

國家圖書館新聞影音資料數位化典藏現況

專文

國家圖書館新聞影音資料數位化典藏現況¹

■宋美珍

一、前言

今日的新聞就是明日的歷史 (Today's News, Tomorrow's History)，每一則新聞就像述說著一個真實故事，透過以文字、聲音及影像鏡頭講述抽象的意識形態衝突、自由主義、恐怖主義、宗教主義、社會事件、經濟發展、自然生態、八卦等，這些片段的新聞，日積月累就形成了我們社會發展的重要歷史紀錄。

新聞依其傳播來源可分為平面新聞與影音新聞，前者最典型的代表就是報紙、雜誌，而後者則包括廣播新聞、電視新聞、新聞資料片與新聞紀錄片等。在電視時代來臨之前，看「報紙」與聽「廣播」是大多數人的新聞消息來源，然而在電視的普及與電子技術的進步推動之下，電視挾其「視」、「聽」雙管齊下的傳播性質，打破閱聽人教育程度與識字能力的藩籬，使得近半世紀以來，電視新聞已經成為現代人類重要的資訊來源之一。

對於這些記錄著人類歷史活動的新聞資料的保存與整理工作，無論在國內或國外都有相關的單位積極的進行中，其中又以平面新聞資料的整理與蒐集最為廣泛，包括書目控制、典藏、剪報、標題索引製作，乃至於製作新聞標題資料庫提供使用者查詢利用，或拍攝微縮進而以數位化掃描等方式進行典藏，無論報刊的發行單位或是典藏的圖書館，對於以上資料的整理與蒐集方式，都已具有相當的經驗；例如目前在國內由行政院國科會所推動的數位典藏國家型科技計畫與行政院文建會所推動的「國家文化資料庫」計畫，就有數個與報紙數位化典藏有關的計畫正分別由國家圖書館（以下簡稱「國圖」）、國立臺中圖書館及世新大學圖書館執行中，其他如聯合報、中國時報、中央日報等報社亦陸續

完成回溯性資料的數位化工程。

至於影音新聞方面，由於這類型的資料是透過有線電或無線電方式進行播送，並未對外發行出版，而原始資料又多以各種不同的錄影（音）媒體母帶存放於原播放單位，如電視臺及廣播電臺；電視臺與廣播電臺對於這些資料各有一套專屬的管理系統，以備所屬企業的新聞製播部門人員參考相關新聞畫面，並提供新聞畫面的剪輯與新聞專題或節目製作利用。在國內圖書館方面，除擁有各種特殊的播放設備的專門圖書館如電影圖書館，收藏有1946年至1985年間「臺影電影製片廠」及「中國製片廠」所拍攝的16mm、35mm及betacam新聞紀錄片之外，大部分的圖書館並無相關收藏，而有系統的進行整理或典藏各種新聞影音資料者亦寥寥可數²。反觀國外，如美國國會圖書館、大英圖書館及法國國家圖書館自21世紀初以來，即設有專責部門有系統的收藏該國的錄音錄影資料，其中也包括電視臺及廣播電臺的錄影錄音資料。其中在新聞影音資料的典藏方面，最具盛名者當屬美國范德彼特大學（Vanderbilt University）的「電視新聞典藏」（Television News Archive），該校自1968年8月5日以來即對美國全國性廣播頻道4大電視網如：ABC、CBS、NBC、CNN的晚間新聞（Evening News）及1989年以來ABC電視臺的夜線（Nightline）新聞節目進行側錄保存典藏，並將新聞內容製作索引、摘要，建立新聞資料庫，除了可透過網際網路免費檢索新聞內容資料庫，並提供透過圖書館館際借閱（Library Loan）的新聞錄影帶複製借閱服務，以便學術研究與教學利用。目前該資料庫中所整理的新聞條目達705,000筆，另外還包括8,000筆新聞特別報導。在媒體格式方式，Vanderbilt大學所典藏的媒

體係以betacam錄影帶為主，尚未進行數位化。其他如Texas Tech University 也收藏CNN World Report 自1987年以來的錄影資料³。

