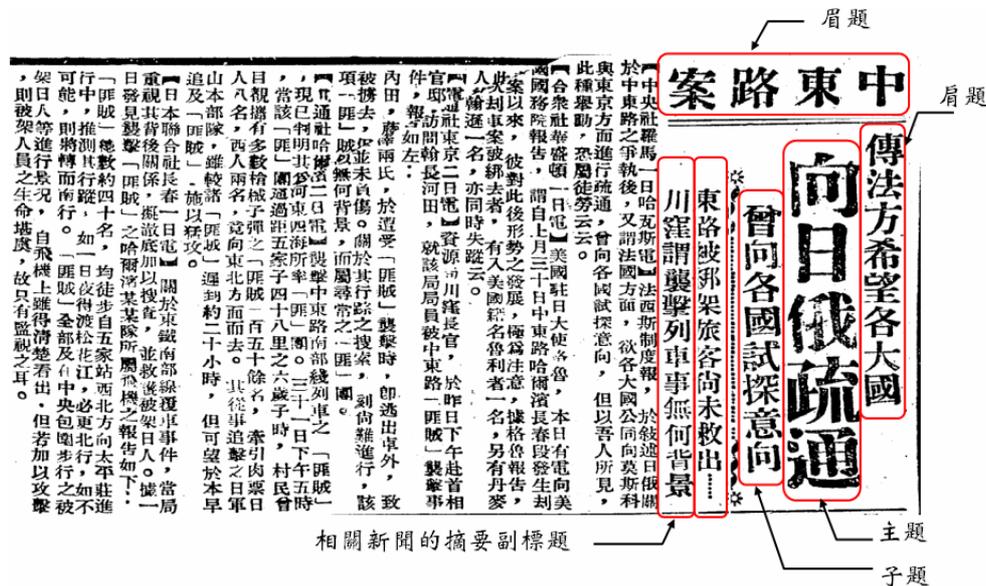
乾淨俐落，方便讀者分辨新聞編排的範圍，容易閱讀。在正式進行OCR時，可以方便切割(Zoning)每則新聞的區域，提高辨識精確率。而歷史性報紙則是採用手工排版、鉛字印刷。有些重要新聞，「世界日報」採用圖二所示的「集合式」報導方式，將多個相關類型的新聞事件共用同一主標題，再以摘要形式的副標題個別予以區隔。排版的方式常有非塊狀的方式，一則新聞有可能被分列成多邊形，影像切割時容易造成OCR辨識引擎的誤判。此外可能因為當時排版技術不成熟，或新聞內容篇幅控制不當，因此常常會有「文轉幾版」的情形產生必須倚賴人工判斷。

此外當時報紙印刷採用鉛字排版，可能因為鉛字不敷使用，亦有可能是以摘要方式將內文重點如圖三所示，以較大字體印刷。造成同一篇新聞稿中可能會發現某幾個字字級特別大或特別小，字與字之間間距又不固定，此種類型資料多見於軍事與

政治相關之新聞，且越後期出刊的報紙，出現比率越低。

(二) 微縮捲片轉換數位圖檔的問題

報紙經過長時間存放，以及使用者日積月累的翻閱，紙張勢必會變黃、變脆，因此容易產生折痕、破損、髒污，而不如原先新發行時的清晰。因此轉製成微縮捲片的過程中，會產生第一次影像品質的降低，再將微縮捲片掃描成TIFF數位圖片格式，由於技術上的限制，會造成第二次對影像解析度的損害。所以，經過實體報紙轉製成微縮捲片、微縮捲片掃描成數位圖檔兩次轉製的過程之後，實際得到的數位圖檔會有許多雜點，也會有整片的髒污、破損，甚至是影像扭曲的情形，影響數位化影像呈現的效果。



圖二：集合式的相關新聞報導模式

資料來源：世界日報，1924年9月3日，3版



文內新聞字體大小不一 非塊狀排版方式

圖三：內容報導模式

資料來源：世界日報，1928年1月5日，3版

綜合上述歷史性報紙原本排版上的特殊性，以及數位化過程中無可避免影像品質低落的原因，導致「世界日報」光學字元辨識的正確率非常低，因此無法採用OCR自動辨識模式，而以人工打字方式建立新聞內容全文。

