

佛學研究資源數位化作業標準與規範

Standards of the Digitalized Buddhist Research Resource Project

杜正民 (Tu A ming) *

一、前言：佛教文獻數位化的意義與基準

當代的電腦網路、資訊科技與數位「載體」(carrier)或「媒介、媒體」(media)，已開始將人類的知識重新以不同的形式呈現。因為載體不同，所以人類對知識取得的習慣也跟著改變，如方便迅速的檢索工具與大型的資料庫，改變了學者的研究方法與速度，但是其背後的數位化作業如何進行，甚至如何深入運用此資料，則較不為多數學者所知道。

譬如，數位文獻資料的標記規範與應用的標準，也就是把人類想要表現的文獻資訊標註給電腦解讀，這種標誌記錄就如古人以句讀的方式，標記於紙本文獻上，或以眉批的方式作內容的摘錄、或以科判的方式作大綱

與全文掌控、甚或以校註校勘等方式作版本的比對與解讀等。這些表現的方式，都需依電腦資訊的特殊方式來做表達。而這些方法，目前多以標記的規範與標準，來對數位化的內容進行標記作業。¹

進而，如常用到的Metadata (含DC, TEI等)，²可以說都是為配合新科技的工具與運用因應而生的。再如，如何運用GIS電子地圖的概念來建構時間與空間的架構，以完整的表達知識結構，利用電子文化地圖從事古知識新資訊的作業 (Spatial Analysis of Historical Texts) 等，且配合此作業，需要有技術的提供與研發等配合，也是目前大部分數位化作業的要項。³

為配合網際網路的資訊時代來臨，人們溝通和處理資料與知識方式的改變，所以佛典的數位化 (Digitalized) 或電子佛典 (Electronic Buddhist Text) 的需求急遽增加，因此有各種佛典的數位化輸入計畫。⁴譬如，中

* 作者為中華佛學研究所圖資館館長。

1 Aming Tu, "Buddhist Studies and Information Literacy" (佛學研究與資訊素養), 「佛教與當代社會學術研討會」(2003.12.5-7), 香港: 香港大學佛學中心。文中提到佛學研究方法與資訊素養——工具、標記、標準等的重要性。

2 Metadata 被稱為「後設資料」，主要是描述資料屬性的資訊，用來支持如指示儲存位置、資源尋找、文件紀錄、評價、過濾等的功能。為有效組織整理網路資源，各研究單位也因目的不同發展出不同領域的 Metadata，以因應電子資源的整理。DC (Dublin Core 都柏林核心集) 與 TEI (Text Encoding Interchange 標記集) 為目前常用的標準。DC 創始於 1995 年，由圖書館、電腦、網路方面的學者和專家在 OCLC 及 NCSA 所聯合贊助的研究會中共同討論而產生的，目的是要讓資源的創造者或出版者等非專業人員能自行建立資源的 Metadata，以利資訊檢索……。TEI 文獻編碼計畫為國際性的計畫，以促進資源的交換為目的，發展一套指導方針，做為電子文獻交換及再利用的準備。規定每篇文獻之前，均應放置所有描述該文獻 (包括來源、版本等) 的標示文字 (即文獻的 Metadata)。

3 譬如「數位典藏國家型科技計畫」於計畫辦公室下設五個分項計畫，分別為：內容發展、技術研發、應用服務、訓練推廣及維運管理分項計畫，協助計畫辦公室相關業務的推動。其中以「內容發展分項計畫」負責本國家型計畫數位典藏內容資料之調查、建置及發展相關事宜。而由「技術研發分項計畫」負責本國家型計畫數位化流程與規範制定、技術研發，及支援各計畫之技術需求。詳細內容請參閱 http://www.ndap.org.tw/1_intro/structure.php。

4 早期以佛典的電子化為訴求，以電子佛典為名。而近時，則因數位載體的關係，所以又名為數位化，本文因為沿用與貫通前後期的作業關係，所以電子佛典及數位化作業混用。

華電子佛典協會 (Chinese Buddhist Electronic Text Association, 簡稱 CBETA, <http://www.cbeta.org>) 採用目前廣為學術界使用的日本大正時代開始編輯出版的藏經《大正新脩大藏經》(簡稱《大正藏》) 為底本, 進行數位化的作業。《大正藏》是以高麗本為底本, 對校宋、元、明三本, 另參照正倉院藏經、敦煌古本及巴利文、梵文經典, 並在校勘欄中記錄了各版本的不同用字等資訊。CBETA 在製作電子佛典的過程中, 將這些校勘資訊以 XML (eXtensible Markup Language 可擴展標記語言) 記錄, 並以 HTML 方式呈現, 藉由校勘資訊做部分的版本還原, 讓使用者可以選擇瀏覽不同版本。此作業過程及其呈現方式或許可作為數位化古籍校勘版本處理技術的參考, 是以下文由佛典電子化的定義⁵與時代意義談起。

