

佛典數位典藏內容開發之研究與建構—經錄與經文內 容標記與知識架構 數位化工作流程簡介

修訂日期：2008/03/027

計畫單位：法鼓佛教研修學院

計畫名稱：佛典數位典藏內容開發之研究與建構—經錄
與經文內容標記與知識架構

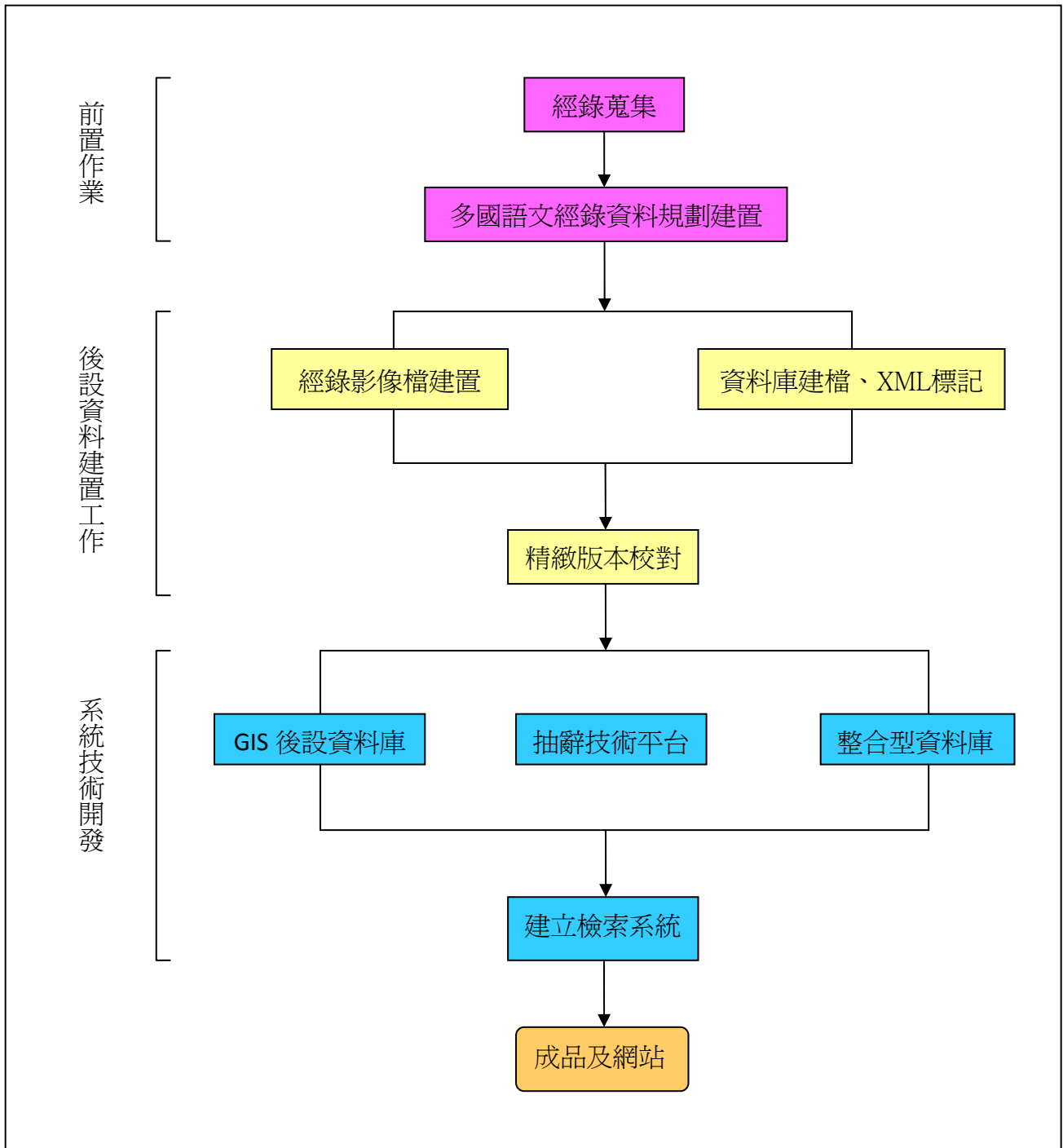
計畫簡介：

計畫將各版藏經目錄、歷代古經錄及佛教文獻資料等，配合當代資訊科技，以超文本 Hypertext 與 Metadata，如 XML / TEI markup 等標準規範，建構佛典知識管理系統。此一符合新時代的佛教經錄與全文資料庫，能在漢文經典文獻保存、知識管理，與學習、流通、推廣上，挹注一股連結歷史發展趨勢的新動力，成為台灣數位典藏與當代資訊化的基礎。

計畫團隊與合作典藏單位——中華電子佛典協會（CBETA）所產出的數位資料有相當深厚合作關係。本計畫以 CBETA 現有的數位典藏為基礎，以歷代佛經版本經錄、法寶總目錄及《法寶義林》等建構數位文獻資料，配合當代資訊科技與 XML TEI Markup 等標準規範進行內容開發，建構佛典知識管理系統，以提綱挈領掌握浩瀚佛典整體內容，服務學界與大眾。

計畫在國科會補助下已順利完成第二年計畫，完成建構多種語言經錄的資料庫內容，經錄文獻的數位化工作向前邁向一大步，完成古人所不能完成的強大功能多語言、多種版本比對連結及全文線上閱讀，計畫內容趨於更臻完整。

經錄與經文內容標記與知識架構數位化工作流程圖



經錄與經文內容標記與知識架構之數位化工作流程說明：

目前館藏的經錄資料分爲：「漢文佛典經錄」與「梵、巴、藏譯經錄」，本計畫除了將歷代中外所有佛典經錄進行全文數位化之外，並規劃建置多國語文藏經資料庫。

一、經錄蒐集

由具佛學文獻背景、且具藏經編輯歷史知識之工作小組同仁，蒐集各種藏經的紙本目錄、工具書、世學目錄中佛典的部分、或至法鼓佛教研修學院之圖書館查詢歷代中外所有佛典經錄。在蒐錄的過程中，難免會遇到有些古代藏錄無完整的經錄，如：宋代部分經錄有缺失，經工作小組與老師討論溝通後，這部分的經錄資料雖不完整，爲呈現經錄完整的歷史架構，故此部分的經錄資料還是會進行數位化的工作。