## 二、Metadata制訂

Metadata是對數位物件的一組描述性資料，一般而言其描述的資訊可以分成兩部分：管理資訊與內容資訊，前者包含數位物件形式、版權宣告、傳遞資訊、製造資訊……等，後者則是描述數位物件的內容本身。Metadata不僅可用來處理資料的調度、儲存和檢索(陳百齡，2002年)，也能作為光學字元辨識(OCR)、語言處理(Linguistic processing)和資料找尋(Search)的指引。(世新大學北平世界日報內容數位化開發計畫工作流程簡介)基於北平

「世界日報」的媒體類型與內容特性，必須選擇適合用以描述新聞資料的標準，以利於完整的呈現北平「世界日報」的內容，以及新聞事件之間的關聯性。目前已發展與新聞描述有關的Metadata，包含由國際出版通訊協會(International Press Telecommunication Council, IPTC)所制訂之NewsML、國際新聞通訊會議制定的新聞業文字規格(News Industry Text Format, NITF)、專門描述運動新聞的SportsML、國際數位企業聯盟(International Digital Enterprise Alliance, IDEAlliance)制訂之產業出版需求標準(Publishing Requirements for Industry Standard Metadata, PRISM)，以及XMLNews組織所制訂之XMLNews等。



圖四：中央社南京電文全文

資料來源：世界日報，1931年2月15日，3版



圖五：法庭旁聽記

資料來源：世界日報，1931年2月14日，7版

上述各Metadata都有一個共同的現象，就是比較偏重描述數位物件的管理資料，但是對新聞內容本身的描述相形之下較少，主要原因在於這些Metadata主要是應用於現代新聞產業，滿足跨平台間流通的問題，因此重點在於標記管理資料，例如版權、資料格式、傳遞方式、出版時間等，而沒有迫切要去描述內容本身。(Cabo & Llavori, 1997, September, pp.645-661) 新聞報導的主題涵蓋範圍廣，內容的呈現並無特定結構，很難明確界定新聞稿所記敘的內容，因此與各類學術、檔案、公文比較起來，設計涵蓋完整新聞應用領域的Metadata時，有其一定的複雜度。

對新聞內容的描述，目前有兩個代表性的類型。一種是「由上而下」，也就是先從新聞文體結構的相關理論，取出一套描述架構，藉此發展Metadata的設計；一種則是「由下而上」，也就是不預設新聞文體的結構，直接從不同領域的新聞稿中，去瞭解新聞元素。或許比較好的方式是選擇折衷，混合「由上而下」及「由下而上」兩個觀點，以分析這個非常基本、卻瞭解不足的領域。(陳百

齡，2002年)

「世界日報」屬於民初的新聞報紙，不論是在版面的編排、標題的標示方法、新聞稿的類型、內文，除了包含新聞事件的人事時地物之外，尚有以不同字體呈現的摘要描述的撰寫格式等，與現代新聞學的規範大異其趣，甚至對新聞的定義，均有異於今日的報紙。茲舉例說明如下：

### (一)民初報紙有許多特殊的新聞文體

現代新聞稿要求專業的寫作方法，因此發展出倒金字塔的寫法，也就是將整個新聞事件的重點寫在導言中，包括人、事、時、地、物等5W1H，導言之後再分段詳細敘述。首先呈現出的導言最為重要，之後各段重要性遞減的倒金字塔文體結構。這樣的專業寫法其實已有一定的結構，但即使如此，如上所述目前各Metadata標準對新聞內容結構，仍然無法完全掌握。

民初報紙著重在資訊告知的功能，對新聞稿寫作方式並無特別要求，「世界日報」的新聞稿格式與類型更是五花八門，可以直接將政府單位、通訊

社或駐外記者的電報如圖四全文刊載上報；也可以將法庭上法官與被告的對話直接如圖五紀錄下來變成一篇新聞；更有按照事件發生的先後順序敘述，彷彿以小說故事的方式來寫作。因此現有的Metadata對新聞稿類型的描述，除了基本的專欄、社論等類型之外，實不足以完全描述「世界日報」。

### (二)民初報紙新聞稿並不署名

現代報紙每則新聞的最前面，一定會標示記者姓名與報導地點之類的署名，也是記者對新聞報導負責的一種表現。但是民初報紙並無此傳統，以「世界日報」為例，只有註明新聞來源、來稿地點及來稿形式。但「世界日報」的副刊、周刊文章，還是有作者署名，因此未來在處理副刊、周刊的資料時，仍需要著錄「作者」一欄。

