

中央研究院近代史研究所檔案館

外交經濟重要檔案數位化工作流程簡介

修訂日期：93/11/29

計畫單位：中央研究院近代史研究所

計畫名稱：近代外交經濟重要檔案數位化計畫

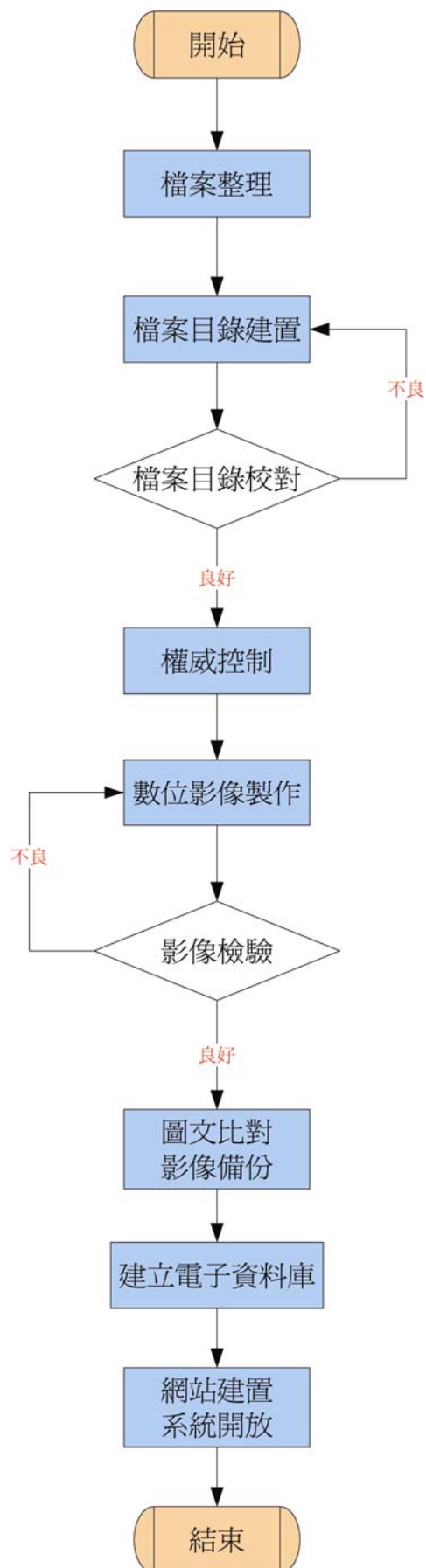
計畫簡介：

中央研究院近代史研究所籌備處於 1955 年 2 月成立，1965 年 4 月正式設所，同年徵得清末以來外交部門的官方檔案，1966 年經濟部亦移交大陸時期經濟檔案，長久以來，近史所檔案館度藏的外交與經濟檔案，為研究清末民國時期外交、財經方面重要史料，數量龐大，利用頻率也相當高。為了長久保存這批史料，近史所檔案館於 1991 年進行檔案微縮的拍攝，1997 年改以數位典藏，至 2004 年止，總計完成約 300 萬頁的影像檔案。