二、臺灣新聞影音資料數位化典藏計畫緣起

「國圖」設有視聽資料閱覽室，館藏視聽資料約有2萬餘件，其中大部分為各類型視聽節目帶與音樂唱片，次為各類型館藏的微縮資料，其中視聽資料帶中有部分Beta帶，已經面臨無播放機器可使用的窘境，而部分的VHS節目帶，也因長期使用，以致有畫面品質不良的情況。「國圖」自1996年起即積極進行各種館藏資料如期刊、報紙、古籍善本、博碩士論文等資料數位化的工作，同時也是參與數位典藏國家型科技計畫的9個機構計畫之一，故無論是在數位化的工作流程或技術處理方面都累積了相當的成果與經驗，因此自2001年初起，「國圖」即開始著手規劃視聽資料的數位化工作。由於視聽資料的重製，有較一般圖書資料更複雜的著作權問題，故初始規劃係以館藏Beta帶及官方出版但已絕版，且原件典藏狀況不佳的小量視聽節目與音樂帶，作為數位化的內容。

2001年5月「國圖」獲得一批來自文從道捐贈的拍攝帶，該批資料內容包括文先生過去從事傳播事業時，所受各界委託拍攝的各類資料母帶，經初步檢視資料帶所附的筆記抄本，發現內容大約包括中視早安新聞、放眼望天下等節目母帶以及大量的新聞拍攝帶，數量約2,000餘捲，由於該批資料的儲存媒體複雜，包括Betacam、1/2吋帶、1吋帶等，且新聞內容多為無聲無旁白的拍攝帶，同一捲母帶中有包含數十則的新聞拍攝資料，在考量「國圖」

並無可供播放的專業設備，而該批資料又頗為珍貴，除急需進行數位化轉製至其他儲存媒體供典藏之外，影帶內容亦需要進行妥善的分析整理，才得以一究其內容。由於「國圖」工作人員對於影音資料的數位化尚缺乏相關經驗，為了解新聞拍攝帶的處理與影音資料數位化的作業流程，乃由閱覽組主任陳昭珍率相關同人於同年10月初赴臺灣電視公司參觀該公司的資料片庫。這次的參訪過程，除瞭解我國第一家開播的電視臺如何典藏與管理所有的節目帶（約5萬捲）與新聞資料帶（約3萬捲）外，並對目前國內電視臺數位化的進程有了初步的認識，最大的收穫則是雙方就是否可能進行電視資料的數位化合作開啓了對話。

在參訪前一個月（2001年9月18日）臺灣北部遭受了近年來最嚴重的水災，在納莉颱風的大雨侵襲下，各企業受創嚴重，其中包括若干電視臺多年來的影帶也付諸積水之中，損失甚劇；臺灣電視公司亦憂心該公司珍貴的影音史料可能有一天亦遭無情水火的災害，因此也正擬以數位化進行保存及利用，故在對談中，雙方對於電視臺所保存的大量新聞資料應該如何進行整理、數位化與提供大眾利用有相當深入的討論，而基於資料共建共享的原則，初步也達成未來可以該電視臺所製播的電視新聞資料帶進行數位化典藏的共識。

然而經「國圖」評估此項新聞資料的數位化典藏作業，所需經費龐大，除「國圖」自有預算外尚需額外的經費補助始能進行，乃於同年11月草擬「臺灣影音資料數位化典藏暨數位媒體隨選視訊服務計畫書」向行政院文建會申請經費補助並於11月30日獲得行政院文建會來函同意補助，該計畫執行數位化典藏的內容，除了前述館藏文從道捐贈的新

聞拍攝帶之外，亦擴大至臺灣新聞影音紀錄的數位典藏，計畫時程從2001年11月至2002年11月止，為期1年。

三、計畫內容

「國圖」在獲得行政院文建會的經費補助後，即積極著手與國內電視臺展開合作標的物洽談，電視臺所製播的新聞資料大致區分為：電視新聞、新聞節目（含談話性、評論性節目）、新聞專輯帶、新聞紀錄片等，在考量經費、時程的限制及授權使用範圍的協商空間和未來使用需求，初步決定以電視臺的晚間新聞及新聞節目帶作為該計畫數位典藏的重點，在參訪幾家較具歷史的電視臺後，並獲得數位化檔案永久典藏權及公共資訊永久使用授權的保證承諾後，決定以臺灣最早的電視臺——臺灣電視公司，也是擁有臺灣最多最完整的晚間新聞資料帶的電視臺，以及開播最早的新聞性節目「華視新聞雜誌」，作為該計畫新聞影音資料數位化典藏內容，以下就計畫執行的內容分別簡述如下：