(一) 佛典電子化的定義與時代意義

近年來, 由於大眾的努力, 網路上結集了不少佛教經典, 及各式各樣的佛學資源與佛學資料庫, 希望能透過網路傳播佛法, 並利用電腦功能拓展佛典的應用範圍及閱讀方式, 讓更多人同霑法益。因此, 對於「佛典電子化」一詞, 本文擬採廣義的解釋, 亦即包含佛教藏經的電子化(或數位化)、當代佛教文獻數位化與佛學網路資源等。

所謂佛學研究資源數位化, 其廣泛的定義為將所有的佛學資料數位化, 自然是涵蓋古今佛教文獻資料, 除當代佛學研究文獻外, 未入藏的佛典以及歷代的大藏經, 都是包含於佛學研究資源數位化的作業範圍, 其中以大藏經的數量最大, 涵蓋面亦較廣, 故本段先自大藏經的編製, 及其標準與規範作一說明。

佛教典籍或名藏經、一切經、契經、三藏等, 包括以釋迦牟尼教說解脫道的方法被紀錄為「經」; 為導向解脫因此而制定須遵守的規則為「律」; 後世對經、律等關於佛教教理與方法的闡述解說稱之為「論」。釋迦牟尼般涅槃後, 弟子們將佛所教說的內容結集匯整, 初期以口傳的方式傳播, 之後以文字記載的方式抄寫彙整

流傳, 也開始佛典大量的傳播。

漢譯佛典的傳譯, 自後漢迄元代已經大部分完成。總稱佛典為「一切眾藏經典」、「一切經藏」、「大藏經」, 但是流通皆賴書寫。直至宋開寶四年(971)始有木版印刷的《開寶藏》傳世, 並頒賜國內各地, 以及日本、契丹、西夏、高麗諸國。此後有遼版之《契丹藏》、《金藏》、《萬壽藏》、《昆盧藏》、《圓覺藏》、《資福藏》、《磧砂藏》等宋朝版本, 以及韓國的《高麗藏》; 元代有《普寧藏》、《弘法藏》等; 明朝刊刻《南藏》、《北藏》等。

簡單的說, 佛典傳入中國後, 經過長時間的傳譯工作、目錄編撰、抄寫經典、石雕刻經、木刻印刷及至近代的排版印刷與現行的電子出版等, 其發展與改變過程中都有規範與標準可資參考。因此, 先就經典發展歷史, 以個人的見解做一知識流程如下:

經典的翻譯 → 經典的理解 → 經典的解說
→ 經典的整理 → 經典的保存 → 經典的研究
→ 經典的發展

參照上述經典發展的流程, 於現今佛典數位化的作業, 依當代的方法與工具, 又可分為內容、標準、技術、研發、推廣、流通、管理等方面來考量。

(二) 佛教文獻數位化基本概念與基準

上述電子化或數位化的內容因為標準、技術、研發、推廣、流通、管理等方式與早期的抄寫、木刻、石雕刻或近代的排版印刷等傳播都有很大的不同, 而有了相當大的改變。譬如, 以抄寫經典為例, 中國自南北朝起已經將諸經歸納為《一切經》, 南齊明帝令人抄寫《一切經》, 陳武帝和宣帝命人抄寫 12 部, 文帝命人抄寫 50 部, 放在諸州各大寺廟裡。北齊孝明帝也命人抄寫 12 部《一切經》, 隋文帝命人抄寫 46 部《一切經》放在各州。可知早期以王室欽定下令書寫給各州的《一切經》, 其數量最多也只可至 50 部, 因為是書寫的關係

5 有關資訊的定義等資料, 可參考謝清俊, 〈對『資訊科技之於學術研究』的幾點看法〉, 「第三屆國際漢學會議」(2000.6.29-7.1), 《漢學研究通訊》第 20 卷第 2 期(總第 78 期), 2001, 頁 58-67。網站相關資料 http://www.sinica.edu.tw/dp/paper/2000/20000629_1.htm, 其中有許多對資訊定義的精闢觀點。

並無法廣為流通。

以數量來說，由早期的抄經的有限數量，至現在幾近無限量的傳播；以時間而言，傳播的時間可說是即時與隨時可自網路取得；進而以空間來說，也可說是無疆界。所以當代的傳播流通速度，已不同於早年的時空限制，可說已是無限制。因此，可知電子時代的數位載體，可以借其無形的能源，突破之前有形的傳播媒體。其影響與重要性可預期，故於本段介紹數位化的基本概念與基準。

