

教育部計畫成果典藏資訊網之規劃與建置

林巧敏

政治大學圖書資訊與檔案學研究所助理教授

E-mail : cmlin@nccu.edu.tw

摘要

由教育部顧問室委託建置之「人文社會科學相關領域中綱計畫成果資訊網」，係教育部顧問室過去補助各校發展之台灣人文社會科學相關領域中綱計畫執行成果之數位典藏網站，網站規劃蒐集整理教育部中綱計畫各項子計畫之成果報告、教材、非書資料、研討會議程與論文集、活動成果紀錄等，典藏內容除一般文獻型資料外，尚且包括計畫執行過程紀錄與各項影音成果。此一典藏資訊網之建置除了由計畫工作小組回溯數位化處理外，同時亦可由各計畫執行團隊自行典藏現行計畫執行過程相關紀錄，以完整保留計畫執行經驗與產出成果。本文將介紹此一數位典藏網站建置過程，說明系統規劃工作、典藏資料內容與特性、數位化原則與系統建置成果，本典藏資訊網不僅可供人文社會學者研究利用，亦有助於計畫執行者經驗交流。

關鍵字：數位典藏 (digital archive)；數位化 (digitization)；
數位化流程 (digital procedure)；數位計畫管理 (digital project management)

壹、前言

台灣地區國家政策為促進經濟發展、配合產業需求，長期以來人才培育偏重先導性之重點科技、醫學及工程科學相關領域的發展。但有鑑於人文內涵與思維係建全社會發展之重要基石，2005年教育部顧問室為厚實人文社會科學領域基礎及國家競爭力，整合各界意見，擬定「2005-2008教育施政主軸」，以「創意台灣、全球佈局：培育各盡其才新國民」為宗旨，自2007年起規劃推動「人文教育革新」、「全球化下的台灣文史藝術」、「新興議題與專業教育改革」及「通識教育」

等中程綱要計畫，2008 年加入「海洋教育先導型計畫」，在科技與人文的預算比例上進行相當幅度的調整，期使未來人文社會科學教育持續進步，莘莘學子在優質的教育環境中，涵泳人文素養，培養社會關懷情操，塑造全方位人才。(教育部顧問室，2009，頁 6)。

此一「人文社會科學相關領域中綱計畫」歷經三年的各項資源投注，教育部顧問室為考量能長期保存各中綱計畫之執行成果及日後計畫之延續，並期在未來能有更充足之應用，於 2009 年 12 月決議成立兩年期程之專案計畫，以建立專屬系統網站，並透過此系統網站之功能設計與網頁內容呈現，將現階段台灣人文社會科學領域各中綱計畫之豐碩成果，加以數位典藏並開放運用。

「人文社會科學相關領域中綱計畫成果資訊網」規劃以典藏應用為基本理念，收集整理台灣人文社會科學相關領域中綱計畫各項子計畫之成果報告、教材、論著、期刊、非書資料、研討會議程與論文集、活動成果紀錄等，將各種文獻、資料進行詮釋資料 (metadata) 建檔及內容數位化掃描，予以數位長久典藏，並提供讀者於網路上查詢、閱覽。本文將介紹此一數位典藏網站建置過程與經驗，期許透過系統建置經驗分享，瞭解相關領域學者專家之意見，俾於第二階段系統功能精進的參考。

貳、計畫工作事項

本計畫分兩年 (2010.03-2012.02) 進行，規劃之工作內容包括：資料整理與數位化、建置系統與網頁服務及建置成果宣導與推廣等三大事項。

一、資料整理與數位化

收集整理人文社會科學相關領域中綱計畫各項子計畫之成果報告、書籍、論著、期刊、非書資料及研討會論文集等，將各種文獻、資料進行詮釋資料 (metadata) 建檔及內容數位化掃描，予以數位永久典藏。

二、建置系統與網頁服務

將內容影像檔或全文電子檔加以彙整建置資料庫，完成「人文社會科學相關領域中綱計畫成果資訊網」，以供透過網際網路查詢計畫成果相關資料書目索引，進而可連結內容影像檔瀏覽全文。系統平台架構包括典藏資料之數位內容管理功能、各計畫單位社群與帳號管理功能、系統及網站管理功能，以及典藏資料統計報表功能等。