（一）臺視晚間新聞

包括臺灣電視公司自1962年10月1日開播以來至2001年共40年的每日晚間新聞數位檔案與新聞詮釋資料。其中高解析度的數位檔案（Mpeg2 6M）由「國圖」及行政院文建會進行永久典藏，低解析的串流播放檔案（250Kbps、56Kbps）則分別授權於全國公共圖書館內部網路與網際網路做永久非營利播放使用。

由於電視臺對於將56kbps的串流檔案作為公共資訊一事有所疑慮，擔心恐影響其商業利益，故有關公共資訊權的利用，期間經過了近4個月的多次

討論協商才達成協議。至2002年4月24日該館依據政府採購法與臺灣電視公司完成簽約；臺灣電視公司至2002年11月以前將製作並提供「國圖」1962年至1971年、1986年至2001年每日晚間新聞的數位化檔案及相關詮釋資料。

（二）華視新聞雜誌

包括「華視新聞雜誌」節目從1971年至1991年間，該館所挑選的182個新聞單元數位檔案及所有新聞單元的詮釋資料。其中高解析度的數位檔案（Mpeg2 6M）由「國圖」及行政院文建會進行永久典藏，較低解析的串流播放檔案（250Kbps、56Kbps）則分別授權於全國公共圖書館內部網路與網際網路做永久非營利播放使用。

2002年6月25日「國圖」依據政府採購法與中華電視臺完成簽約；中華電視臺須於2002年11月以前將以上節目數位化檔案及相關詮釋資料提供「國圖」。

（三）館藏新聞拍攝帶與資料帶

原文從道所捐贈的資料帶，包含中國電視公司的「早安新聞」資料帶、「放眼望天下」的拍攝帶及其他新聞拍攝帶，前兩項由於在著作權歸屬方面似尚有問題無法釐清，故暫不進行數位化作業，而以文先生所拍攝的新聞拍攝帶約1,000餘捲，作為數位化的內容。

四、新聞影音資料數位化作業流程

由於影音資料的數位化設備投資費用甚鉅，且各電視臺在資料安全的考量下，對於所擁有的珍貴母帶，亦不同意在電視臺以外地點進行數位化重



製，故該計畫的數位檔案製作係由「國圖」直接委託資料所有權人亦即臺視及華視，進行資料數位化轉製工作，至於館藏新聞帶部分則是透過公開招標方式，委託廠商代為轉置，該項作業已先於2000年12月完成決標，由中華電視臺得標，負責該項資料的數位化作業。

雖然該計畫影音資料數位化的工作都委由廠商進行，然而「國圖」在各項資料委託發包之前，必須就資料數位化的各種檔案格式、製作規範，及新聞資料內容分析的詮釋資料（metadata）格式等訂定相關標準規範，以作為各電視臺在進行數位化檔案製作時作業規範與依據，這些標準亦同時作為該館招標規範書與驗收依據的文件。以下就該館在進行新聞影音資料數位化的各項前置作業與工作流程分別說明如下：

（一）數位檔案轉製單元

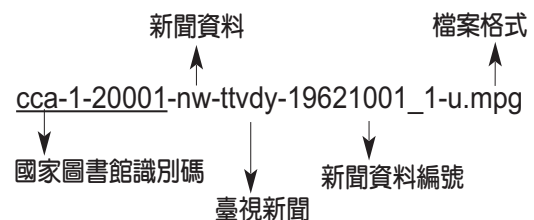
由於該計畫所進行的3種新聞資料各有特性，且原始儲存媒體也各有不同，故在數位化檔案轉製過程，施作方式或有差異，例如臺視晚間新聞係以新聞播出帶的每則（clip）新聞為數位化檔案產出單元，華視新聞雜誌亦以每集之中播出的不同單元，分別轉製數位檔案。至於館藏新聞拍攝帶則由於每捲母帶自成主題（例如橋樑、交通、國民學校……等），以每捲影帶為單元，產出1個數位檔案，故該項數位檔案較前兩項大。

（二）相關標準與規範

配合該計畫的內容，「國圖」參考「資料數位化與命名原則規範」⁴，訂有3項作業標準與規範，包括：新聞影音數位化檔案格式標準、新聞影

音數位檔案命名原則及新聞影音詮釋資料格式。

- 1.新聞影音數位化檔案格式標準：依據數位檔案的使用目的，新聞影音錄影資料所轉製的數位檔案格式分為典藏格式及服務格式。前者檔案格式為Mpeg2/6M，後者則區分為串流寬頻格式與窄頻格式，檔案格式分別為ASF 250Kbps及ASF 56Kbps；故每則新聞資料將產出3個不同格式的數位檔案。
- 2.新聞影音數位檔案命名原則：針對3種新聞資料特性並參考「資料數位化與命名原則規範」訂定本數位檔案命名原則表。檔案命名結構分為機關代碼、資料群組代碼、資料群組次級代碼、資料編號、檔案格式及附屬檔名，例如1962年10月1日第1則晚間新聞的典藏格式檔案命名為：