佛典數位化作業與一般文獻數位化作業類似，以電子佛典的經驗而言，則因作業有其特殊的意義，是以就佛典數位化作業簡述於下：如佛教藏經分為經、律、論三藏般，同樣的，於佛典數位化作業的三大部分，與傳統的經律論功能極為類似，也就是就「內容」而言，進行佛教文獻數位化的主體與內容，與匯集導向解脫教說的「經」都是同為主軸。就「標準」而言，為能讓數位化的內容得以發揮最大的功效，如資源共享、後設資料的紀錄等，佛學文獻數位化的規範與標準，與以實際操作的內容有所遵循，而設立規範之「律」一樣的重要。進而，就「技術研發」而言，依規範標準完成數位主體的內容，則需要好的技術、程式、研發與資訊開發等作業來配合，所以佛學文獻數位化的技術與研發，與為讓「經」及「律」發揮最大效用的「論」有同樣的效能。

如下所示，數位化的內容是廣義的「經」，它是數位化作業的主軸；而規範與標準則是「律」，因為它是數位化作業之所依；而為讓這些內容與標準能做更好的呈現，就需要有如同讓經、律發揮最大功能的「論」一般的技術與研發。此三大項目可稱之為佛典數位化作業的三基準。至於，其他如推廣、應用、服務等活動，或可像法會般，於適當時機作推廣活動與教育訓練。

經——佛陀導向解脫的教說為主軸 = 佛教文獻數位化的主體與內容
律——實際操作經內容所應遵循的規範 = 佛學文獻數位化的規範與標準
論——讓經與律能發揮最大效用的方法 = 佛學文獻數位化的技術與研發

是以，如前述所論佛典數位化作業的三大基準，本

文將以佛教大藏經與佛典經錄為內容，討論數位化作業的標準與規範之重要性，最後則以佛典數位化的技術為進一步研發的考量。

（三）簡述佛學研究數位資源現況

自 1994 年臺大佛學研究中心建構「佛學網路資料庫」至今，已完成「佛學書目資料庫」、「佛學全文資料庫」等二十餘項佛學網路資料庫，這些資源以蒐集與組織當代文獻資料為主。隨即自 1998 年起，開始佛教原典文獻的數位化作業。CBETA 於 2000 年底完成《大正藏》電子化基本作業，同時建立電子資料庫各項「標準化」作業，奠定漢文佛典電子文獻製作規範。2001 年至 2004 年間，除前述當代文獻的蒐集與原典文獻的電子化作業繼續更深入開展外，也開始佛教文物資料數位化作業，如國科會「數位博物館計畫：玄奘西域行」、國科會國家數位典藏「臺灣佛教數位博物館：蓬萊淨土遊」與國科會「《瑜伽師地論》數位資料庫的建構與研究」和蔣經國國際學術交流基金會「臺灣佛教文獻數位資料庫的建構與研究」、「《法華經》多種語文版本數位資料庫的建構與研究」等計畫；還有「漢文電子佛典資料庫」、「佛教文獻數位資料庫」及「臺灣佛教數位資料庫」等涵蓋古今的佛典數位資料庫建置完成，可說已經累積有相當的資料與經驗。因此下文擬將就臺灣此豐富的佛典文獻數位化經驗，以及佛學數位典藏內容及標準與研發等資料作一簡介。

二、經——佛教文獻數位化的主體與內容

（一）佛典編藏的標準與規範

1. 依千字文編號，便於查閱與管理

當今，漢文所保存的古印度佛籍，為數最多，而且也最為完備。流傳的藏經大致以《開元錄》入藏目錄為基礎有所增刪。除明《嘉興藏》、《永樂北藏》、《永樂南藏》外，增加中國撰述 400 餘種，保存了大量的中國著作。它們大都作有目錄，並依千字文編號，是一便於查閱的目錄。採千字文編號是藏經管理的良好標準與規範。

2. 依函列卷次、逐卷規範，提綱挈領

大藏經數量浩瀚，為能全體掌握，於是就須有簡明扼要的解題目錄。如宋仁宗天聖二年（1024）遵式撰《教藏隨函目錄》，敘述諸部著作大義；宋徽宗崇寧三年（1104）惟白依《開元錄》入藏次第撰《大藏綱目指要錄》，逐函分列卷次，逐卷略錄義例，鉤玄提要，使讀者能瞭解宏綱；崇寧四年王古撰《大藏聖教法寶標目》；明神宗萬曆四十一年（1613）寂曉撰《大明釋教彙目義門》，卷首的〈釋例〉，即分部的說明。此書按天台五時判教，依明南北藏，先後更置，分為8部，每經之下寫示經文大意及論疏旨趣，漢人釋經的注疏，即附於同部之後，有利於查閱。萬曆四十六年作者將上書節為簡本，名為《大明釋教彙門標目》4卷，每書之下，以一二語概括介紹，頗便檢閱。訂定標準，依函列出卷次，再依卷說明大意，提綱挈領，規範清楚。