以下就數位化工作流程做一說明。

數位化工作流程說明

數位化工作流程圖



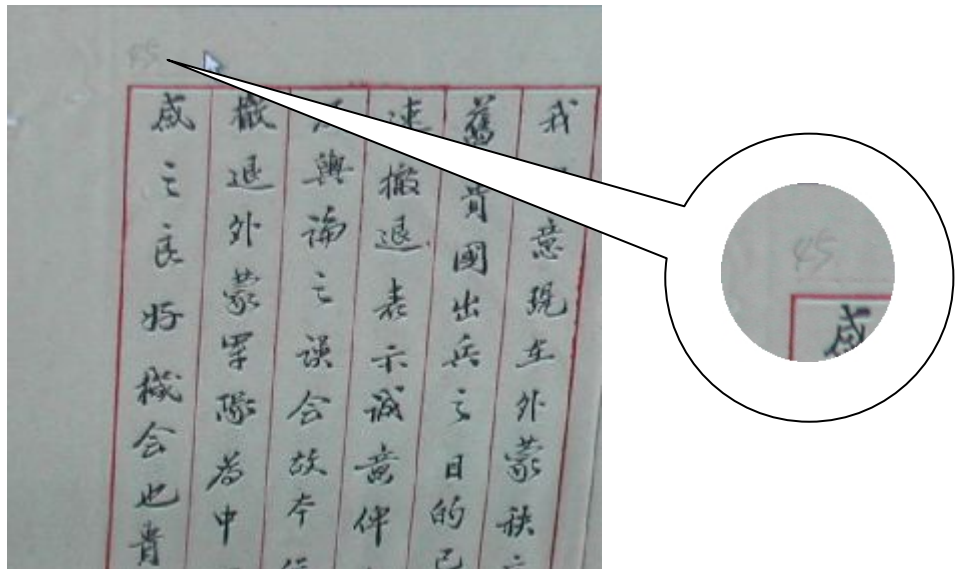
更新時間：2004/11/30

一、檔案整理

原檔案已按全宗、系列、冊層級作好整編，為配合影像製作必須逐冊拆卷、編寫每冊文件號及每冊文件之流水編號。檔案拆卷（圖一）前須檢查全宗號、系列號、宗號、冊號與本宗函目是否相符，將文件逐頁攤開，編頁作業依《外交檔案編頁規範》進行，編頁時使用 2B 鉛筆，將頁碼寫在檔案原件的左上角（圖二）。長寬均大於 A3 的檔案，整編時先將大於 A3 文件抽出，並於抽出文件的位置放入一張註記「A0」的白紙，依續編上流水號，並將抽出的檔案登錄在清單上。



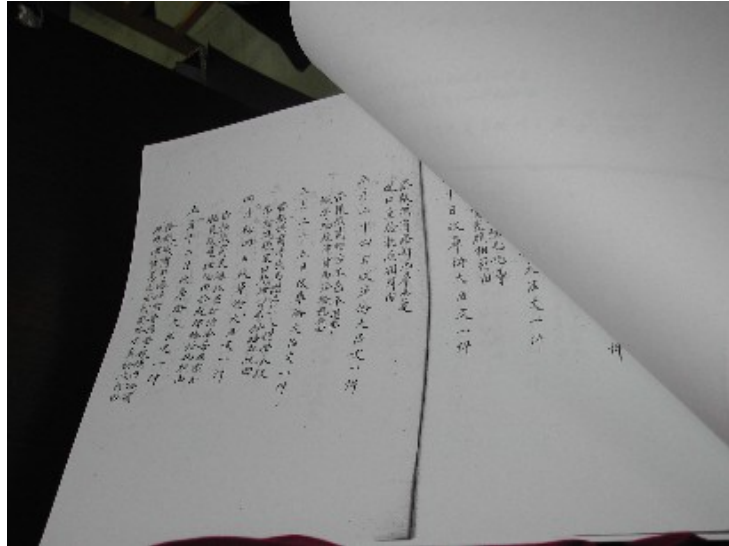
圖一：拆卷



圖二：編頁

二、檔案目錄建置

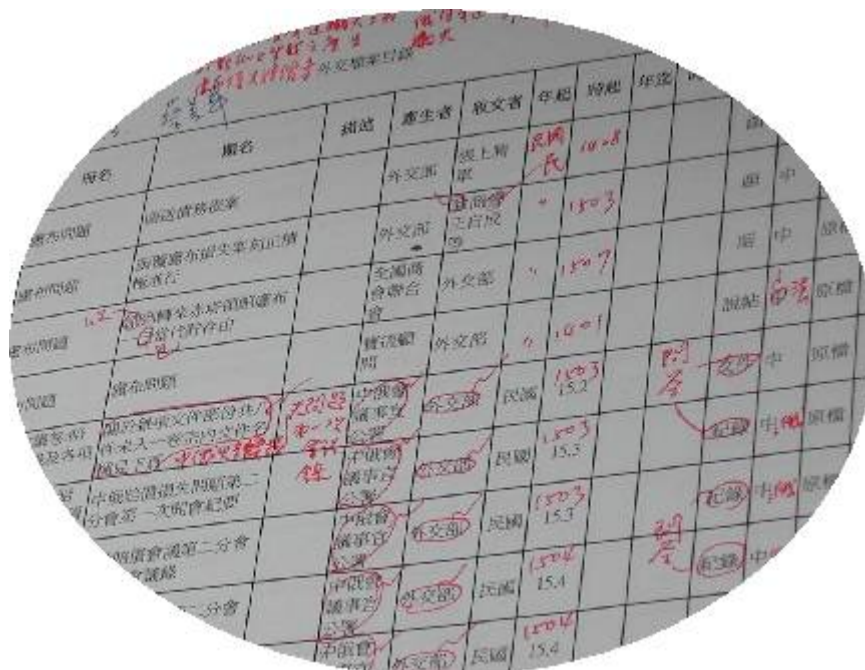
由檔案館影印每冊的檔案目錄（圖三），挑選 Metadata 中的基本欄位，依《外交檔案目錄建檔流程及規範》委外製作，採 Excel 格式建檔，完成一校一改後交還檔案館。