鑑於部分計畫已建置擁有資訊網站，為期本資訊網達到最大整合服務功能，第二年規劃重點在於建置提供整合檢索服務，遵循 OAI 協定將中綱計畫其他服務網站納入檢索範圍。

三、建置成果宣導與推廣

完成建置之網站內容將與各相關機構合作，進行系統內容之推廣與使用宣導，分別針對人文社會科學相關領域中綱計畫成員及其他學術機構人員進行推廣宣導說明，對於部分尚未取得著作權授權之資料，亦將透過宣導方式爭取合法授權，以利資訊的傳播與學術應用。

參、典藏資料內容與特性

本系統典藏資料範圍為教育部顧問室及其所屬人文社會科學相關領域中綱計畫 2007 年至 2010 年所產出之成果報告及其相關資源為主，包括「人文教育革新中綱計畫」、「全球化下的台灣文史藝術中綱計畫」、「通識教育中綱計畫」、「新興議題與專業教育改革中綱計畫」、「海洋教育先導型中綱計畫」、「創造力教育計畫」及「先導型計畫及其他活動」等。(歷年計畫結構圖如圖 1 所示)

因各計畫典藏物件類型紛雜、檔案格式互異，為瞭解系統建置容量與工作量預估，經 2010 年 1 月 6 日與各項子計畫助理開會討論並說明物件數量統計調查表填寫方式後，於 1 月 14 日彙整回收各子計畫調查表填寫典藏數量統計，估計人文社會科學相關領域中綱計畫各項子計畫成果資料，分為 2009 年之前回溯處理與 2010 年新產生之待數位化物件，預估系統數位化資料類型與掃描工作量統計，包括紙質類文獻 341,935 頁、51GB；圖像資料 100 張、364GB；錄音資料 42GB；錄影資料 320GB；其它類型電子檔 60GB。