3. 入藏次第清楚、分類條列，組織嚴密

南明桂王永曆八年（1654）智旭撰《閱藏知津》，總目4卷，本文44卷，分經、律、論、雜藏四大類。經藏分大乘小乘，大乘中依天台判五時教次第而略有改易，分為華嚴、方等、般若、法華、涅槃五部，方等中又分顯說與密咒二門，密咒包拓秘密部諸經及儀軌。律藏也分大小乘，在小乘律之下，附載疑似雜偽的律。論藏分大小乘，大乘論中，又分釋經論、宗經論、諸論釋諸部，下又別為西土、此土二科。雜藏分西土撰述、此方撰述。西土撰述中附外道論、疑偽經。此方撰述分儀儀、淨土宗、台宗、禪宗、賢首宗、慈恩宗、密宗、律宗、纂集、傳記、護教、音義、目錄、序贊詩歌及應收入此土撰述15部。其分類條分縷析，組織嚴密完整。⁶依入藏次第清楚編排，再依此規範，做為分類條列標準，使整部藏經得以有系統的組織。

由上述資料可知自古以來佛典的編輯有相當精緻的規範與標準的設置，不論是經文大意的描述、論疏旨趣的建置、釋經的注疏的收編等皆能以經錄做概括介紹，組織嚴密完整，入藏次第清楚，逐函分列卷次，逐卷略錄義例，使讀者能瞭解宏綱，並以千字文編號管理等，

依此規範做為分類條列的標準，利於查閱，頗便檢索，此皆為古德規範清楚的利益。

（二）CBETA 數位佛典與傳統藏經的關係

佛典藏經的發行與漢譯的工作不斷，正是促使佛教在中國知識普及化與廣為流傳的因素之一。20世紀末的資訊時代，因為網際網路改變了人們溝通和處理資料與知識的方式，所以佛典的數位化需求急劇增加，因此有各種佛典的數位化輸入計畫，而CBETA的工作即是其中一項重要的作業。以下將簡述其與紙型印刷藏經的關係，以及如何進行當代的規範與標準。

CBETA 電子佛典初期以《大正藏》與《卍續藏》為底本。《大正藏》全稱《大正新脩大藏經》，是在日本大正13年（1924）由高楠順次郎和渡邊海旭發起，組織大正一切經刊行會，小野玄妙等人負責編輯校勘，1934年完成印行。全藏分3個部分：正藏55冊，續藏30冊，別卷15冊（內圖像12冊，總目錄3冊），共100冊。收入佛籍總數冠于各種大藏經，計正藏、續藏、圖像和總目錄共3493部，13520卷。編輯體例別創一格，經藏分為阿含、本緣、般若、法華、華嚴、寶積、涅槃、大集、經集、密教十類；律藏包括彌沙塞部、摩訶僧祇部、曇無德部、薩婆多部、解脫戒經（迦葉遺部）和菩薩戒等；論藏分為釋經論、毗曇、中觀、瑜伽、論集等五類；雜藏分為經疏、律疏、論疏、諸宗、史傳、事彙、外教和目錄等八類。續藏30冊中除最末一冊為古逸和疑似兩類，收敦煌古逸經、律、論疏和疑偽經外，其餘29卷收入日本著述的續經疏、律疏、論疏、諸宗四類及悉曇。圖像12冊收入日本各寺院所藏歷代有名佛教畫像和密宗明王像、金剛像及各種曼荼羅圖等363種。總目3冊收入中國歷代各版藏經目錄和日本各寺院所藏的寫本和刻本藏經目錄以及大正藏勘同錄、大正藏總目錄、總索引、譯著目錄等77種，《大正藏》為學術界及民間流通最廣的漢文大藏經。

《卍續藏經》為日本藏經書院於完成《卍正藏經》的編印後，前田慧雲主持，中野達慧負責，廣泛搜集中國和日本歷代未入藏的佛教典籍彙編成書，又名《續

6 相關全文請參照蘇晉仁，〈中國佛教經錄〉（摘錄自《佛教數字圖書館》<http://210.52.11.162/dzjji/zhongguofojiaojinglu.htm>）。或參考徐建華，〈中國歷代佛教目錄類型瑣議〉，《佛教圖書館訊》第29期（2002.3）等有關中國歷代佛教目錄資料。

藏》或《藏經書院續藏經》，內容上至六朝遺編和唐宋章疏，下迄清代著述之缺佚。在編纂過程中，曾得到金陵刻經處楊文會、蘆山寺式定和各寺院居士們的支持，在1905-1912年印行。內容包括950餘人的譯著1659部7143卷（總目5卷），分為十門六十三類共三編151函，每函5冊（包括目錄1函1冊），24開線裝本，4號鉛字排印，版式同《卍正藏經》。但印就後，存書和《卍正藏經》一同被焚，流傳不多，甚為可惜。⁷之後有臺灣的影印本及日本再版的三欄本等之發行。