圖三：檔案目錄影印稿

三、檔案目錄校對

一校稿完成後，由檔案館負責人員合併、調整電子檔列印目錄，進行二校。二校人員需根據《外交檔案目錄校對流程及規範》配合檔案原件進行目錄校對，並註記收發欄中人名、機關等相關資訊，以利權威控制作業。



姓名	職名	籍貫	產生者	收文者	年起	時起	年迄	原檔
	國務院秘書		外交部	張上智	1908			原中
	新加坡布羅失業糾正債權通告		外交部	蕭商學	1903			原中
	新加坡轉業非英領事通告		全國商會聯合會	外交部	1907			原中
	廣布問題		實業顧問	外交部	1901			原中
	關於新加坡文書部共同		中領事館	外交部	1903			原中
	自來水一公司因文件名		中領事館	外交部	1903			原中
	關於上海中領事館		中領事館	外交部	1903			原中
	中領事館失物招領第二		中領事館	外交部	1903			原中
	分會第一次開會紀要		中領事館	外交部	1903			原中
	中領事館第二分會		中領事館	外交部	1904			原中
	中領事館		中領事館	外交部	1904			原中
	中領事館		中領事館	外交部	1904			原中

圖四：校對目錄



圖五：參考檔案原件進行二校

四、權威控制

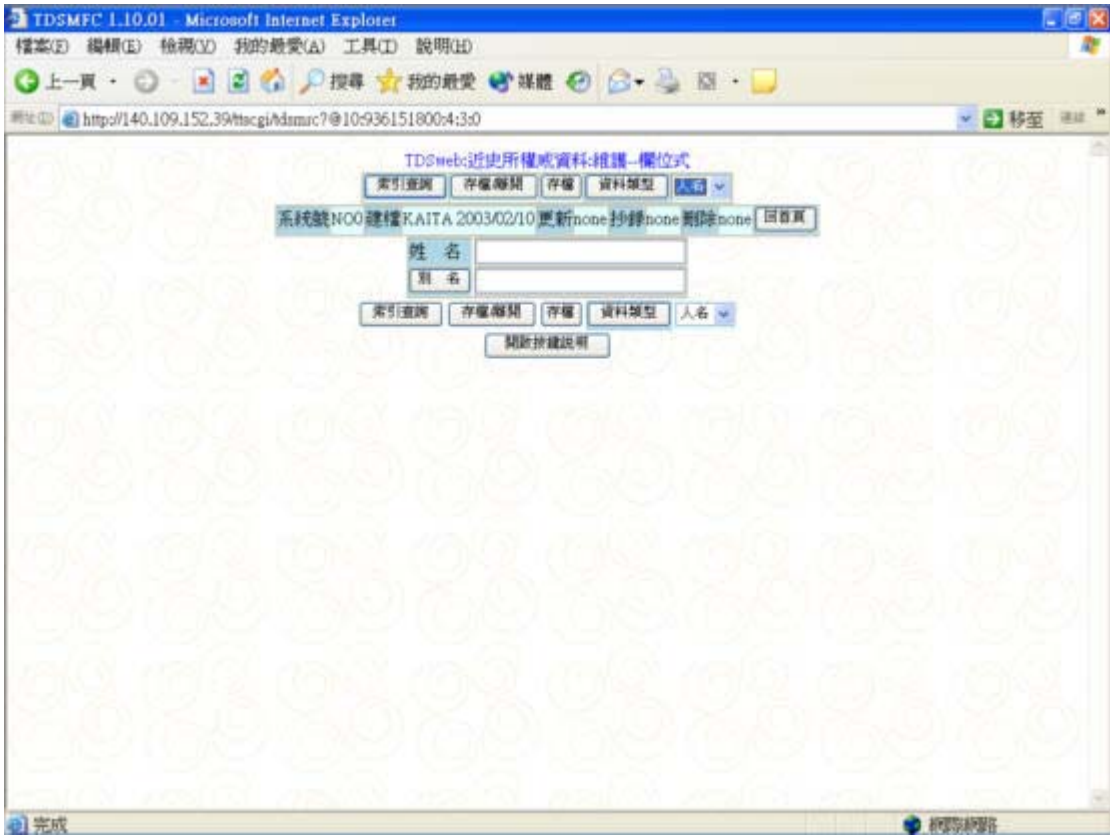
由於外交檔案原始文件收發者多只註記簡稱、職銜，為方便使用者查詢，必須對文件人名、地名作權威控制之加值。外交檔案權威款目包括「機關」名稱、「人名」之別名或字號、「職銜」以及「地名」等，這些訊息來自目錄中「收發文者」、「描述」、「冊名」、「題名」等欄位，作業人員根據原始文件查考相關資料（圖六），補填人名及職銜全稱（圖七）。權威控制作業項目包括（1）整理機關、人名、地名、職銜行政資訊；（2）製作權威對照表；（3）將權威控制欄位與別名檔在檢索系統中連行連結。（圖八）此項工作須查考史籍文獻，以具有史學背景、熟悉檔案內容的專業背景者擔任為佳。