經錄原件館藏樣貌。



其他館內珍藏資料。

二、多國語文藏經資料庫規劃建置

在數位化工作進行的同時，與資訊人員討論並規劃建置多國語文藏經資料庫。

多國語文藏經資料庫包括：

1. 漢文藏經目錄
2. 藏文藏經目錄
3. 梵文藏經目錄
4. 巴利語藏經目錄
5. 滿文大藏經目錄

因佛典經錄文獻來源具多樣性，各種多語言佛典之經錄，因時間、空間、人為等因素，產生多語言經錄內容與編排方式的差異。所以，規劃建置多語文經錄時，必須依文本經錄性質，做不同資料庫欄位的建置。基本上，多語文經錄的共

通欄位有經名、作譯者、翻譯年代、頁數等。

三、經錄影像檔建置

將蒐集整理好的經錄資料，紙本資料的部分做掃描、委外翻拍，轉換成電子檔，集合電子檔資料的部分做重新編排整理。

(一) 自行掃描

(1) 掃描：

將蒐集整理好的經錄資料，紙本資料的部分做掃描或委外翻拍（下面會介紹），轉換成電子檔。因有些經本太重或太大不好處理，故視資料狀況使用進行掃描。

掃描要點如下：

1. 掃描。
2. 抽樣查看掃描品質——有無線條或歪斜不清者。
3. 掃描完畢後，就奇數頁與偶數頁檢查有無漏頁。
4. 編頁碼——先編奇數頁後編偶數頁，然後合併。
5. 抽樣檢查頁數正確與否。
6. 轉檔。
7. 燒錄。
8. 燒錄完成後，瀏覽檔案，若有缺漏或無法開啓的檔，加以修改或補齊。
9. 歸檔。
10. 清潔掃描器。

經錄資料掃描時，先掃成典藏級圖檔，解析度 300dpi、tif 檔之較高階影像，色彩模式為灰階或黑白，可供日後依不同目的降階應用。為了便於之後的校對工作進行與網站瀏覽呈現，再將圖檔另外轉成 Tif-g4 黑白格式，考量轉成 Tif-g4 黑白格式，是因這種格式檔案容量小，且畫質清晰。

房山石經(隋唐刻經)第一冊目錄

一	妙法蓮華經(七卷)	末附五十三佛名(姚秦鳩摩羅什譯)	隋、唐	一
二	維摩詰經(三卷)	(姚秦鳩摩羅什譯)	隋、唐	七七
三	大方廣佛華嚴經淨行品華嚴百四十願(二卷)	(東晉佛陀跋陀羅譯)	隋、唐	八六
四	佛說溫室洗浴衆僧經(一卷)	(後漢安世高譯)	隋、唐	八八
五	金剛般若波羅蜜經(一卷)	三藏菩提流支在胡相國秦太上父宣公弟譯	隋、唐	八八
六	勝鬘獅子吼一乘大方廣經(二卷)	(劉宋中印度求那跋陀羅譯)	隋、唐	九五
七	無量壽經(一卷)	(蕭齊曇摩伽陀耶舍譯)	隋、唐	九九
八	佛說觀彌勒菩薩上生兜率陀天經(二卷)	(劉宋沮渠京聲譯)	隋、唐	一〇五
九	無量壽經優婆塞提舍願生偈(一卷)	(梁嚴梨豆菩薩造、元魏菩提流支譯)	隋、唐	一〇八
一〇	受菩薩戒法(一卷)	出《菩薩地持經》(北凉曇無讖譯)	隋、唐	一〇八
一一	賢劫千佛出賢劫經(一卷)	隋、唐	一〇九
一二	八戒禪法(二卷)	(劉宋沮渠京聲譯)	隋、唐	一一四
一三	十方佛(一卷)	出自《十方滅闍冥經》(西晉竺法護譯)	隋、唐	一一五
一四	三十五佛并懺悔文(二卷)	出《決定毗尼經》(西晉鳩摩羅什譯)	隋、唐	一一五
一五	佛臨般涅槃略說教戒經(二卷)	(姚秦鳩摩羅什譯)	隋、唐	一一六
一六	大王觀世音經(一卷)	隋、唐	一一八
一七	文殊師利所說摩訶般若波羅蜜經(二卷)	(梁曼陀羅仙譯)	隋、唐	一一九
一八	彌勒下生成佛經(一卷)	(姚秦鳩摩羅什譯)	隋、唐	一二九
一九	請觀世音菩薩消伏毒害陀羅尼咒經(二卷)	(東晉竺難提譯)	隋、唐	一三二
二〇	佛說七種施因緣經(一卷)	(失譯人)	隋、唐	一三六
二一	佛說甚深大迴向經(一卷)	(失譯人)	隋、唐	一三七

(二) 委外翻拍：以滿文藏經作業為例

(1) 將《滿文大藏經》紙本目錄翻拍

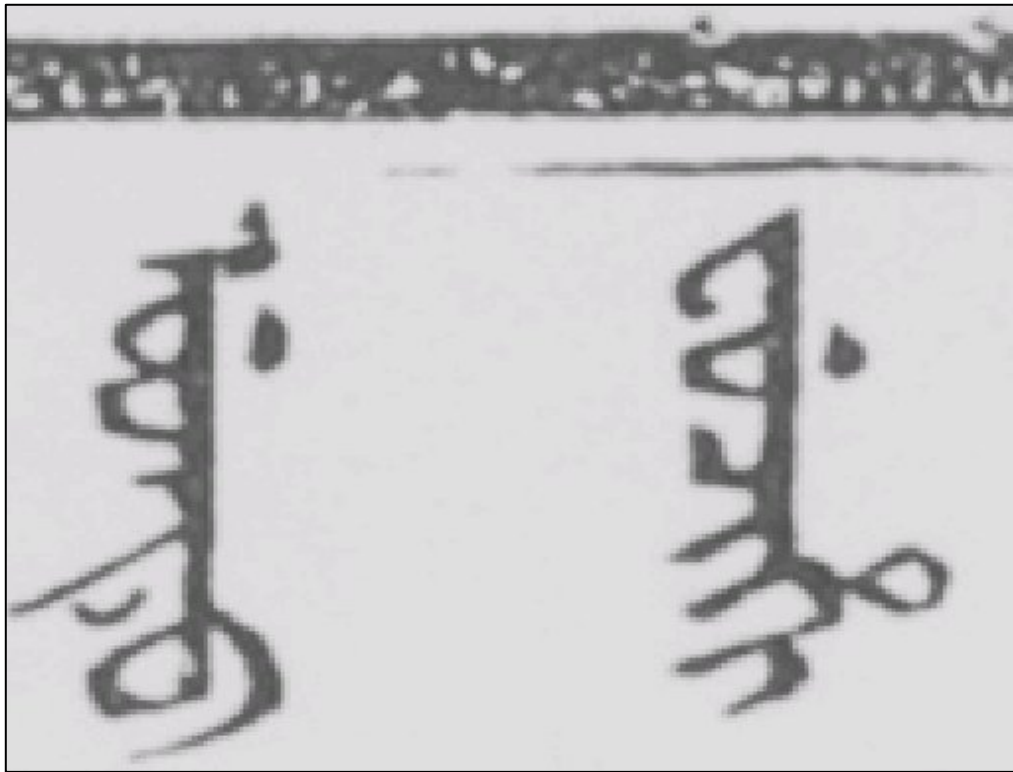
為提供使用者比對查詢，最理想翻拍是直接以「A2 平台式書本掃描器」，進行掃描，解析度與效果皆比照相機好。但受限於經費與設備不足，只好退而求其次，將《滿文大藏經》紙本目錄委外專人進行翻拍。委外人員使用超高解析度照相機進行翻拍工作，由於有些經目原本印刷字跡就不太清楚，需翻拍多次，才能達到令人滿意的效果。