也就是說《卍續藏經》目前有京都藏經書院的《卍大日本續藏經》、臺北新文豐影印的《卍續藏經》、以及日本東京國書刊行會再版的三欄《卍新纂大日本續藏經》，共三種版本發行流通。但其版式各有不同，如京都藏經書院的《卍大日本續藏經》採用「編、套、冊」格式，計三編，合150套，共750冊。⁸臺灣新文豐影印本，以《卍大日本續藏經》為底本，改以150冊雙欄發行，每一冊對應到舊編的一套，每頁以阿拉伯數字編號，較易檢閱。東京國書刊行會的再版三欄《卍新纂大日本續藏經》係由《大日本續藏經》、《日本校訂大藏經》中國撰述部，及若干新增補的典籍所成。經文有88冊，加上總目錄、索引、解題，共計90冊。版面一頁三欄，臺灣有白馬影印本流通。CBETA電子經文資料庫，是以新版三欄《卍新纂大日本續藏經》為底本。

自1998迄今，CBETA積極建置佛典文獻數位資料，以網路及光碟發行。今後，CBETA電子佛典集成除《大正藏》與《卍續藏》外，並擬廣收未入前述二藏的佛典典籍，期能配合新的數位媒體發行一套古來較完整的藏經集成，以達到流傳普及、方便使用的功能，成為當代及未來必備的佛學資料。

三、律——佛教文獻數位化的規範與標準

（一）古代「經錄」與規範

由上文可知佛典在中土的發展與重要性，但與數位化的標準與規範最有關的部分，應可由佛教的經錄著手，故下文擬從佛典現存經錄瞭解自漢魏以來佛教經文的發展與現況，由漢文佛教藏經目錄，介紹臺灣建構中的佛典電子化現況。

1. 佛教目錄學在中國目錄學之位置

梁啟超於〈佛教目錄學在中國目錄學之位置〉中所說：「佛學諸家經錄……，覺其所用方法，有優勝於普通目錄之書數事：一曰歷史觀念甚發達，凡一書之傳譯淵源，譯人小傳、譯時、譯地、靡不詳述；二曰辨別真偽極嚴……；三曰比較甚審……；四曰蒐采遺逸甚勤……；五曰分類極複雜而周備……。」⁹由此文不僅可見佛教目錄學的特點與重要，也可看出其對中國目錄學的影響。

雖然佛教經錄在中國目錄學有其獨特的位置，然由於種種原因，撰釋家往往無法掌握佛家經錄，造成疏漏等現象。如陳垣於《中國佛教史籍概論》所言：

尤所注意者，《四庫》著錄及存目之書，因《四庫提要》於學術上有高名，而成書倉猝，紕繆百出，易播其誤於眾。如著錄……《開元釋教錄》而不著錄《出三藏記集》及《歷代三寶記》，猶之載《唐書》〈經籍志〉而不載《漢志》及《隋志》也。其弊蓋由於撰釋家類提要時，非按目求書，而惟因書著目，故疏漏至此。¹⁰

7 相關資料及全文請參照童瑋，〈漢文大藏經〉（摘錄自《中國大百科智慧藏》<http://library.mit.edu.tw/web/Content.asp?ID=62502>）。或參考李富華、何梅著，《漢文佛教大藏經研究》（北京：宗教文化出版社，2003）等有關漢文藏經研究資料。

8 第一編：1-95套、第二編：1-32套、第二編乙：1-23套，而其裝訂方法為：每5冊合為一函，每函約500-550葉（一葉為兩面/頁），一葉有4版（分別為右上、右下、左上及左下四版，或稱第一至第四欄）。

9 梁啟超，〈佛教目錄學在中國目錄學之位置〉，《梁啟超集》（北京：中國社會科學出版社，1995年），頁100，（《飲冰室專集》67）；或《現代佛教學術叢刊》第40冊（臺北：大乘文化，1980年），頁21-52。

10 陳垣，《中國佛教史籍概論》（臺北：新文豐版，1983年），頁1。

由此可知佛教目錄學之殊勝與難以掌握。是以，下文擬從經錄開始介紹，以期對佛教經錄及經文有一概括的瞭解。

2. 經錄的編輯

中國佛典綜合目錄——「經錄」的編纂，在中國目錄學史上，為重要的貢獻。僅就目前現存資料而言，始自梁僧祐《出三藏記集》中所包含的東晉道安的《綜理眾經目錄》及至隋代編的各種經錄，唐初道宣所編的《大唐內典錄》，唐中葉圓照編成的《貞元釋教錄》為止，共有十數種經錄。¹¹