H	I	J	K
收文者	機關權威	職銜權威	人名權威
各國使臣	外務部		
外務部		美國公使	田貝 Charles Denby
北洋大臣		北洋大臣	裕祿
美國田使		美國公使	田貝 Charles Denby
外務部		總稅務司	赫德 Robert Hart
北洋大臣		北洋大臣	裕祿
總稅務司赫德		總稅務司	赫德 Robert Hart
外務部		英國公使	賈訥樂 Claude Maxwell MacDonald
北洋大臣		北洋大臣	裕祿

圖六：權威控制資訊著錄表格(部分)



圖七：參考資料，如「辛亥以後十七年職官年表」、
「近代來華外國人名辭典」



圖八：大鐸系統權威著錄畫面

五、數位影像製作

(一) 檔案

數位影像製作採招標委外作業，為確保數位化影像之品質與完善的處理流程，並保護檔案原件不致受損，廠商必須將機器設備進駐近史所檔案館，進行掃描工作（圖九至圖十一）。影像規格為 300dpi、TIFF 格式，另轉製一份品質壓縮參數 75% 之 JPEG 光碟備分。由於影像規格高，掃描存檔長，操作人員可同時多機進行，唯同一宗檔案只能於同一部機器處理，以避免錯誤。影像掃描存檔後，交由後端裁切人員裁切、存檔。以上作業均須符合《外交檔案掃描規範》。