在確認《滿文大藏經》紙本目錄翻拍前，先進行不同數位檔規格之相關測試：

第一種方式：

相機直接拍攝成 JPG 檔，每張大小約 2691KB。

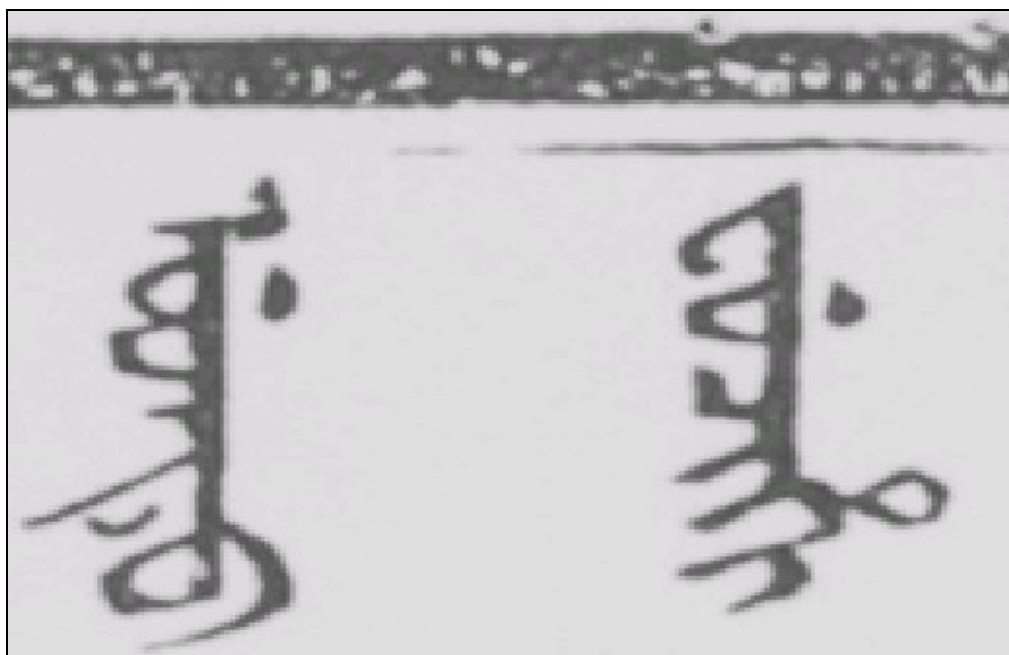
所得到的結果為：



第二種方式：

從相機原始格式（the NEF raw file）轉成 JPG 檔，每張大小約 9870KB。

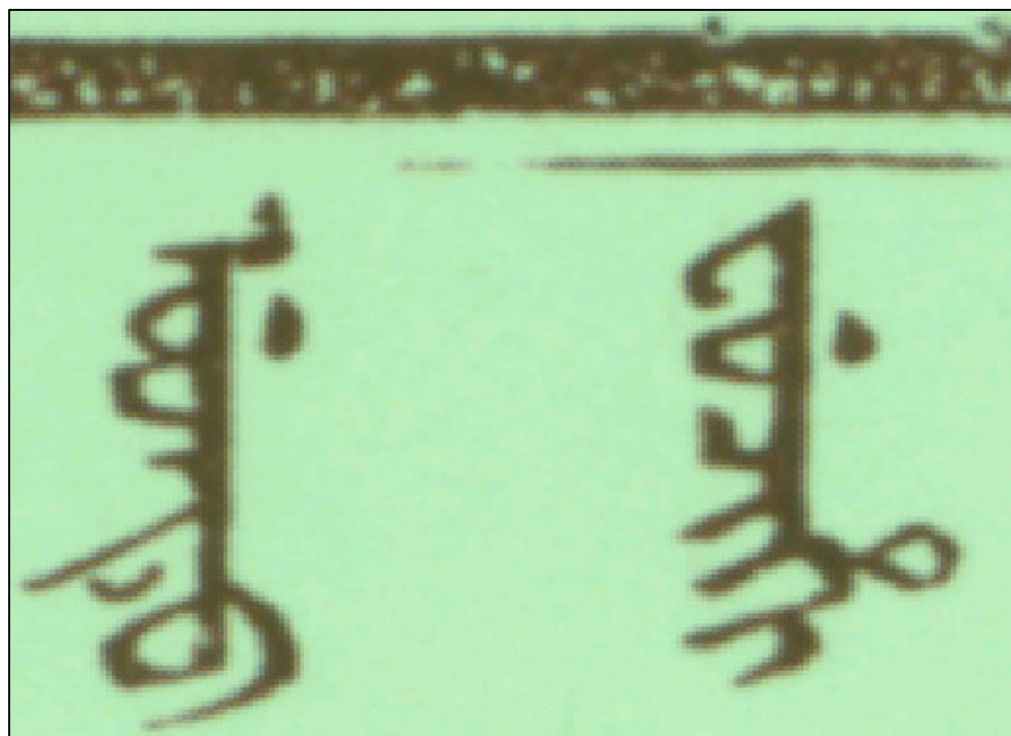
所得到的結果爲：



第三種方式：

從相機原始格式（the NEF raw file）轉成 TIFF 檔，每張大小約 29,883 KB。