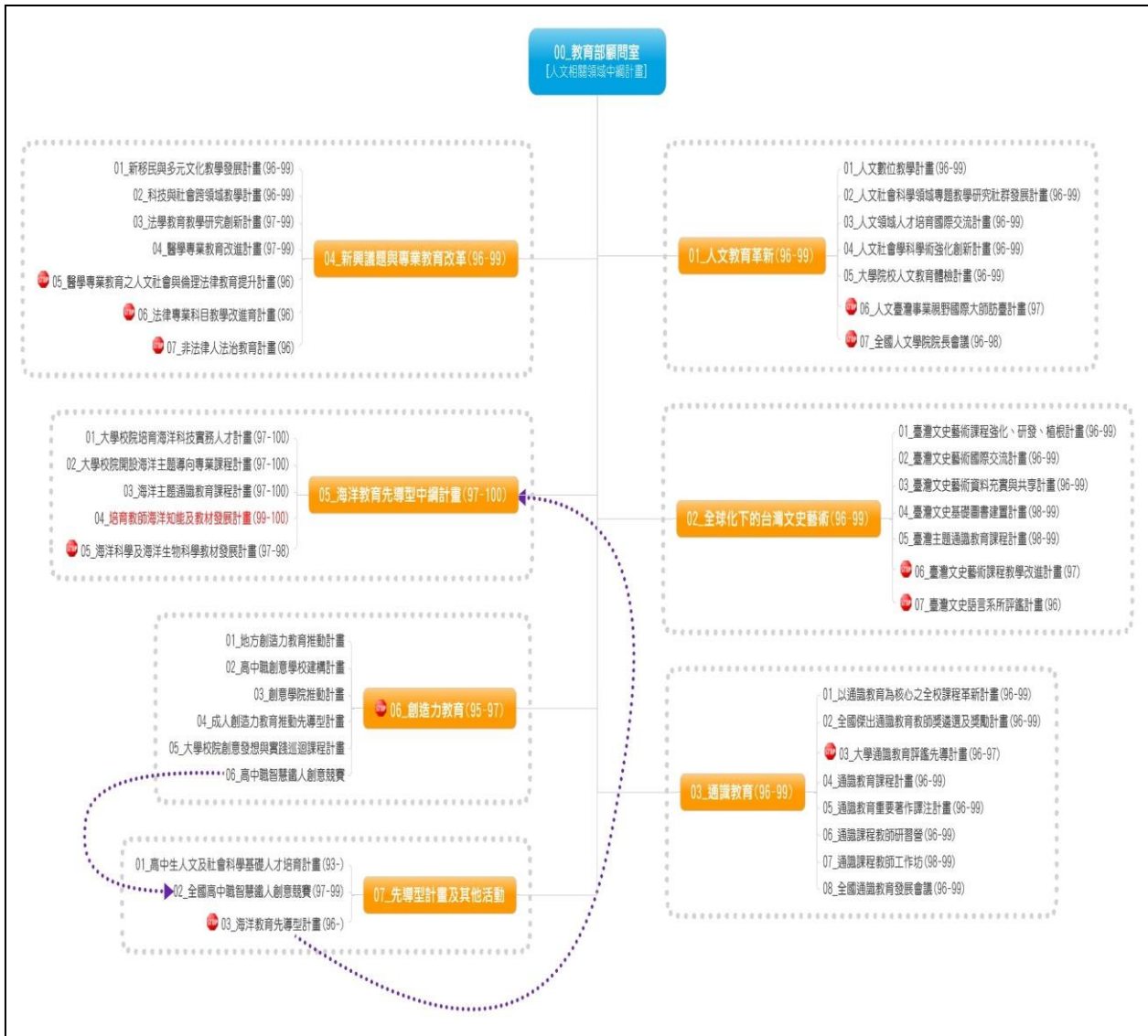


圖 1：人文社會科學相關領域中綱計畫歷年執行計畫關係結構圖

本計畫數位典藏資料特性，在資料規格上，因各計畫產出資料以紙本報告書為主體，故在尺寸上以 A4 較多，惟各計畫分由不同主持人進行，其成果報告缺乏一致性體例，故成果報告不僅頁數多寡差異極大，紙質與版式也不盡相同；在資料類型上，因各計畫辦理業務不同，故計畫產出資料類型迥異，但仍以紙質文獻及多媒體為主；而在資料保存狀態上，因所收錄資料為近三年所產生，故數位物件保存狀態較佳。

為考量未來數位化成果應開放使用，故數位物件之收錄原則，以沒有著作權疑慮者，且不涉及個人隱私及經費預算內容者為主，收錄時如已有電子檔，優先採用電子檔轉置，惟繳交電子檔內容需與紙本

內容相符，並避免典藏內容與紙本最終確定版內容有所出入，否則寧願重新掃描影像。報告書以年度報告、期末報告書為主，因期中報告為過程文件，暫不予典藏，若有計畫之期中報告具特殊典藏價值者，亦可列入收錄範圍。各類型資料收錄原則整理如下表：(表 1)

表 2 各類型資料收錄範圍與收錄原則說明表

資料類型		收錄範圍說明	收錄原則
紙質 類文 獻	報告書	指各計畫、徵件或活動執行之成果或年度報告書，如：全球化下的臺灣文史藝術中綱計畫所補助之「臺灣社會西化與現代化的發展歷程」期末報告書、教育部顧問室人文社會科學相關領域計畫 98 年度報告書等。	<ol style="list-style-type: none"> 1. 紙質文獻資料以排好版之電子檔或印刷出版品為主，若無法取得上述之電子檔或出版品，則可典藏相關之內容文稿。 2. 報告書以年度報告、期末報告書為主，因期中報告為過程文件，暫不予典藏，若有計畫之期中報告具特殊典藏價值者，亦可列入收錄範圍。 3. 部分計畫出版之專書因出版市場考量，該專書出版品僅典藏封面、導讀及目次。
	文宣品	指計畫、活動或會議執行時所製作之宣傳品，如：人文教育革心中綱計畫簡介摺頁、全球化下的臺灣文史藝術中程綱要計畫 DM 等。	
	會議論文	指各計畫所辦理各項會議之論文集。	
	手冊	指計畫過程產出之各項活動或會議手冊，如：96 年度北區通識課程教師研習營研習手冊、97 學年度第一學期通識教育課程計畫期初座談會暨教學助理研習會會議手冊、第一屆全國通識教育發展會議大會手冊等。	
	出版品	指各計畫所製作之出版品，如：通識教育中綱計畫委請國立編譯館與國立政治大學出版社合作翻譯發行之《培育人文》等此類專著或期刊。	
	其他	指非上述之相關資料產出，如：教育部全國傑出通識教育教師獎遴選及獎勵要點、教育部辦理補助大學校院推動以通識教育為核心之全校課程革新計畫徵求事宜等有關計畫辦理之程序性文件。	
圖 像	照片	指計畫、活動或會議執行時所拍攝之照片。	<ol style="list-style-type: none"> 1. 圖像資料繳交以數位檔