此命名原則的特點如下：

- （1）將資料數位化的過程與詮釋資料分析工作分開執行，節省作業時間。
- （2）依檔案格式代碼可用以判斷檔案格式與服務範圍，例如u為典藏格式，b為串流寬頻服務格式，n為串流窄頻服務格式。
- （3）此外依此檔名可辨識資料來源單位，並回溯找到數位物件。
- （4）未來若加入國際既有命名系統（如URN、DOI）時，能直接加入國家識別碼，而成為國際間唯一號碼。

3.新聞影音詮釋資料格式新聞資料內容的描述，多由5W1H六元素所組成，亦即人物（Who）、時間（When）、地點（Where）、事件（What）、原因（Why）及經過（How）。該館在編訂新聞資料詮釋資料格式以前，曾就臺視及華視內部的新聞資料庫進行比較分析，在考量新聞資料的特性，並參酌DC的架構後，訂定了此詮釋資料標準，為使電視臺在未來進行詮釋資料處理時，能直接轉出資料庫中已有的資料，再進行其他資料項目的補充，該館乃先就兩公司的資料庫格式與本詮釋資料格式，整理完成資料轉出欄位對照表（Crossing Table），提供兩家公司作為資料轉檔的參考工具。

（三）製作流程

該計畫的數位化作業分別於臺視及華視公司內部進行，各公司依據「國圖」提供的各種數位化作業規範與標準，配合公司內部資料管理作業，訂定數位化作業流程，其流程大致雷同。以下僅以臺灣電視公司處理臺視晚間新聞為例，說明各部分工作流程。

1.新聞影帶數位化檔案轉製

（1）檢查原始母帶品質：以臺視新聞資料帶為例，其原始母帶包括16mm film、3/4吋U-matic錄影帶及1/2吋Betacam錄影帶，其中早期的新聞帶（尤其是3/4吋 U-matic與16mm F帶），常有影像品質不良，如跳動、拉線或影像上下顛倒的情形，故在轉檔前需檢視影帶品質，若有影像品質不良者，須先進行影帶清洗或影像調整程序，方進行後續數位化作業。

（2）影帶時間碼擷取：由於晚間新聞播出帶係以天（日期）為單位，而「國圖」所要求的數位檔案係以每則新聞為單位，故必須先將類比錄影帶資料內容依每則新聞作區隔，登錄每則新聞段落時間碼的起點與終點，在登錄過程中，需將不相關資料濾除，例如每天新聞開播主播所說的歡迎詞、氣象預報、片尾工作人員一覽表字幕等特殊內容。本時間碼擷取完成後，即可以每則新聞為單位，建立影帶資料時間碼工作表，以達到自動依序影帶數位化編碼與轉檔的自動化作業。

（3）影帶數位化轉檔：依據前項完成的影帶時間碼擷取工作表，自動依序擷取每則新聞內容，同時配合「國圖」需求同時產出廣播級的高解析品質及網路傳輸的低解析品質的不同檔案格式。

2.詮釋資料分析

將完成的數位影音內容對照臺視原有新聞資料庫，並將對照表所列資料自動轉入詮釋資料庫中，工作人員需核對新聞影音內容與詮釋資料內容是否相符及正確性，若原件缺乏詮釋資料內容，需進行資料新增作業。

3.資料品質檢核

就數位化檔案以程式抽檢影音品質，若品質有瑕疵，須重新轉錄。

4.數位檔案光碟製作

就影像品質無誤的數位化檔案依「國圖」所列光碟燒錄規範，進行數位檔案光碟與詮釋資料光碟燒錄及產品包裝作業。

5.報表列印

依新聞日期依序列印詮釋資料報表。



五、計畫成果與應用服務

該計畫至2002年11月第一階段結束，產出的數位化成果及相關資料庫應用服務簡述如下：

(一) 計畫成果

1. 臺視晚間新聞

完成1962年至1971年，1987年至2001年，共計25年的臺視晚間新聞數位化檔案典藏，原始檔案總時數長度約為207,263分鐘，新聞內容詮釋資料共147,041條，產出數位檔案則有典藏級（Mpeg2）DVD-R光碟片2,300餘片資料量約8T、網路播放格式（56K, 250K）CD-R共500餘片，資料量約320G。