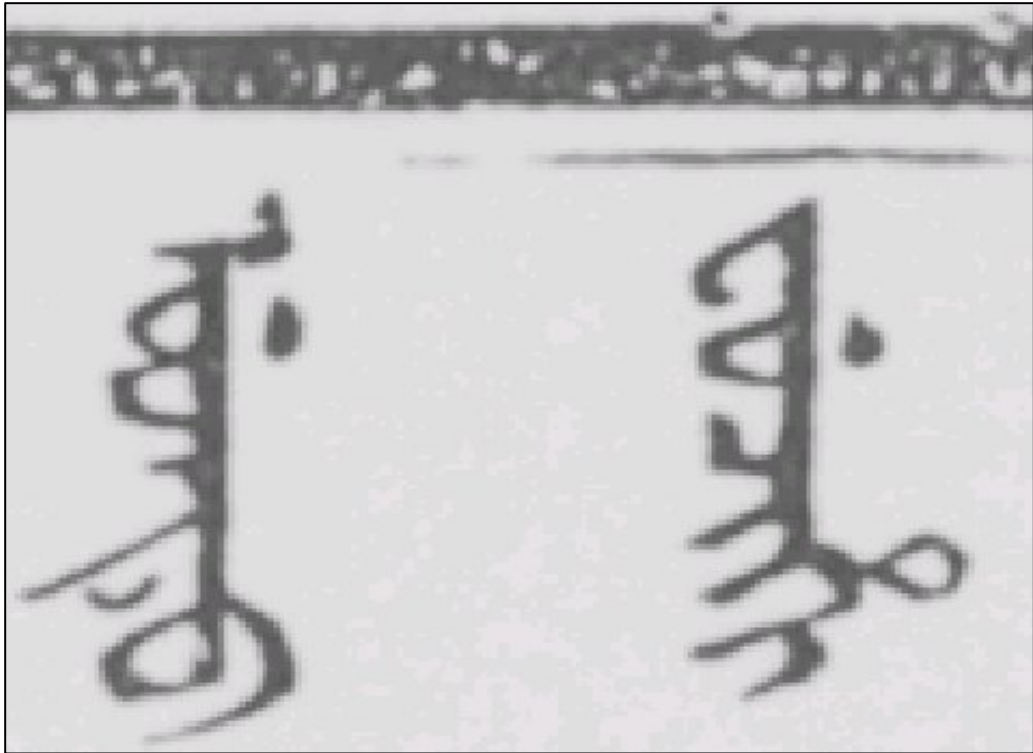
所得到的結果爲：



第四種方式：

從相機直接拍攝成 JPG 檔，再轉成 TIFF 檔，每張大小約 29,424KB。

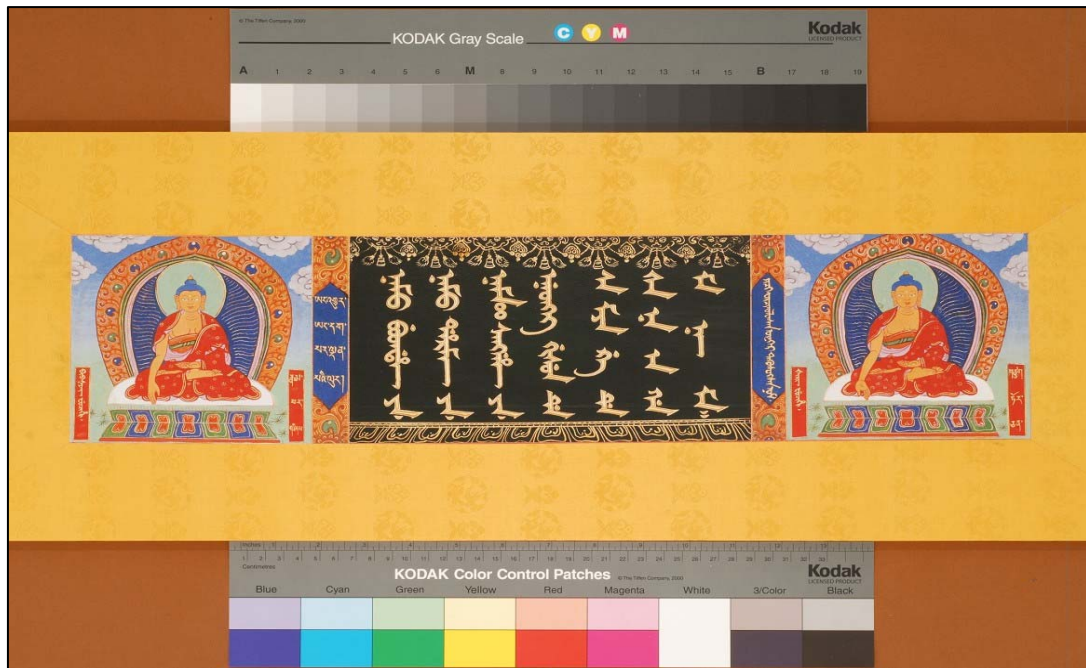
所得到的結果爲：



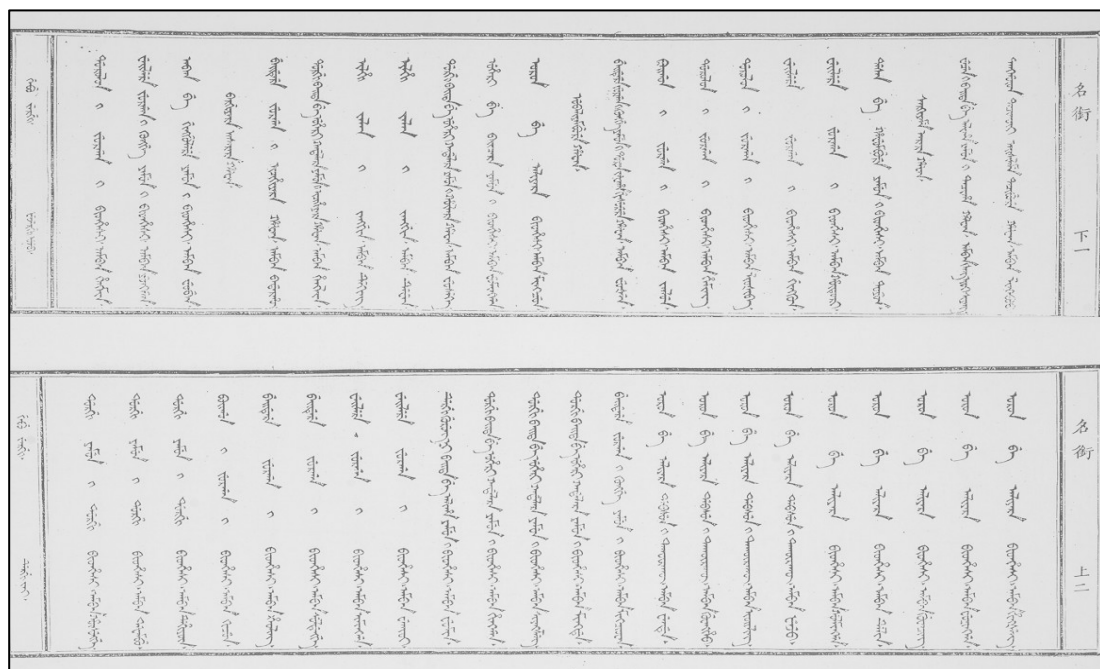
最後得到的結論為：若是為做成資料庫與上網連結，視覺效果與檔案大小，第一種最理想，故之後都是使用第一種數位檔規格進行經錄數位化工作。