資料	圖片	指計畫、活動或會議執行時所產出之圖片，如：LOGO、活動橫幅等。	<p>為主。</p> <p>2. 圖像資料應以清晰為原則，解析度至少為150dpi以上。</p> <p>3. 圖像資料繳交應清楚描述資料內容，以利典藏。</p>
	海報	指計畫、活動或會議執行時所製作之海報。	
多媒體資料	錄影	指計畫、活動或會議執行時所拍攝之錄影資料。	<p>1. 多媒體資料內容以清晰可辨為原則。</p> <p>2. 多媒體資料請於清單說明資料內容要旨，以利判讀並進行後續數位化工作。</p>
	錄音	指計畫、活動或會議執行時所錄製之聲像檔案。	
	簡報	指計畫、活動或會議執行時所製作之簡報檔。	
網站		指各計畫、徵件或活動執行產出之相關網站，如：全國大學院校人文教育體檢計畫資料庫 (http://www.hee.tw/)、第八屆智慧鐵人創意競賽活動網站 (http://ironman.creativity.edu.tw/) 等。	<p>1. 本計畫對於網站典藏係進行網站之描述與連結。</p> <p>2. 本計畫連結之網站以未失效者為主。</p>
業務資料		本項資料是否典藏由各計畫決定，如有各計畫辦公室因執行業務產生之相關紀錄，如：會議通知、會議紀錄、通訊錄、網站或活動之規劃資料及具通用性之公文等，如考量具有後續辦理業務參考價值，可擇要加以典藏。	<p>1. 各計畫業務相關資料可選擇公開或不公開，不公開之物件於系統上僅為該計畫內部人員可瀏覽使用。</p>

典藏資料可選擇公開或不公開，囿於部分資料因涉及補助經費編列、計畫審查過程，其內容並不宜公開，如內容屬不公開之物件，僅開放該計畫有權限之內部人員瀏覽使用，其餘未限制使用之數位化物件，將開放一般大眾於網路瀏覽。

肆、數位典藏與系統建置

數位典藏是以網際網路為媒介，以文化資產為內涵，利用資訊技術為工具，所建置之數位資訊庫。本計畫經由下列程序完成數位典藏

工作：

一、制訂工作流程與規範

（一）實體物件的分析與造冊

首先需要進行實體物件的分析與選擇，瞭解需要數位化物件的保存狀況、數量估計、材質尺寸、媒體形式，基於經費與效益因素考量，需要衡酌典藏物件的性質，根據藏品價值、使用者需求、原件保存的風險、智慧財產權歸屬問題等原因，決定數位化的優先順序。
(Blodgett, Jan et al.,2007,PP.15-17)

本計畫預計數位化的實體物件散置於各研究計畫主持人手中，各中綱計畫主持人對於數位典藏的認知不一，需要透過面談說明的方式，逐一與各計畫主持人或助理說明數位典藏的意義與收錄的資料範圍，經為期半年（2010.03-2010.08）的訪談過程，一邊訪談瞭解實體物件數量形式，並釐清智財權問題後，一邊編造需要數位化的物件清單。