2. 華視新聞雜誌

完成1981年至1991年，共182個新聞單元資料的數位化檔案及詮釋資料，原始檔案總時數約為3,165分鐘，新聞單元詮釋資料共1,741條，產出數位檔案則有典藏級（Mpeg2）DVD-R光碟片45片資料量約180G、網路播放格式（56K, 250K）CD-R共12片資料量約7G。

3. 館藏新聞資料帶

完成館藏文從道捐贈新聞資料帶約1,057捲轉製為數位化檔案，總計產出40,980分鐘、155,220單元的數位化檔案及14,668條新聞拍攝內容詮釋資料，資料量約3T。

以上所完成的各項數位化檔案，依使用目的分別儲存於DVD-R光碟片及CD-R光碟片，其中Mpeg2典藏格式分別存放於「國圖」視聽片庫及行政院文建會國家文化資料庫磁碟陣列，250K及56K則依授權範圍分別上載「國圖」數位典藏資料庫及行政院文建會文化資料庫的磁碟陣列。

(二) 應用服務

目前該計畫所完成的成果包括新聞影音數位檔案及相關詮釋資料，並已陸續匯入「國圖」數位典藏系統中，並透過「國圖」「臺灣記憶」（Taiwan Memory）系統（圖9）及視訊隨選服務系統（VOD）服務網頁，提供使用者透過網路線上免費查詢相關新聞的詮釋資料，並可直接觀看不同播放速率的新聞影片內容（圖10至圖13）。



圖9 臺灣記憶系統首頁



圖10 新聞影音資料檢索畫面



圖11 新聞影音資料檢查查詢結果顯示畫面



圖12 新聞影音詮釋資料



圖13 新聞影音資料線上播放 (56K串流)

六、結語

該計畫未來將繼續於2003年底前將臺灣電視公司開播以來的電視新聞進行完整數位典藏，這些數位化的影音新聞將是40年來臺灣人民的生活、政治、經濟、外交、社會生活、科技發展等活動的全紀錄，除了可重現40年來的歷史記憶，提供研究人士利用外，亦希望以此為基礎，逐步有系統的完整收集並典藏本國重要的新聞影音資料。

回顧過去一年來計畫執行的過程，除體會到新聞影音資料的整理與典藏工作，須投入龐大的經費與人力之外，在執行過程中，更發現其他如：影音資料數位化格式標準、新聞詮釋資料標準、新聞主題詞語表的編製、人名權威控制、如何建立單一檢索平臺以及合作典藏機制等議題，除了圖書館界之外，尚須結合新聞媒體產官學界共同合作，方能共謀一完整的全國性新聞影音資料數位典藏策略。

後記

該計畫得以順利進行，除了感謝行政院文建會提供經費補助外，尚感謝臺灣電視公司前總經理胡元輝先生及前視聽處李慧慧處長的大力支持與協助促成本計畫。此外，承臺灣電視公司李榮林處長、林開南主任及中華電視臺黃日春處長、歐陽瑞主任、謝偉淦主任、周梅盛先生等在數位化技術方面的支援與配合，特此致謝。

【註釋】

1. 國家圖書館新聞影音數位化工作係為91年度「臺灣影音資料數位化典藏計畫」的工作成果之一，該計畫承行政院文建會提供經費資助。

2. 針對臺灣新聞片片目的書目整理，參見李道明、盧非易主編，《臺灣新聞 / 紀錄片片目》（臺北市：行政院文建會，民國88年）。
3. 有關 Vanderbilt University 的 TV News Archive 的詳細介紹參見 <http://tvnews.vanderbilt.edu/>。
4. 「資料數位化與命名原則規範」係數位圖書館標準規範之一，由教育部於民國91年12月27日發布。

【參考書目】

數位典藏國家型科技計畫內容分項計畫新聞主題小組。《中文新聞內容數位化研討會會議手冊》。臺北市：該組，民國92年3月。

林開南。《臺灣電視公司影帶資料數位化作業程序書》。臺北市：臺灣電視公司，民國91年。

Vanderbilt University. "Television News Archive."
"Vanderbilt University. <http://tvnews.vanderbilt.edu/>.

Hane, Paula J. "Vanderbilt Improves Television News Archive." *Information Today* 19:1 (October 2002) .