(2) 建置《滿文大藏經》影像檔資料庫

照原書目錄的函號，進行基礎資料庫的建檔工作，依序建立圖檔資料庫，並進行校對。



《滿文大藏經》封面翻拍影像檔。



《滿文大藏經》內文翻拍影像檔。

四、資料庫建檔、XML 標記

(1) 各版藏經經錄等基本資料的輸入繕打與建檔：

先從漢文藏經目錄資料開始基礎的資料庫建檔工作，主要是將經目資料全文輸入至 Access 對應資料庫，建立欄位及內容統一格式的資料，這階段

所著錄的基本欄位包括：權威碼、冊數、經號、經名、卷數、朝代、作譯者、頁數、備註。

冊數	經號	經名	經名備註	書前目錄別名
007	0001	大般若波羅蜜多經		
007	0002	放光般若經		放光般若波羅蜜經
007	0003	摩訶般若波羅蜜經		
007	0004	光讚經		光讚般若波羅蜜經
007	0005	道行般若經		道行般若波羅蜜經
007	0006	小品般若波羅蜜經		
007	0007	摩訶般若經		摩訶般若波羅蜜經
007	0008	大明度經		大明度無極經
008	0009	勝天王般若波羅蜜經		
008	0010	金剛般若波羅蜜經		
008	0011	金剛般若波羅蜜經		
008	0012	金剛般若波羅蜜經		
008	0013	能斷金剛般若波羅蜜多經		
008	0014	佛說能斷金剛般若波羅蜜多經		
008	0015	金剛能斷般若波羅蜜經		
008	0016	佛說清淨普曜無上清淨分衛經		一名
008	0017	佛說仁王般若波羅蜜經		仁王護國般若波羅蜜經
008	0018	寶相般若波羅蜜經		
008	0019	摩訶般若波羅蜜大明呪經		
008	0020	般若波羅蜜多心經		
008	0021	文殊師利所說摩訶般若波羅蜜經		
008	0022	文殊師利所說般若波羅蜜經		
008-009	0023	大寶積經		
009	0024	大方廣三昧經		
009	0025	佛說無量清淨平等覺經		
009	0026	佛說阿彌陀三耶三佛薩樓佛檀過度人天經		佛說阿彌陀經
009	0027	佛說無量壽經		
009	0028	阿[門]@((人/人*人))佛國經		佛說阿[門]@((人/人*人))佛國經
009	0029	佛說大樂十法經		
009	0030	佛說普門品經		
009	0031	文殊師利佛土嚴淨經		
009	0032	佛說觀胎經		
009	0033	法鏡經		佛說法鏡經
009	0034	郁迦羅越問菩薩行經		
009	0035	佛說幻士仁賢經		幻士仁賢經
009	0036	佛說決定毗尼經		

漢文藏經後設資料著錄。

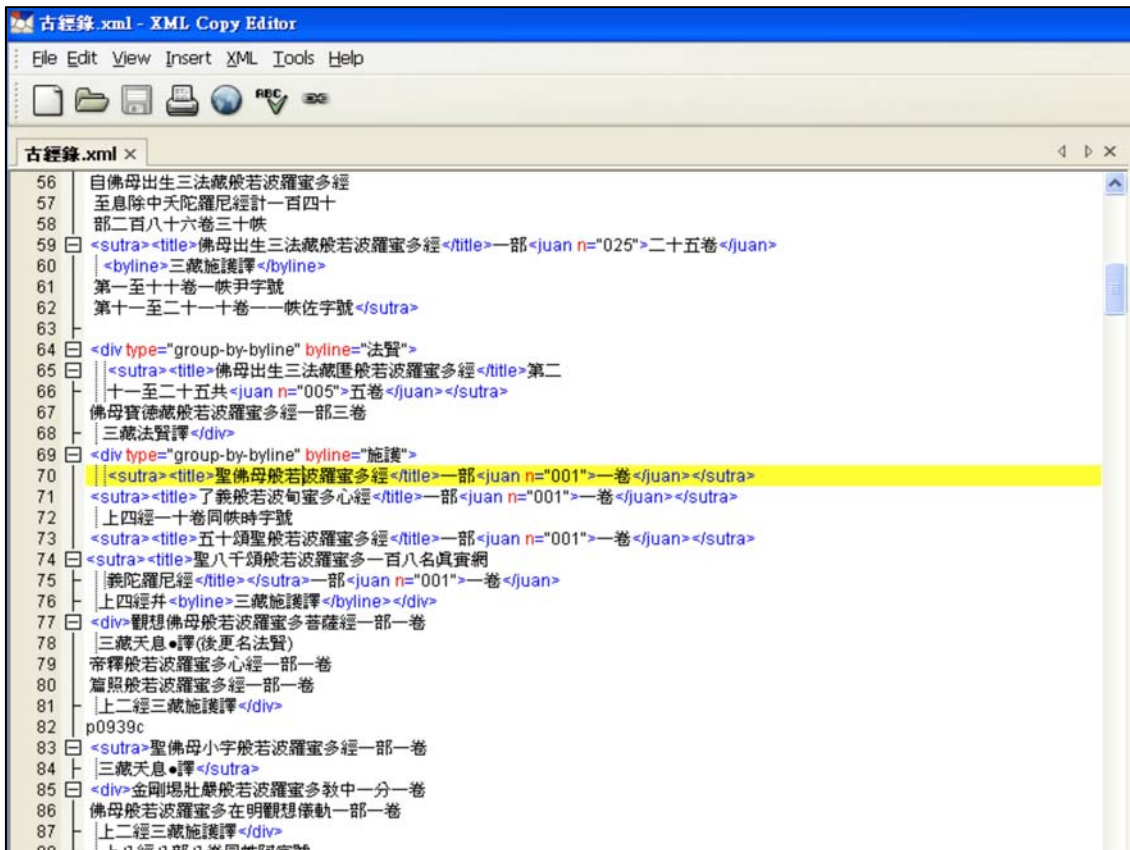
(2) 《滿文大藏經》經錄等基本資料的輸入繕打與建檔

工作小組先討論、建立起欄位及內容統一的格式後，搭配影像數位檔資料，進行基礎資料庫建檔工作。有些經目原本印刷字跡就不甚清楚，需查閱其他的工具書，協助判讀。

《滿文大藏經》的後設資料著錄，工作小組直接以 xml 檔建立數位化工作，不再用資料庫 Access，主要的後設資料欄位包括有：函號、滿文經題、中文經題、藏文經題、頁數、卷數、圖檔瀏覽等五項。