（二）決定數位化方式

對於文字性資料之數位化處理，需要根據物件特性與需求決定數位化方式，如果是以典藏原貌為重點，較節省經費的方式是採影像掃瞄或數位攝影方式，將物件掃成影像檔；如果強調內容文字的檢索需求，可以採用人工逐字輸入內文，或是以光學文字辨識(OCR)轉置為文字檔。

本計畫經實體物件狀況與數量分析後，對於紙質物件，因考量有原樣數位保存物件之必要，且衡酌經費與時效性，遂決定以數位掃瞄方式為主，依物件特性分別以平台式掃瞄器或饋紙式掃瞄器，進行數位化；至於為數有限之非紙質物件，如原件已是影音數位檔，則進行影音檔轉置，或是依規格將影音帶轉置為數位檔案形式。

（三）建立數位化工作流程

數位化是持續進行的工作，為確保不同時間以及各項工作參與者對於數位化工作過程有一致的瞭解，需要研訂數位化工作手冊，說

明作業程序、步驟、規範、品質要求與檢核標準等事項，做為作業指導原則與後續人員教育訓練參考。本計畫數位化工作程序與步驟，自掃描作業規劃、進行物件掃描與最後影像檔匯入系統等完整作業流程，預為規劃並繪製作業流程圖如下：(圖 2)