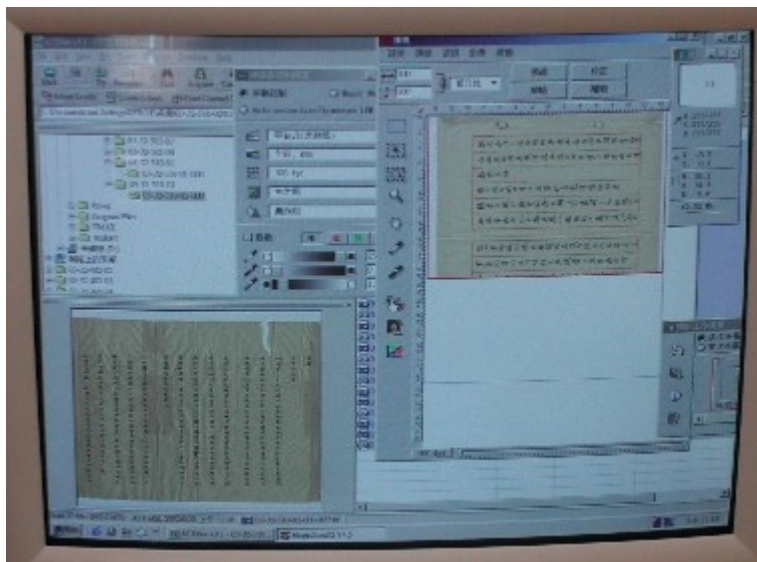
圖九：掃描現場



圖十：影像掃描



圖十一：影像掃描



圖十二：掃描時的電腦畫面

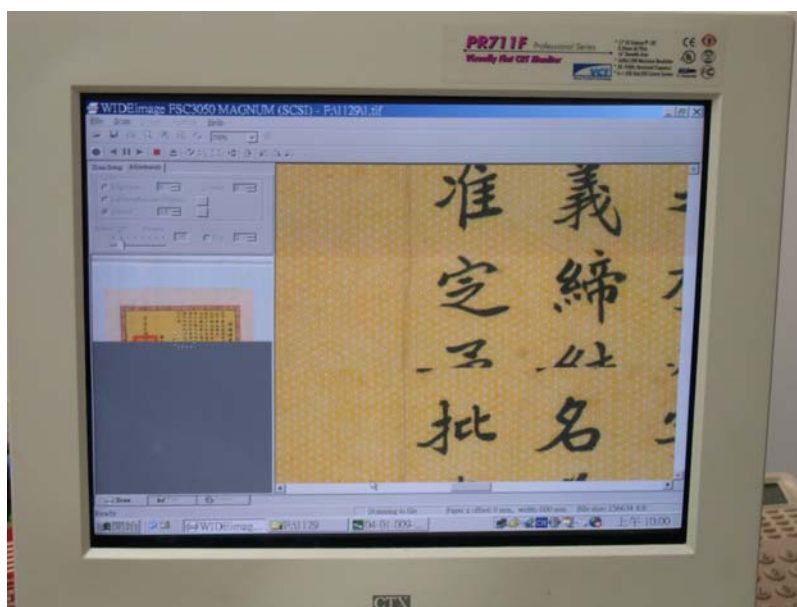
(二) 大圖

所謂大圖為長寬尺寸超過 A3 (42*29.7cm) 的資料，包括表單、地圖、契約、證書等。

編頁時註記「A0」的白紙掃描後檔案會特別小，依檔案大小即可從系統中辨識大圖的位置，大圖掃描完成後，將影像插入。(圖十三、圖十四)



圖十三：大圖影像掃描



圖十四：大圖掃描時的電腦畫面

六、影像檢驗

廠商完成初步數位影像製作後，須作預檢（圖十五、圖十六），再交由檔案館進行全數查核（圖十七），廠商更正錯誤後，檔案館再進行複查，複查時採十分之一影像抽驗，影像檢驗標準必須符合《外交檔案影像檢驗規範》。

檢驗過程中爭議性最大的為色彩，色差的產生可能來自多部掃描機有多種色調，也可能來自於同一台掃描機不同時間所產生的掃描色差。因此檔案館每月必須定時校色。校正方式為掃描色彩導表，一來比較不同機器的色度是否一致，二來每一台機器所掃描的色表須與前次所掃描的色表作

比較，以看出同一台機器前後的色差度。



圖十五：影像校正



圖十六：正在進行影像校對的圖檔



圖十七：館方人員進行驗片

七、圖文比對及影像備分

檢驗修正後影像，以宗為單位，儲存於 100G 抽取式硬碟交付館方資訊人員，製作一式兩份的磁帶備份（圖十八、圖十九）。一份存於檔案館，一份存於中研院計算中心。廠商並同時提供 JPEG 影像以硬碟交付館方，俟檔案館完成三校目錄後，進行圖文比對。

數位化製程中由於專業分工的需要，目錄與影像往往採取分別製作的方式產生。數位影像的命名原則通常由館藏號而來，故檔案目錄中的館藏號資訊一方面是影像掃描製作的依據，另一方面則是影像備分及系統整合時的校正依據。

檔案館以「件」為單位進行圖文比對，每一筆目錄應產生一筆該件的影像檔；比對時，可自實體檔案的電子目錄中擷取出館藏號欄位，製成「應掃檔案文號索引」，再自掃描後的影像檔名中產生「掃描檔案文號索引」，兩個索引檔的產生方式如圖二十。

圖文比對後，由檔案館人員更正錯誤影像檔名，上傳中研院計算中心主機後交還廠商燒錄光碟（圖二十一）。



圖十八：磁帶機與磁帶



圖十九：右側兩台即為大容量空間的磁碟陣列

1.目錄中的館藏號欄位：						1.掃描後產出之影像檔(表資料夾)：
全宗	系列	宗號	冊號	件號	...	\03-32-001-01-001\
03	32	001	01	001	...	03-32-001-01-001-001.JPG
03	32	001	01	002	...	03-32-001-01-001-002.JPG
03	32	001	01	003	...	03-32-001-01-001-003.JPG
2.修改格式製成「應掃檔案文號索引」：						\03-32-001-01-002\
03-32-001-01-001						03-32-001-01-002-001.JPG
03-32-001-01-002						03-32-001-01-002-002.JPG
03-32-001-01-003						\03-32-001-01-003\
						03-32-001-01-003-001.JPG
						03-32-001-01-003-002.JPG
						03-32-001-01-003-003.JPG
2.選用件層級製作「掃描檔案文號索引」：						03-32-001-01-001
						03-32-001-01-002
						03-32-001-01-003