(3) 古經錄全文 XML 標記

目前經錄中，包括「漢譯經錄」與「梵、巴、藏譯經錄」，其中「漢譯經錄」又分為「現存藏經目錄」與「古經錄」兩種，「現存藏經目錄」是使用前面所提的資料庫 Access 建置而成。為順應潮流，本計畫之工作小組在「古經錄」資料庫建置時，使用 XML 作標記，建立符合國際標準的 XML 語言經錄資料庫，將經錄資料作另一種呈現。



古文經錄 XML 標記。

五、精緻版本校對

(一) 一次校對

該計畫利用電腦檔案比對，即同一份經文內容，由兩個版本予以輸入，然後以檔案比對程式找出兩者差異，再以看圖校對方式進行訂正，產生一份超越一般人工校對水準之經文檔。

首先，收集兩份同一經文但輸入來源不同之純文字電子檔。若有一頁一頁的小檔，可利用「檔案合併程式」，將兩檔各自所含小檔之純文字檔案合併成大檔，以利文書編輯處理及後續比對作業的進行。

將合併成大檔之兩檔匯入「檔案比對程式」，執行第一次兩檔比對。比對後產生一個主要差異檔。有差異的地方，再用人工對回原典做再次確認。

(二) 二次校對

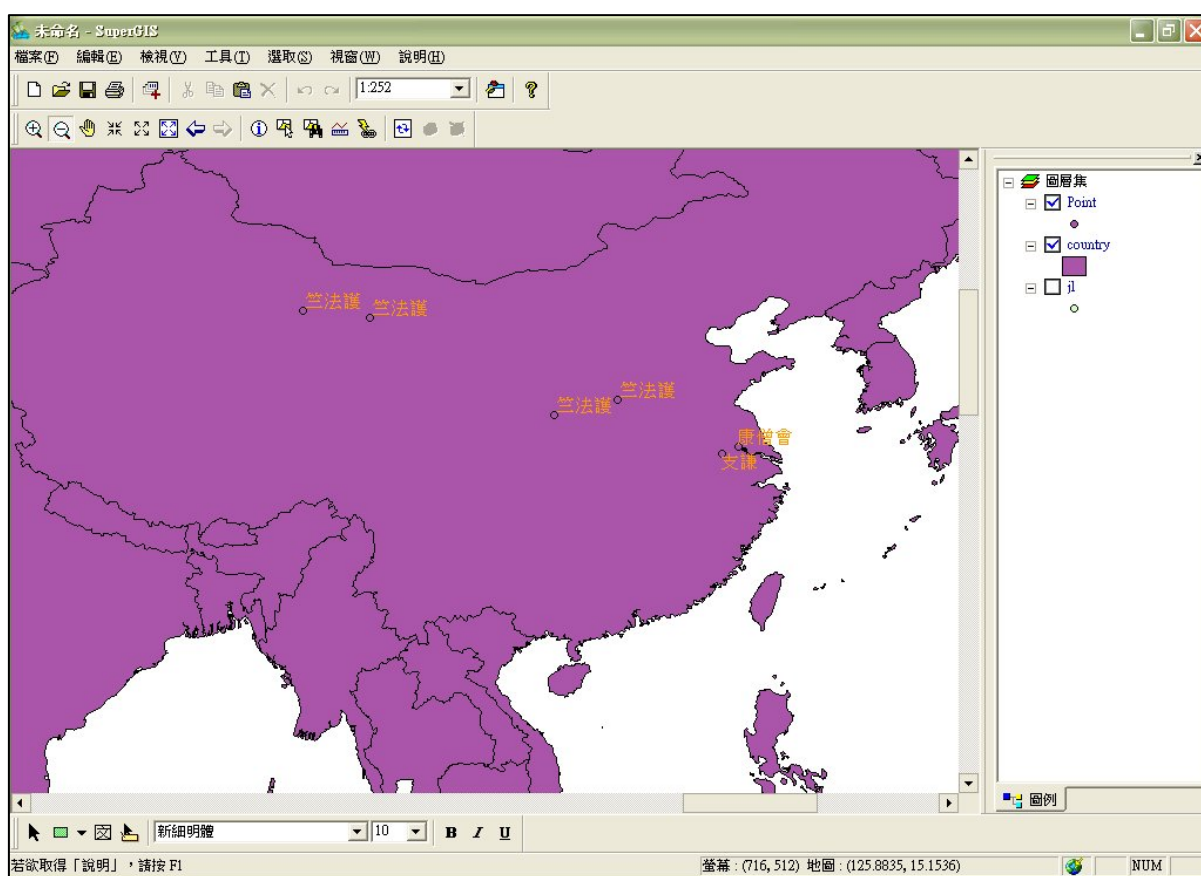
針對書前目錄與經書內文進行比對，由於書前目錄可能有誤，仍需進一步與內文比對確認，比對後之差異檔，交由兩位熟識經文之經驗人員各自利用「看圖校對程式」，以差異檔比照原書掃描圖檔予以訂正。

該兩位人員訂正後交回的兩份校對完成檔，再以程式執行第二次檔案比對，比對後會產生一數量較小之差異檔。將此差異檔交由一位人員進行最後把關，方

法也是以差異檔比照原書掃描圖檔看圖校對。

六、建置經錄 GIS 後設資料庫

建置時空地理資訊檢索系統。方便讀者從地圖，就能知道譯經者的地點，在加上時間軸，整體呈現就更加立體化。但由於經錄的作譯者有些同名，同名的作譯者，需查閱相關的工具書，協助判斷是否為同一人，或是另有其人，相當耗時。此系統的時間與空間軸呈現方式，目前尚在測試開發中。