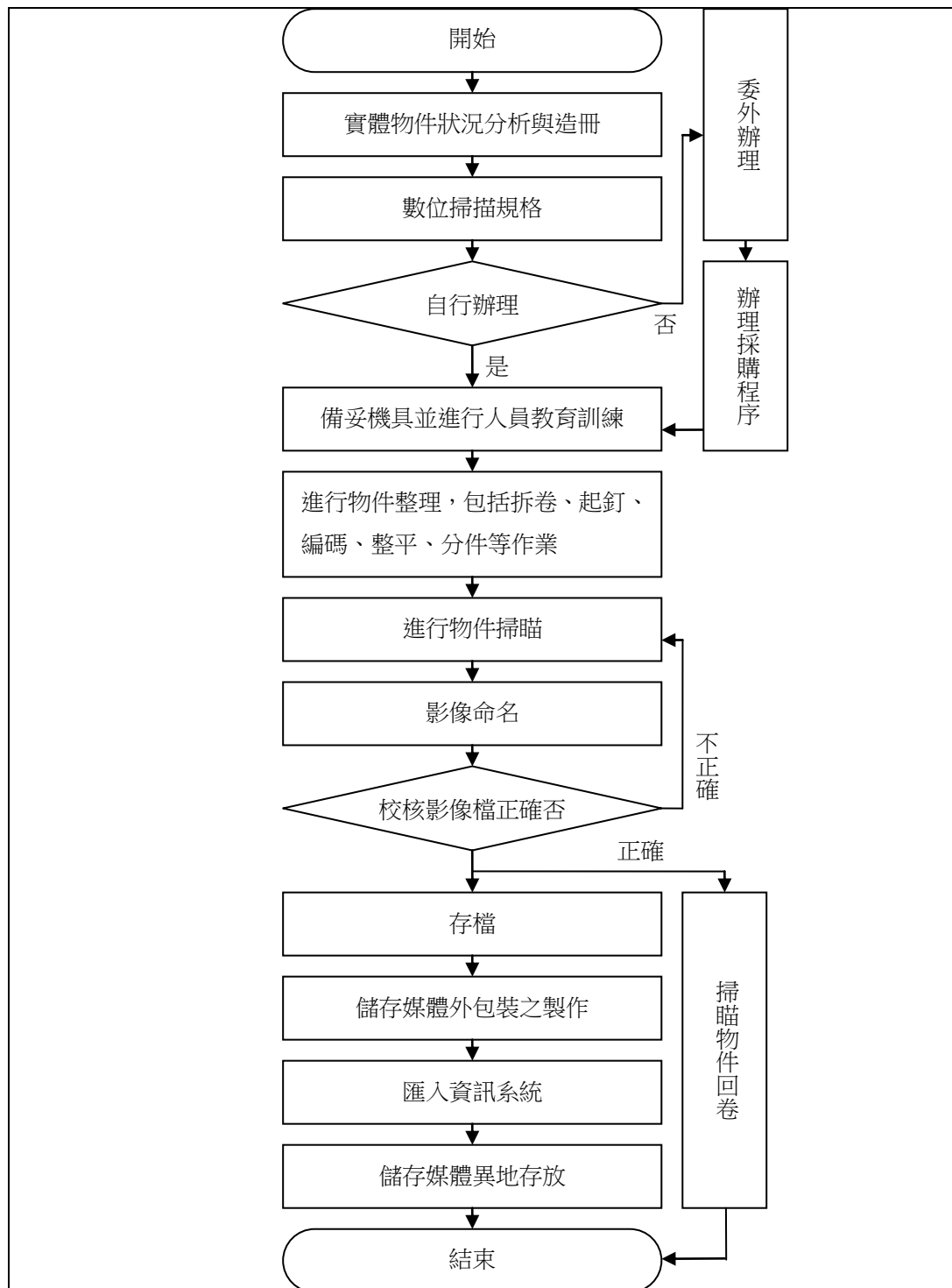


圖 2 本計畫數位掃描作業流程圖

因本計畫數位掃瞄除特殊物件外，主要採行委外由廠商辦理，有關影像品質要求與檢核標準，均訂於招標規範書與合約中，要求廠商依循。

二、影像數位化與 Metadata 建檔

(一) 影像數位化作業方式

本計畫因是教育部委託辦理的工作，並無持續人力投入，故影像掃瞄採用委外辦理方式，並依據行政機關辦理委外作業要求，依據台灣的《採購法》規定進行招商程序，先擬訂招標需求規範書，上網公告徵求廠商，經評選由最優廠商進行作業。

(二) 訂定數位化格式與作業規範

需要訂定之數位化格式規範，包括：檔案儲存格式(TIFF、JPEG、GIF)、影像解析度(200dpi-500dpi)、數位檔案命名原則與詮釋資料格式。數位化的格式因應用途需求，產生兩種格式數位檔：

1. 典藏用

(1) 彩色影像規格

檔案格式 TIFF；全彩(RGB)；不壓縮；解析度 300DPI。

(2) 灰階影像規格

檔案格式 TIFF；灰階；不壓縮；解析度 300DPI。

(3) 黑白影像規格

檔案格式 TIFF；黑白；解析度：300DPI。

2. 瀏覽傳輸用：

(1) 彩色影像規格

檔案格式 JPEG；全彩(RGB)；壓縮 90%；解析度 150 DPI。

(2) 灰階影像規格

檔案格式 JPEG；灰階；壓縮 90%；解析度 150DPI。

(3) 黑白影像規格

檔案格式 JPEG；黑白；解析度：300DPI。

掃描完成之電子檔採一物件一檔名設計，檔案命名為了能提供各子計畫的助記性與階層的邏輯性，檔名設計採四層結構，其命名原則為：計畫編號-資料類型代碼-DO 流水號-DO 用途代碼.jpg，例如：人文教育革新中綱計畫（01）人才國際交流計畫（03）的報告書，其檔案命名為 0103-Dr-001-i.jpg。¹

廠商根據規範於期限內完成所交付之數位化影像製作工作，經掃描完成之影像，進行校驗。若有遺漏或不符合要求之資料，應無條件修正，直至合乎要求為止。

經過檢測完成之影像檔，轉錄於 DVD-R 光碟片。燒錄後之光碟片，印上本校全銜及資料名稱（例：國立政治大學《典藏資料數位化掃描》）、光碟編號、資料類型、檔案格式及資料量等。

（三）建立 Metadata 格式與建檔

數位化物件的內容需要利用詮釋資料(Metadata，又稱後設資料或元資料)格式加以描述著錄，因應不同藏品特性需求，各專業領域已發展不同之詮釋資料格式，Metadata 類型的歸納不乏相關著作，可參閱中央研究院數位典藏系列圖書的整理。（數位典藏國家型科技計畫，2004；蔡永澄等，2007，頁 39-45）