「經錄」道安稱為「眾經目錄」，隋代費長房編纂的《歷代三寶紀》，是由經錄與佛教年表合併而成，乃未用「經錄」之名，但三寶紀以外的隋代諸經錄，則使用了「眾經目錄」的名稱。唐代的道宣，以儒教道教之典籍作為外典的佛教立場出發，編成《大唐內典錄》。智昇則編纂《開元釋教錄》，不拘於經名，改為綜合的佛教書籍目錄的名稱。開元十八年（730）由長安西福寺智昇（668-740）所編纂的《開元釋教錄》20卷，可說是道安綜理眾經目錄以來的中國經錄之決定版。

編纂經錄的目的，主要是將流傳到中國的梵語及其他外語原典的既譯經典，不管其現存與否，盡知所能地蒐集、登錄，並將譯者名、譯出年代明確者，按照大小乘、經律論的次序整理，譯者不明者即作為失譯，失譯經中若確知其搜得之地區者，乃依地區別予以編集。¹²

3. 經錄的分類與規範

梁啟超對佛教經錄的規範，可簡約為「分類、真偽」及「譯經年代、譯者」兩大類：

經錄之學，至隋而殆已大成，綜其流別，可分兩派：其一，專注重分類及真偽，自僧佑，李廡以下皆是，

至隋《法經》集其成……。其二，專注重年代及譯人，竺道祖以下凡以朝代冠錄名者皆是，至隋費長房集大成。¹³

佛經目錄自東晉道安《綜理眾經目錄》，經隋代諸經錄至唐初道宣參考歷代經錄編製《大唐內典錄》，並加以合理的分類與組織。如道宣於序言所說：「今總會群作以類區分，合成一部開為十例，依條顯列無相奪倫。」經由道宣《大唐內典錄》及智昇《開元釋教錄》¹⁴這些資料可知，古經錄已於書目、卷數、譯出年月、譯場、譯者以及存缺等情形皆已有相當的規範，這些規範即成為後代匯集整套大藏經之所依。

4. 經錄與藏經

如前述中國第一部刻本大藏經是宋代的《開寶藏》，此藏刻成之後121年，刊刻於福州的《崇寧藏》又告完成。而《崇寧藏》之後，迄清代的《龍藏》為止，中國各朝代所刊刻的大藏經，在印度文獻方面，仍陸續有零星的增補。內容包含新發現的同本異譯、同一譯典的不同版本以及宋元新譯等佛典。

雖然歷代藏經版繁多，但相關經錄可從下列四部文獻尋得：《大正藏》第55冊、《昭和法寶總目錄》3冊、蔡運辰編《二十五種藏經目錄對照考釋》（臺北：中華佛教文化館，1983）、童瑋編《二十二種大藏經通檢》（北京：中華書局，1997）。除《大正藏》第55冊及《昭和法寶總目錄》3冊含有大量的目錄外，《二十五種藏經目錄對照考釋》依據《昭和法寶總目錄》所收，酌加增補，且製成表格，便於查索。此書共包含《開寶》、《崇寧》、《昆盧》、《圓覺》、《資福》、《磧砂》（甲乙二版）、《普寧》、《金藏》、《南藏》、《北藏》、《嘉興》、《龍藏》、《麗藏》、

11 現存經錄如下：（錄自《大正藏》第55冊）(1)東晉道安《綜理眾經目錄》（《出三藏記集》卷5）；(2)梁僧祐《出三藏記集》卷4、卷5；(3)隋費長房《歷代三寶紀》卷15；(4)隋法經等《眾經目錄》卷7；(5)隋彥棕等《眾經目錄》卷1；(6)唐靜泰等《眾經目錄》卷1；(7)唐道宣《大唐內典錄》卷10；(8)唐明佺等《大周刊定眾經目錄》卷15；(9)唐智昇《開元釋教錄》卷18；(10)唐圓照《貞元釋教錄》卷28。

12 摘錄自牧田諦亮著，楊白衣譯，〈疑經研究——中國佛教中之真經與疑經〉，《華岡佛學學報》第4期，頁284-306。（譯自牧田諦亮著，《疑經研究》第一章「中國佛教之疑經研究二·中國佛教中之真經與疑經」）

13 同註9。

14 智昇編纂的《開元釋教錄》計「合二十卷，開為總別。總錄括聚群經；別錄分其乘藏。二錄各成十卷，就別更有七門。」

《縮刻》、《頻伽》、《卍字》等十餘種刊本大藏經目錄及若干種佛書目錄所含的佛典書名。《二十二種大藏經通檢》為大陸編《中華藏》時所做，對前書有所增補修訂，亦為一值得參考的經錄工具書，然上述兩書內容簡潔，無法建立流傳與版本等關係。