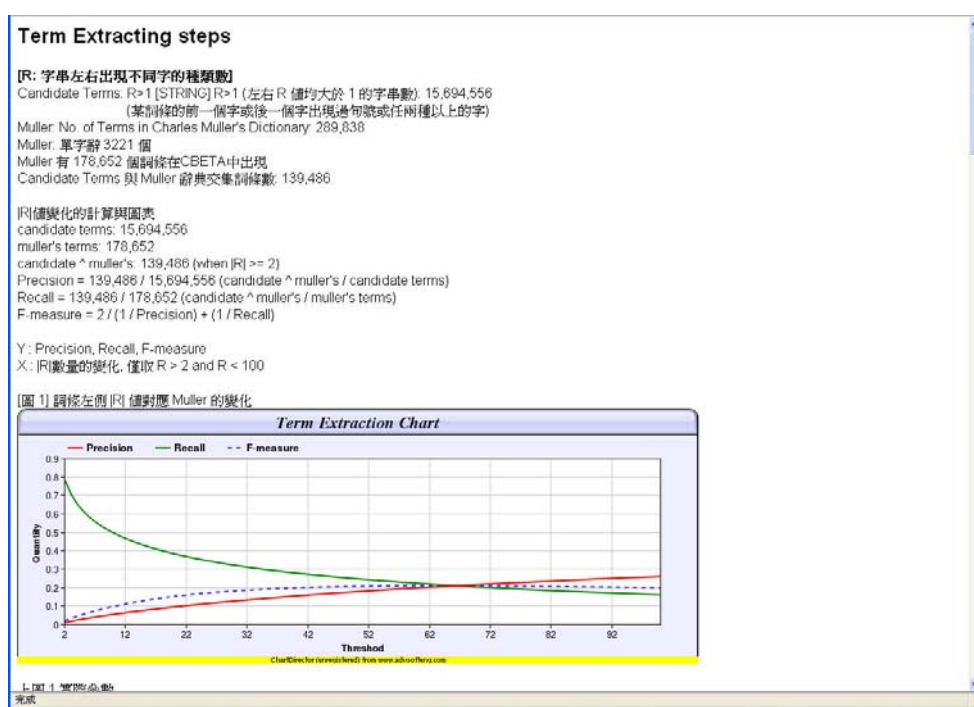
時空地理資訊檢索系統。

七、經錄抽辭技術平台

工作小組中的資訊人員與文獻處理人員密切配合，研發出一套方便佛教學者建立知識架構的環境與研究平台 --- 「抽詞」技術平台，以時間資料庫作量化的排序。

「抽詞」技術平台的開發，使佛教學者可以透過它對龐大的數位佛教資料庫，進行統計分析 (Statistical Analysis)、資訊檢索及抽取 (Information Retrieval and Extraction)、文件分類與分群 (Document Classification and Clustering)、資料

探勘 (Data Mining) 等各項工作，此平台提供研究者不同於傳統佛學的研究方法及更多樣的參考資源和結果。

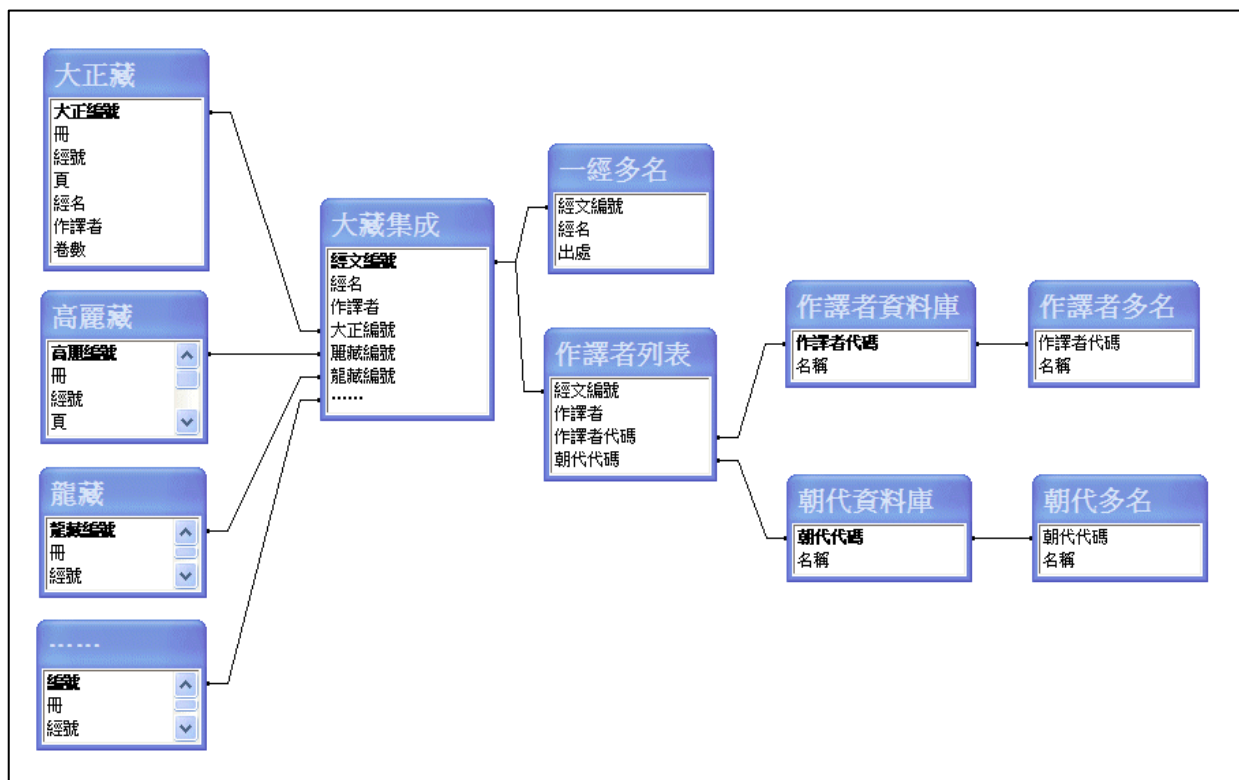


「抽詞」技術資料分析。¹

八、整合型經錄資料庫之建構

經錄資料庫是指從單一版本經錄資料庫，到建構整合多語言經錄資料庫。佛典經錄所涉獵文獻龐大，從紙本至不同刻本，一部佛經在不同的版本中就有不同的出處。因此除了整理每版數千筆經錄的各版藏經為一資料庫外，工作小組將多年來開發建置的經錄資料庫作精緻處理、作串連，並建立各版藏經的相關網絡關係，集結成一個總資料庫，記錄唯一經號、經名、譯者、卷數的總資料庫，並記錄與各藏各經的關聯及交叉比對等等，使讀者從佛經目錄資料庫中檢索到某部經在各版藏經的版本記錄訊息，若各經錄資料全數開發建置完成後，將會是全世界獨一無二，完整的佛經經錄數位資料庫。