本計畫對於 metadata 的選擇，是採用都柏林核心集 (Dublin Core 簡稱 DC)，因本計畫物件內容性質殊異，需尋求一個簡單明確，又可同時描述會議論文集、年度報告書、文宣摺頁、手冊、照片、圖片、海報等多元資料類型之 metadata，DC 因其具備描述物件共同之基本欄位又具延展擴充性，不失為最佳選擇。經計畫工作小組多次討論修訂 DC 基本元素，並嘗試著錄本計畫不同類型物件，最終於 2010 年 7 月確定本計畫採用之 metadata 著錄欄位，第一層著錄欄位包括：資料類型 (type)、計畫名稱 (title)、主題與關鍵詞 (subject)、描述 (description)、創作者 (creator)、其他貢獻者 (contributor)、出版單位 (publisher)、計畫期程 (date)、識別資料 (identifier)、來源 (source)、

¹ 資料類型共 2 碼（第一碼：教育部(O)、主計畫(M)、子計畫(D)，第二碼：報告書(r)、文宣品(p)、會議論文(c)、手冊(m)、其他(o)）；DO 流水號共 3 碼；DO 用途代碼共 1 碼，取 internet 之首位字母，i。

資料格式 (format)、關連 (relation)、作品語文 (language)、權限範圍 (rights) 等。

三、系統功能規劃與設計

規劃建置之系統以數位物件典藏管理功能為主，兼顧使用檢索需求，主要分為前端使用者服務及後端資料管理維護功能。

(一) 前端使用者服務功能

前端使用者服務主要功能，包括：

1. 提供使用者由計畫分類、點閱率排行、會議或活動類型、圖像/網站/多媒體類型線上瀏覽典藏內容；
2. 提供使用者簡易/進階查詢方式，查詢結果可條列式顯示，並可直接點選影像檔瀏覽全文影像；
3. 提供一般網站服務，如：最新消息、網站地圖、常見問題、網站連結等功能。

(二) 後端資料管理維護功能

後端管理功能，主要包括：

1. 提供資料管理，能進行詮釋資料維護與校核，並提供平台管理者或社群管理者可批次匯入詮釋資料與數位物件；亦能設定數位物件在前台的開放與存取權限。
2. 提供網站管理，允許平台管理者設定 DTD 各欄位於前台的顯示名稱與呈現方式；可自行設定開放於讀者端進階查詢功能中，所能查詢的欄位；能進行常見問題與相關網站之新增、修改及刪除，可設定是否開放和自訂排序規則。
3. 提供系統管理，平台管理者能設定 IP 範圍，做為控制前後台登入或是控制前台瀏覽 DO 的限制；亦包含系統執行所需的參數設定，於系統異動時，可由管理者直接調整設定。

4. 提供統計報表，可依 DTD 統計特定區間(日、週、月、季、年)入庫的資料量與網站使用統計，並提供報表。

伍、計畫成果與效益

本計畫執行期程自 2010 年 3 月開始，透過訪談中網計畫各計畫主持人或研究助理的方式，瞭解實體物件形式與數量，並釐清智財權問題後，著手編造數位化物件清單；同時規劃數位化作業流程與數位掃瞄規格需求，由計畫助理與工讀生進行數位資料整理、影像掃瞄檢核作業。系統功能之開發設計，最初由研究團隊擬具系統功能需求書，對外徵求廠商並據以開發系統，得標廠商經與本計畫主持人及助理多次討論，確認系統功能設計細節後，據以開發完成系統，施作過程將近六個月，但數位掃瞄作業仍持續回溯資料進行中，已掃瞄完成者陸續上傳系統，提供檢索查詢使用。

具體而言，本計畫執行結果的效益在於：

一、不受時地限制提供使用服務

本計畫建置之「人文社會科學相關領域中網計畫成果資訊網」，提供使用者可透過網際網路，不限時間與地點，隨時可上網查找資料，將過去分散各校各地之中網計畫成果，透過本資訊網站的建置，提供研究人員與一般民眾可不限時地的檢索利用各項計畫成果資訊。