總之，很多佛教重要文獻，都是由於《大正》、《卍續》二藏的問世，才得以重見天日，以補充現存宋清間諸部大藏經所收的中印佛典的不足。這也是《大正》、《卍續》二藏最主要的一項價值。¹⁵是以CBETA電子佛典，即以此兩套大藏經為基礎進行數位化作業，以下簡介當代的規範與標準如何應用於此古文獻。

（二）當代 Metadata 於佛教數位文獻的運用

自 1998 年以來 CBETA 的電子化作業，即開始採用 TEI/XML (SGML) 標記語言。同時「中華佛學研究所」(CHIBS) 自 1999 年於中央研究院協辦的「國際電子佛典協會」(Electronic Buddhist Text Initiative, EBTI) 會議中，曾與多位以 Metadata 為標準的工作單位聯繫，商洽有關佛教文物 Metadata 的規則與實務，及應用軟體 (Metadata Editor and Search) 的取得與應用等後，即開始進行 Metadata 於佛教電子文獻運用的測試。

隨著參與多次「電子文化地圖協會」(Electronic Cultural Atlas Initiative, ECAI) 國際會議中，瞭解到 Metadata 於佛學文獻資料的電子化的重要性，尤其是佛學書目的管理，以及佛教文物圖像的電子化，皆可採用「都柏林核心集」(Dublin Core) 的 15 個元素¹⁶作為標準建立先製資料，以做檢索與交換之準則。

因此在國科會的「數位博物館計畫——玄奘西域行」、「國家數位典藏計畫——蓬萊淨土遊」以及「佛

學數位圖書館暨博物館 (DBLM)」等計畫中即著手進行 Metadata 的作業。¹⁷目前有很多數位化計畫，均採用 XML 做為 Metadata 或全文資料的標記語法。

1. XML (eXtensible Markup Language) 標記語言於電子佛典的應用

在佛典電子化的過程中，如果僅只是做文字的電子化，那便喪失了原來書本資料所擁有的許多寶貴資訊，例如：經名、年代、作者、譯者、校勘等等。因此 CBETA 的大正藏電子化工作，使用 XML 標記技術將這些寶貴的資訊記錄下來，以便後續進行更廣泛的學術研究應用。CBETA 早期使用 SGML 做電子佛典標記，隨著標記語言及網路工具的發展，於 1999 年已改為 XML。XML 的一大優點就是它可以自訂標籤 (Tag)，因此，各自可以按照自己特殊需求訂定新的標籤。有了共同的標記語言 XML，就可以用同樣的標記語言、標記格式來定義各自不同的標籤名稱。

2. TEI (Text Encoding and Interchange) 標記集於佛典數位化的應用

如前述，XML 是一個允許使用者自訂標記的語言，為使佛教經典能有一套共同的標準標記，以利將來的資訊交換，CBETA 採用國際標準 TEI，為了達到資源共享的交換目的，TEI 整理出一套標記集 (Tag Set)，以利電子文獻的分享與交流。

然而，由於 TEI 並不是針對中文文獻所設計的，在 CBETA 的實做過程中，發現仍然有許多標記是 TEI 所未提供的，於是 CBETA 儘量採用 TEI 標準的標記，若實在沒有適用的標記，便嘗試著自己制定適用於中文佛典的標記，期待能為未來佛典的標準標記有所貢獻。¹⁸因

¹⁵ 參考藍吉富，〈刊本大藏經之入藏問題初探〉，《中華佛學學報》第 13 期 (2000)，頁 167-178。

¹⁶ 都柏林核心集是 1995 年 3 月由國際圖書館電腦中心 (OCLC) 和 National Center for Supercomputing Applications (NCSA) 所聯合贊助的研討會。都柏林核心集的 15 個基本項目有：主題和關鍵詞 (Subject)、題名 (Title)、著者 (Creator)、簡述 (Description)、出版者 (Publisher)、其他參與者 (Contributors)、出版日期 (Date)、資源類型 (Type)、資料格式 (Format)、資源識別代號 (Identifier)、關連 (Relation)、來源 (Source)、語言 (Language)、涵蓋時空 (Coverage)、版權規範 (Rights)。摘錄自吳政叡，〈從元資料看未來資料著錄的發展趨勢〉，《資訊傳播與圖書館學》3 卷 2 期 (1997.12)，頁 46-47；或 <http://mes.lins.fju.edu.tw/pub/j-inf-comm-lib-4-2/meta3.htm>。

¹⁷ 本段引用杜正民，〈簡介 Metadata 於佛教電子文獻的應用——以 TEI 與 DC 實務作業為主〉，《佛教圖書館館訊》第 32 期 (2002.12)，頁 26-40。

¹⁸ 《中華電子佛典協會新聞電子報》第 5 期 (1999.9.24)。

此，CBETA 依《大正藏》為底本，根據國際標準的 TEI 規則所制定的標準標記集，採用 XML 國際標準標記語言等，以作為國際間交換的作業準則，目前已完成第 1 冊至第 55 冊及第 85 冊的《CBETA 電子大藏經》分享於國際間。