目前世界日報「後設資料動態管理系統研發計畫」，參考國內外現行新聞Metadata標準，如NITF、XML News、PRISM、RSS等，以Dublin Core及NewsML定義之元素結構為主，並且考量「世界日報」的專有特色，發展出一套適合描述歷史新聞的Metadata，並依據發展之Metadata設計儲存、檢索與管理新聞資料的資訊系統。

### 三、Metadata著錄

Metadata制訂完成之後，接著執行Metadata著錄。將相關資料填入Metadata欄位中，並連結相關圖檔，封裝成爲數位物件。「世界日報」的Metadata著錄，除了提供一套著錄標準流程與規範之外，著錄者還需具備歷史的知識背景與新聞的知識背景。

「世界日報」是民初的報紙，報導的新聞主題以及使用的詞彙、語法、標點符號都和現代大不相同，因此如果著錄者沒有相當的歷史知識背景，在著錄資料時很難得心應手。例如，1931年2月12日第3版有一則簡報內容爲：「東北邊署參議員富占魁、昨日由瀋陽來平、林家訓昨假擷英番菜館、爲富氏洗塵、林邀于學中等作陪云……」，如果著錄者不熟悉電報的文法和當時的慣用詞彙，那麼很難從這段文字擷取重點，著錄在適當的欄位中。

除了歷史知識背景之外，著錄者還需要對新聞基礎知識有所瞭解。現代報紙標題形式清楚，很容易分辨出何者爲主題、何者爲子題，但「世界日報」當時的標題用法，很容易讓著錄者混淆。如圖六所示的標題形式，依照一般人的邏輯，因爲主題應該是字級最大、最醒目的「財力如何？」。但是如果著錄者對新聞編輯學有所認識，就會知道在



圖六：標題

資料來源：世界日報，1931年2月12日，3版

主題之上還可以有眉題、肩題等如圖二所示的副標題，用以輔助主題的語意。因此著錄者就應該瞭解這個標題是：「肩題：財力如何？」、「主題：張學良對晉軍改編變通辦法當以財力為準」、「子題：宋哲元商震等已通電就職」，而將資訊著錄至適當的欄位。

#### 四、搜尋結果的呈現

「世界日報」具有相當豐富的史料價值，可提供歷史研究者查找歷史資料的管道來源。然而在搜尋結果的呈現上，如果僅依照現代大多數報紙資料庫的方式，將無法滿足歷史研究者的需求。

舉例來說，現代報紙對標題最基本的要求就是要清楚表明新聞重點，從標題當中，就能完整表達該則新聞的內涵，所以資料庫在回傳搜尋結果時，可以僅顯示主題、子題，就能讓使用者尋得需要的資料。

但是「世界日報」的標題卻與現代報紙的標題規範不同，當時並不要求標題表達整個新聞的重點，雖然大部分的標題都能夠表達新聞重點，但仍有些標題可能與內文無關，或是無法讓讀者確實瞭解新聞內容，例如：「西使即來平」（1931年2月12日，3版）、「前日省府例會」（1931年2月12日，5版）等，都讓讀者無法從標題直接判斷內容。

此外，由於「世界日報」會將電報內容直接刊載，因此有些內文極短的電報，可能因為重要性並不高，所以集中刊登於一個區塊，標題即為「簡報」或「國府命令」，但其實是由許多簡短電報所組成。

所以，在查詢結果的呈現上，不能僅讓使用者由標題去猜測新聞的內容，理想狀態必須同時以新聞內容摘要的方式描述該則新聞，但由於摘要的撰寫必須具備更高的歷史知識背景，本計畫現有人力資源無法配合，有賴未來與歷史研究者的跨領域合作。目前折衷的方式應是提供新聞的導言，讓使用者從導言判斷內容是否符合其需求。

#### 肆、相關議題探討與建議

民初歷史性報紙除了在數位化過程中，需特殊處理的程序外，從數位化過程的實務經驗中，可歸納出下列幾項議題：

##### 一、全版影像的提供

Deegan 等人認為歷史性報紙 (Historical newspaper) 數位化，必須提供使用者兩個主要的功能：(Deegan, King, Men, & Steinvel, 2001, July)

##### (一)可讀性(Readability)

指讀者可以閱讀與掌握歷史本文。建議讓讀者直接觀看到最真實的原件，也就是全版圖檔，但考慮到全版圖檔篇幅太大、檔案容量也大，會造成網路頻寬的負擔，因此僅提供切割後的單篇新聞圖檔。