二、設計多元檢索方式便捷查詢

本資訊網提供簡易與進階查詢兩種方式，允許使用者透過單一欄位查詢全部的詮釋資料欄位，也可以組合多個特定檢索欄位，以布林邏輯進行欄位間的組合運算；考量本計畫典藏物件屬性，特別設計多元瀏覽檢索功能，包括：計畫分類瀏覽、點閱率排行瀏覽、會議/活動類型瀏覽、圖像/網站/多媒體類型瀏覽功能，並可直接連結影像檔。不論是直接輸入條件值的查詢方式或是線上瀏覽點選的方式，係考量使用者檢索習性，提供使用者便捷的檢索介面設計。

三、典藏內容開放加速資訊傳播

過去中網計畫各計畫執行成果主要由各計畫執行團隊使用，除

部分計畫另有建置計畫網站公開計畫資訊外，多數分由各計畫辦公室典藏保管，計畫成果的流通與資訊傳播功能有限，透過本典藏網站的建置，將原本分散各處之計畫成果加以彙整並提供單一檢索平台，從而可促進各計畫成果內容的資訊傳播。

四、數位化提供資料備份的方式

數位化不僅可減少對於原件調閱的損害，更因為數位檔案複製容易，可複製分存異地，亦可達到資料異地備份的功效。中綱計畫各子計畫分由各校各主持人執行，雖執行期尚未年代久遠，但幾經研究人員易手，不少成果報告與書面資料僅剩孤本，數位化之後可輕易複製多份，降低孤本保存的風險。雖然數位媒體的保存持久性問題，尚無一勞永逸的解決方式，但在數位科技的迅速發展與普及下，以數位方式儲存資訊是未來發展趨勢，數位典藏增加資料備份分擔風險是本案成效之一。

陸、結語

本計畫接續第二年所進行之新增整合查詢功能，將可介接更多相關計畫系統的資訊，豐富本計畫網站可查詢資料之範圍，同時於系統開放線上服務後，能展開使用者檢索詞彙與檢索歷程分析，以更加瞭解本網站的使用者行為與使用需求，以提供本系統未來功能改善之參考。藉由本數位典藏資訊網建置過程，有關計畫執行經驗與心得最大的感觸在於：

- 一、 典藏系統可以不只是典藏功能，適度建立提供有權限使用者對自身資料管理的功能；
- 二、 系統建置完成後需瞭解真正使用者與使用需求，需保留使用記錄，做為未來系統使用分析評估；
- 三、 著作權問題應儘早因應，在資料完成之際就應該取得避免日後典藏處理時機的困難；
- 四、 過去典藏以珍稀性資源為主，因技術與儲存費用降低，典藏的概念不再僅限於過去資料，有業務功能與後續使用價值的

資料，在產生之際就可以同步進行數位典藏保管，反而可在資料產生最完整之際就維護資料的完整性，以免時過境遷反而難以取得原件；

五、 典藏可以由文件產生者在文件產出之際就納入數位典藏，由文件產生者分擔詮釋資料建檔中內容主題性欄位，可以達到文件內容的最佳詮釋；

六、 如何建立典藏物件之間內容關連性的建立，需要找出相關判定的欄位，本計畫嘗試由題名與計畫層級，建立典藏物件之間的關連，提供使用者檢索瀏覽的機制。

參考書目

洪淑芬(2004)。文獻典藏數位化的實務與技術。臺北市：數位典藏國家型科技計畫訓練推廣分項計畫。

陳雪華、陳昭珍、陳光華合編(2001)。數位圖書館XML/Metadata 管理系統。臺北市：文華。

教育部顧問室(2009)。教育部顧問室人文社會科學相關領域計畫98年度報告書。臺北市：教育部。

蔡永橙等著(2007年)。數位典藏技術導論。台北市：國立台灣大學出版中心、中央研究院。

數位典藏國家型科技計畫(2004)。數位典藏技術彙編。Retrieved August 20, 2010, from <http://www2.ndap.org.tw/eBook/showContent.php>

Blodgett, Jan et al. (2007). *Guidelines for Digitization*. North Carolina : State Library of North Carolina. Retrieved September 20, 2010 from <http://www.ncecho.org/dig/digguidelines.shtml>

Hughes, Lorna M. (2004). *Digitizing Collections: Strategic Issues for the Information Manager*. London: Facet Publishing.