圖二十：圖文比對



圖二十一：光碟備份

八、建立電子資料庫

外交檔案 Metadata 經檔案館與 Metadata 小組協商並比對 EAD 款目後，完成 Metadata 需求規格書 0.3 版，並由計算中心開發著錄系統。外交檔案分為全宗、系列、卷、件四層著錄格式，其欄位屬性及其著錄系統參見圖二十二至圖二十九。

項目名稱	英文名稱	資料型態	大小	必填	多值	屬性	提供者	
全宗號	Fonds Number	Varchar	2	*		下拉式選單	填表者	
全宗名	Fonds	Varchar	16	*		下拉式選單	填表者	
來源	Origin	Varchar	40	*			填表者	
館藏地	Repository	Varchar	40	*			填表者	
使用限制	Restriction	Varchar	2			單一選擇選單	填表者	
版權	Copy Right	Varchar	20				填表者	
傳記歷史註	Autobiography/ History	Text	-				填表者	
範圍與內容	Content	Text	-				填表者	
時間	朝代	Time	Dynasty	Varchar	10	*	下拉式選單	填表者
	年代	Period	Varchar	36				
實體描述	尺寸	Physical	Dimensions	Varchar	15			填表者
	範圍	Description	Extent	Varchar	10			填表者
處理資訊	著錄者	Catalog	Cataloger	Varchar	10	*	系統自動產生	系統
	著錄時間	Info	Catalog Date	Varchar	20		系統自動產生	系統

圖二十二：【全宗】需求欄位總表

全宗號：	<input type="text"/>
全宗名：	<input type="text"/>
來源：	<input type="text"/>
館藏地：	<input type="text"/>
使用限制：	<input type="text"/>
版權：	<input type="text"/>
傳記歷史註：	<input type="text"/>
範圍與內容：	<input type="text"/>
時間	朝代： <input type="text"/> 年代： <input type="text"/>
實體描述	尺寸： <input type="text"/> 範圍： <input type="text"/>
處理資訊	著錄者： <input type="text"/> (系統自動產生) 著錄時間： <input type="text"/> (系統自動產生) 修改者： <input type="text"/> (系統自動產生) 修改時間： <input type="text"/> (系統自動產生)

圖二十三：【全宗】著錄系統

項目名稱	英文名稱	資料型態	大小	必填	多值	屬性	提供者	
全宗號	Fonds Number	Varchar	2	*		下拉式選單	填表者	
檔案附屬層級	系列號	Description of Subordinate	Series Number	Varchar	2	*	全宗號之關聯下拉式選單	填表者
	系列名	Components	Series	Varchar	40		系統自動產生【註一】	系統
到館日期	Acquisition Date	Varchar	20	*			填表者	
範圍與內容	Content	Text	-				填表者	
時間	朝代	Time	Dynasty	Varchar	10	*	下拉式選單	填表者
	年代		Period	Varchar	36			填表者
實體描述	尺寸	Physical	Dimensions	Varchar	15			填表者
	範圍	Description	Extent	Varchar	10			填表者
處理資訊	著錄者	Catalog	Cataloger	Varchar	10	*	系統自動產生	系統
	著錄時間	Info	Catalog Date	Varchar	20		系統自動產生	系統

圖二十四：【系列／副系列】欄位需求表

紐交所外文檔案 - [系列/副系列] 層次

全宗號：

檔案附屬層級 系列號：
 系列名： (系統自動產生)

到館日期：

範圍與內容：

時間 朝代：
 年代：

實體描述 尺寸：
 範圍：

處理資訊 著錄者： (系統自動產生)
 著錄時間： (系統自動產生)
 修改者： (系統自動產生)
 修改時間： (系統自動產生)