¹ <http://211.23.77.92/BuddhistTermExtract/newMuller.php>。



資料庫關連結構圖。

九、建置檢索系統及介面

依各藏經的編纂語言分類，個別設計檢索介面，提供多語言、多版本比對、多檢索方式的強大功能查詢介面，更方便使用者作查詢檢索。

十、成品及網站服務

工作小組將數位化完成的全文資料與影像資料進行整合，設計使用者檢索介面，轉換成普及網路版放置網路上，供學術研究者或社會大眾能快速從經錄網站找到經籍的出處與藏經編輯流傳的關係。結合現代科技網路與資料庫的技術，使傳統紙本佛典經錄的功能，大大躍昇。



The screenshot displays the website's search interface. At the top, a banner reads "佛教藏經目錄數位資料庫" (Digital Database of Buddhist Tripitaka Catalogues) with the English translation below it. A navigation bar contains links for "首頁" (Home), "漢文" (Chinese), "尼科耶阿含經錄" (Nikaya Ahimsa Catalogue), "梵文" (Sanskrit), "藏文" (Tibetan), "滿文" (Manchu), "西文譯經" (Western Translations), "GIS時空地理" (GIS Time-Space Geography), and "相關知識" (Related Knowledge). On the left, a tree view titled "古經錄分類" (Ancient Catalogue Classification) lists various categories such as "出三藏記集", "眾經目錄 (法經)", "歷代三寶記", and "開元釋教錄". On the right, a search box titled "古經錄資料庫檢索" (Ancient Catalogue Database Search) features input fields for "經名" (Sutra Name), "朝代" (Dynasty), and "作譯者" (Translator/Compiler), along with "開始查詢" (Start Search) and "清除" (Clear) buttons.

佛教藏經目錄數位資料庫檢索畫面。

※ **製作單位：**數位典藏與數位學習國家型科技計畫

拓展臺灣數位典藏計畫 數位內容建置與整合子計畫

法鼓佛教研修學院 「佛教藏經目錄數位資料庫」研究計畫

※ **文字撰寫：**拓展臺灣數位典藏計畫 數位內容建置與整合子計畫

— 文獻與檔案主題小組助理 陳美智

法鼓佛教研修學院 「佛教藏經目錄數位資料庫」研究計畫

— 林心雁、鄭寶蓮

※ **圖片拍攝：**拓展臺灣數位典藏計畫 數位內容建置與整合子計畫

— 文獻與檔案主題小組助理 陳美智、林芳志

法鼓佛教研修學院 「佛教藏經目錄數位資料庫」研究計畫

— 鄭寶蓮

※ **圖片提供：**法鼓佛教研修學院 「佛教藏經目錄數位資料庫」研究計畫

※ **圖文編輯：**拓展臺灣數位典藏計畫 數位內容建置與整合子計畫

— 文獻與檔案主題小組助理 陳美智

致謝：

感謝法鼓佛教研修學院 「佛教藏經目錄數位資料庫」研究計畫主持人杜正民老師、林心雁與鄭寶蓮小姐，撥冗指導及提供實地拍攝與簡介編寫。並感謝法鼓佛教研修學院其餘相關人員之協助。

程序	工作內容	操作人員（數量、專業能力之要求）	硬體（名稱、版本、價格）	軟體（名稱、版本、價格等）	依循標準（技術規範、成品規格、品質要求…等）	耗時	總結（困難、缺失、特色…等）	成本估算	備註
6	將《滿文大藏經》紙本目錄翻拍	委外進行	PC 及超高解析度照相機	Photoshop 等影像編輯軟體	為提供使用者比對查詢。		有些經目原本印刷字跡就不太清楚，需翻拍多次，才能達到令人滿意的效果。		
7	建置《滿文大藏經》影像檔資料庫	工作小組	PC	Microsoft Windows XP、看圖軟體	照原書目錄的函號，依序建立圖檔資料庫，並進行校對。		基礎資料庫建檔工作，進行順利。		
8	《滿文大藏經》經錄等基本資料的輸入繕打與建檔	工作小組	PC	Microsoft Windows XP、Oxygen、PERL、WinCvs	建立欄位及內容統一格式		基礎資料庫建檔工作，進行順利。唯獨有些經目原本印刷字跡就不甚清楚，需查閱其他的工具書，協助判讀。		
9	建置經錄 GIS 後設資料庫	工作小組	PC	Microsoft Windows XP、Microsoft Excel、Access 或其他文書編輯軟體	建立時空地理資訊檢索系統。		同名的作譯者，需查閱相關的工具書，協助判斷是否為同一人，或是另有其人，相當耗時。		
10	經錄與藏經全文的量化統計	資訊人員	PC	Oxygen、PERL、PHP、Java、Python	建構知識管理模式。		資訊人員與文獻處理人員需密切配合。		
11	古經錄全文 XML 標記	工作小組	PC	UltraEdit、PERL、WinCvs、SP、Python、Microsoft Office	Big5, CP950, Unicode, XML, TEI. 利用程式將簡單標記轉為 XML 標記。		以符合國際標準的 XML 語言建立經文資料庫。		

程序	工作內容	操作人員（數量、專業能力之要求）	硬體（名稱、版本、價格）	軟體（名稱、版本、價格等）	依循標準（技術規範、成品規格、品質要求…等）	耗時	總結（困難、缺失、特色…等）	成本估算	備註
12	所有經錄資料庫總集成	資訊人員	PC	Java、PERL、MySQL、PHP、WinCvs等	記錄唯一經號、經名、譯者、卷數的總資料庫，並記錄與各藏各經的關聯及交叉比對等等。		將會是全世界獨一無二，完整的佛經經錄數位資料庫。		
13	建置藏經目錄檢索系統及使用者介面	資訊人員	PC	Java、PERL、MySQL、PHP、WinCvs等	依藏經編纂語言分類，個別設計檢索介面。		提供多語言、多版本比對、多檢索方式的強大功能查詢介面。		
14	成品及網路服務	資訊人員	PC	Java、PERL、MySQL、WinCvs等	設計使用者檢索介面，提供經錄讀者索引服務等。		學術研究者或社會大眾能快速從經錄網站找到經籍的出處與藏經編輯流傳的關係。		

註：若程序多於七個，請複製本表使用

調查人：林心雁

調查地點：法鼓佛教研修學院

調查日期：2007/12/28