##### (二)可找性(Searchability)

指使用者能夠檢索新聞，並且瞭解新聞與新聞之間的關連性，而這是圖檔無法達到的效果，必須由圖檔轉成文字檔，並加以適當描述，甚至配合權威控制，釐清數位物件之間的關係，以確保使用者能夠找全與找準。

國際上一些大型的歷史性報紙數位化計畫，也傾向提供原始報紙的數位影像檔，像紐西蘭的 NIUPEPA 計畫 (Apperley, Cunningham, Keegan & Witten, 2001, May, pp.86-87)、布魯克林公共圖書館 (Brooklyn Public Library) 的 Brooklyn Daily Eagle 計畫 (Brooklyn Public Library, 2004)、猶他報紙數位化計畫 (Utah Digital Newspapers, UDN) (The University of Utah, 2003)，也都為了維持原始報紙的格式與完整性，將原始報紙的數位影像提供使用者閱覽。

對歷史研究者而言，全版影像的提供也非常重要。因為歷史研究必須追求第一手的史料，數位檔案資料庫的查詢只能算是一種輔助，在數位化的過程中，隨時有可能因為人為疏失，而產生資料錯誤、資料遺漏等問題，如果歷史研究者僅搜尋數位文字檔，那麼很有可能造成錯誤的引用，因此如果能夠直接看到原始新聞事件的數位影像，將能夠增加資料的可信度。

## 二、廣告、副刊、社論的數位化價值

每一份新聞報紙都是對當時社會的紀錄，「世界日報」的貢獻即在於以客觀的角度反映普遍民意。除此之外，「世界日報」除了陳述民初政經、社會的新聞事件之外，還有由張恨水先生主編的副刊。對所有的歷史研究者來說，廣告、副刊和社論都極具研究價值。廣告表現當時的風土民情；副刊展示藝文界的發展；社論則可窺知民意的意向，因此除了新聞報導之外，報紙的廣告、副刊和社論也是歷史新聞報紙數位化的重點。

## 三、數位化可能造成的資訊遺失

數位化能夠延續歷史性報紙的生命，讓讀者繼續以數位形式瀏覽、閱讀、加值與應用。但目前的數位典藏過程幾乎是「由上而下」的，也就是先由計畫執行機構制定一套數位典藏的方針，再根據這一套方針去設計數位化的方法與步驟。但是在這樣的過程中，如果一開始的方向不夠縝密，後續的問題將會接踵產生。

例如現代對開報紙兩個版面之間的中縫會夾雜報社聯絡方式、廣告等資訊，「世界日報」亦有中縫的設計，裡面印有許多廣告。計畫執行時，由報紙的微縮捲片轉為單版圖檔，版面之間的中縫便被裁切而遺漏掉，因此無法在數位化時加以處理與描述。同樣的問題亦存在於Metadata的定義。由於

計畫執行初期對民初報紙版面、體裁、風格的認識有限，造成制定的欄位屬性不敷所用。因此，進行歷史性報紙的數位化之前，對該報的瞭解相當重要，在未詳加考慮的情形下，即可能僅收納已知的資訊類型，而排除許多有用資訊。

## 伍、結論

歷史性報紙隱含某個時代豐富的歷史資訊，對歷史研究和新聞學研究有其重要性，而且因為實體報紙品質損害嚴重、某些歷史性報紙發行報社已不存在或沒有專職的單位負責進行管理，為確保社會共同文化資產的延續，所以急需投入技術、經費、人力，將歷史性報紙數位化，以數位形式延續其生命、增加其利用價值。

其次，歷史性報紙在新聞文體、署名方式上都與現代報紙不同，無法適用專為現代報紙描述工作而設計的Metadata標準，因此必須考量「世界日報」與現代報紙的諸多差異，制訂出專屬於中文歷史性報紙描述的Metadata標準，描述的內涵不僅止於新聞事件的分析，亦須包含內容相關之詮釋。標準制訂後，「世界日報」新聞寫作含有許多與今日不同的語法及背景知識，若沒有一定程度的瞭解，即使有一套統一的著錄規則，著錄人員也無從下手，因此對著錄的規範與訓練有賴與歷史學門、新聞學門的跨領域合作。許多數位典藏機構在執行數位化的過程，亦有相同的需求，必須倚賴學者專家協助資料的辨識與著錄方能完成，此與圖書館編目作業遵循編目規則即能著錄書目的模式有所不同。因此如何加速與強化歷史性新聞資料的著錄工作，仍有待與歷史學者及新聞專家的共同努力，方能對新聞資訊做正確與完整的詮釋。