圖二十五：【系列／副系列】著錄系統

項目名稱	英文名稱	資料型態	大小	必填	多值	屬性	提供者		
全宗號	Fonds Number	Varchar	2	*		下拉式選單	填表者		
檔案附屬層級	系列號	Description of	Series Number	Varchar	2	*	全宗號之關聯 下拉式選單	填表者	
	宗	Subordinate Components	Sub	Subject	Varchar	3		填表者	
			ject	Number	Varchar	50		填表者	
	冊	冊號	Fol	Folder	Varchar	2		填表者	
			der	Number	Varchar	60		填表者	
冊名	Folder	Varchar	60			填表者			
描述	Description	Text	-				填表者		
地名權威控制	Toponym Authority Control	Varchar	30		◎		填表者		
時間	起	Date	Begin	Dynasty	Varchar	4		下拉式選單	填表者
				Period	Varchar	4		下拉式選單	填表者
				Year	Varchar	3			填表者
				embolism	Int	2		【註二】	填表者
				Month	Varchar	2			填表者
	迄	End	Dynasty	Varchar	4		下拉式選單	填表者	
			Period	Varchar	4		下拉式選單	填表者	
			Year	Varchar	3			填表者	
			embolism	Int	2		【註二】	填表者	
			Month	Varchar	2			填表者	
參考資源	Description	Text	-				填表者		
件數	Quantity	Int	4				填表者		
館藏位置	Stack Area	Varchar	40				填表者		

圖二十六：【卷】欄位需求表

近代所外文檔案-[卷]層次

全宗號：

檔案附屬層級：系列號：

宗：宗號

宗名

冊：冊號

冊名

描述：

◎ 地名權威控制：

時間： 起： 朝代
年號
年
閏
月
迄： 朝代

圖二十七：【卷】著錄系統

項目名稱	英文名稱	資料型態	大小	必填	多值	屬性	提供者			
全宗號	Fonds Number	Varchar	2	*		下拉式選單	填表者			
檔案附屬層級	系列號	Description of Series Number	Varchar	2	*	全宗號之關聯下拉式選單	填表者			
	宗	宗號	Subordinate Subject	Varchar	3			填表者		
		宗名	Components Subject	Varchar	50			填表者		
	冊	冊號	Folder	Varchar	2			填表者		
		冊名	Folder	Varchar	60			填表者		
	文	文號	Item	Varchar	3			填表者		
		題名	Item	Varchar	100			填表者		
館藏號	Call Number	Varchar	16	*		系統自動產生【註三】	系統			
描述	Description	Varchar	200				填表者			
產生者	Originator	Varchar	50		☉		填表者			
收文者	Recipient	Varchar	50		☉		填表者			
別名權威控制	機關權威	Authority Control	Agency Authority Control	Varchar	20	☉	填表者			
	職銜權威	Control	Title Authority Control	Varchar	20	☉	填表者			
	人名權威	Control	Name Authority Control	Varchar	20	☉	填表者			
時間	起	朝代	Date	Begin	Dynasty	Varchar	4		下拉式選單	填表者
		年號	Period	Varchar	4			下拉式選單	填表者	
	年	Year	Varchar	3				填表者		
	閏	embolism	Int	2			【註四】	填表者		
	月	Month	Varchar	2				填表者		
	日	Day	Varchar	2				填表者		

圖二十八：【件】欄位需求表

近所外文檔案 - 【件】層次

全宗號：

檔案附屬層級： 系列號：

宗： 宗號

宗名

冊： 冊號

冊名

文： 文號

題名

館藏號： (系統自動產生)

描述：

產生者

收文者：

別名權威控制 機關權威：

圖二十九：【件】著錄系統

九、網站建置、系統開放

檔案館建置有「近代外交經濟重要檔案數位典藏計畫」網頁，(網址 <http://dipeco.sinica.edu.tw>)，提供外交檔案數位典藏之相關文件及規範，並與計算中心合作開發管理、檢索系統。



圖三十：近代外交經濟重要檔案數位典藏計畫網頁

- ※製作單位：中央研究院近代史研究所 近代外交經濟重要檔案數位化計畫
數位典藏國家型科技計畫 內容發展分項計畫
- ※文字撰寫：中央研究院近代史研究所檔案館 莊樹華
數位典藏國家型科技計畫 內容發展分項計畫 溫淳雅
- ※圖片拍攝：數位典藏國家型科技計畫 內容發展分項計畫 溫淳雅、陳嘉萍、
陳美智
- ※圖文編輯：數位典藏國家型科技計畫 內容發展分項計畫 溫淳雅、劉華珍

致謝：

感謝莊樹華小姐、張凱達先生、林威姝小姐、黃國斌先生的悉心指導，並蒙檔案館同仁提供協助，特此致謝。