(收稿日期：94年3月8日)

(本研究係數位典藏國家型科技計畫—技術研發分項公開徵選計畫「後設資料動態管理系統研發計畫」,計畫編號: NSC 94-2422-H-128 -001 之摘錄,承蒙國科會提供補助經費與相關研究所需之協助。計劃執行過程中並獲得內容發展分項計畫「北平世界日報內容數位化開發計畫」相關之資源與數位化經驗之指導協助,謹致謝忱。)

## 參考書目：

- 世新大學北平世界日報內容數位化開發計畫工作流程簡介。上網日期：2005年2月15日。網址：[http://content.ndap.org.tw/main/doc\\_detail.php?doc\\_id=242&class\\_vision=16](http://content.ndap.org.tw/main/doc_detail.php?doc_id=242&class_vision=16)
- 李學軍、鄧小川(2001年)。報紙文獻的數字化建設。四川圖書館學報, 119, 69。
- 何銘傑(2001年, 5月)。一網看盡五十年—聯合知識庫的建置與應用。全國新書資訊月刊, 2001年5月, 11。
- 余顯強(2004年, 7月)。後設資料動態管理系統研發計畫簡介。電子通訊, 3(7), 上網日期：2005年2月14日。網址：[http://www.ndap.org.tw/1\\_newsletter/content.php?uid=1028](http://www.ndap.org.tw/1_newsletter/content.php?uid=1028)
- 林淑芬、宋美珍(2002年, 6月)。期刊報紙數位化問題淺探。國家圖書館館刊, 91年(1), 198。
- 金秉英(1998年, 8月)。當代論才子, 如公復幾人—憶成舍我先生。在報海生涯：成舍我百年誕辰紀念文集(頁248)。北京。
- 姚雪痕(2003年, 5月)。網絡報紙是否會取代傳統報紙。上網日期：2005年2月20日。網址：<http://www.cgan.net/science/discussion/htm/epaper.htm>
- 馬之驢(1986年, 10月)。新聞界三老兵。台北市：經世書局, 頁166, 178, 189。
- 陳百齡(2002年)。讓電腦也能辨識新聞：新聞內容標記問題初探。中華傳播學會2002年會。上網日期：2005年2月15日。網址：<http://ccs.nccu.edu.tw/oldccs/con2002/conworks/11B-1.doc>
- 張琪玉(1999年)。報紙文獻是一種極為豐富而未被充分開發的信息源。圖書館雜誌, 2, 7。
- 鄭植榮(2004年, 3月)。報紙：數位時代的永恆太陽。在數位媒體的發展與再造研討會論文集。上網日期：2005年2月21日。網址：<http://mol.mcu.edu.tw/~ebook/research/pdf/paper/01.pdf>
- 數位典藏國家型科技計劃：新聞主題小組簡介。上網日期：2005年2月14日。網址：[http://content.ndap.org.tw/main/vision\\_brief.php?class\\_vision=16](http://content.ndap.org.tw/main/vision_brief.php?class_vision=16)
- Allen, Robert B. & Schalow, John. Metadata and data structures for the historical newspaper digital library. In Proceedings of the Eighth International Conference on Information and Knowledge Management. Retrieved February 15, 2005, from: <http://portal.acm.org/citation.cfm?id=319971&dl=ACM&coll=GUIDE>
- Apperley, Mark, Cunningham, Sally Jo, Keegan, Te Taka & Witten, Ian H. (2001, May). NIUPEPA: a historical newspaper collection. Communications of the ACM Archive, 44(5), 86-87.
- Brooklyn Public Library (2004). Historic newspaper: Brooklyn Daily Eagle online (1841-1902). Retrieved February 18, 2005, from <http://www.brooklynpubliclibrary.org/eagle/>

Cabo, María José Aramburu & Llavori, Rafael Berlanga (1997, September). An approach to a digital library of newspapers. Information Processing and Management, 33(5),. 645-661.

Deegan, Marilyn, King, Edmund, Men, Alon & Steinvel, Emil (2001, July). The British Library newspaper pilot. Digitization and Preservation Online Resource Center. Retrieved February 23, 2005, from: <http://digitalcooperative.oclc.org/hdigitize/BritishLibraryNewspaper.html>

The University of Utah (2003). Utah Digital Newspapers. Retrieved March 1, 2005, from: <http://www.lib.utah.edu/digital/unews/about.html>