CBETA 電子佛典主要目的，就是要利用資訊科技的易於保存、複製、傳播、編輯、再製等電子媒體的便利性。然而從紙本到電子檔的這個過程中，期能保存紙本中的資訊，如大正藏的編排格式、經名、譯者、作者、品名、偈頌、附文、校勘等資料。因此必須在電子檔裏用「標記」的方式記錄大正藏中的各種資訊。

TEI 的標記集裏有許多屬於各種文獻所共有的，可以與電子佛典共用的標記。因此 CBETA 標記進行的方式，原則上參照與採取 TEI 已有定義的規範。如有不適宜者，會對所採用的 TEI 屬性 (Attributes) 修改後使用。如遇 TEI 規則所無者，如漢籍或佛經特有的「經」、「卷」、「品」與「會」等問題則自訂新標記，然後報給 TEI 協會參照或修訂。

至於 CBETA TEI 的工作流程，為工作進行方便，大致分為校對組的簡單標記與研發組的 TEI 標記兩部分進行：所謂「簡單標記」，就是在進入正式的 XML 標記之前，CBETA 校對組會提供一個稱為「簡單標記版」的經文電子檔，如於行首資訊加上經號 (N)、卷名 (J)、作者行 (A)、段落 (P) 等基本標記，提供給研發組作為 TEI 標記作業的基礎。而研發組的 TEI 標記，則是使用轉檔程式，將「簡單標記版」轉為「XML 版」然後經過詳細的檢驗比對與檢查。再以此作為基礎 XML 經文檔，以便最終的成果輸出。

也就是依此 XML 經文檔，以各轉檔程式產生各種不同格式的經文檔，以滿足使用者不同的需求。例如：產生 Normal 與 App 兩種純文字版，以通用字表達缺字，產生 HTML Help 及 DOC 版則以 Unicode 表達缺字，並提供更好的編排等功能。CBETA 現階段的工作重點在於電子佛典的基礎建立，目前對電子佛典所做的標記還是相當有限。希望能進一步運用這些標記，提供方便使用的功能與工具。

3. 都柏林核心集 (Dublin Core)¹⁹ 於佛教數位文獻的應用

在執行的國科會計畫中，有關佛學文物資料庫文物圖像的表達，即是以專業領域共同之 Metadata 作為標準，並參照國內「數位博物館資源組織與檢索相關規範工作小組」所規範之 Dublin Core 以及國外「博物館資訊交換協會」(the Consortium for the Computer Interchange of Museum Information, CIMI)²⁰ 的各項標準所擬定。進而，為讓「佛學數位圖書館暨博物館」的十萬筆書目資料，能充分發揮其功能，因此也擬採用 Dublin Core 標準，以其 15 個元素記錄所有電子化資料的屬性，以之作為先製資料，以利作業人員或使用者運用這些項目來管理電子文獻資料及文物項目，並將參照機讀編目格式 Z39.50 作為檢索與交換的標準。

(三) CBETA 佛典內容與標準規範的關係及其重要性

CBETA 自 1998 年 2 月 15 日起開始《大正藏》電子化作業，於 1998 年年底發表《大正藏》第 5 冊至第 10 冊網路版與光碟版。經一年多的時間完成漢譯佛典「經、律、論」計 32 冊六千多萬字作業，於 1999 年 12 月發行《大正藏》第 1 冊至第 32 冊。同年，CBETA 亦完成建立電子資料庫的「標準化」、「國際化」作業規範，此項作業奠定了國際間漢文電子文獻製作的規範。至 2000 年底，CBETA 已大致完成《大正藏》第 1 冊至第 55 冊及第 85 冊初稿，繼而於 2001 年 4 月發表全文電子檔。其中，有「雙行小字」及「缺字」等特別的處理難度與困難，這些技術的克服與處理，也是國際間漢字處理的一大突破。至此，經過三年二個月的努力，終於完成「大正藏普及版」電子化初步作業。

除上述各項的作業外，CBETA 同時也於 2001 年度積極進行「校勘版」的輸入、校對、標記、缺字處理等工作，此項工作與標準規範有相當大的關係。於 2002 年底完成所有校勘資料電子檔的輸入及電子化初步作業。2003 年，CBETA 成立五週年時，正式發表「校勘版」作業的初步成果。於《大正藏》的電子化作業的同

19 關於都柏林核心集相關資料，請參考下列網址：<http://purl.org/DC>

20 關於博物館資訊交換協會相關資料，請參考下列網址：http://www.cimi.org/documents/meta_bestprac_v031.html

時，CBETA 也積極進行《卍續藏》基本輸入與測試作業，2002年初正式進行《卍續藏》電子化作業，於2003年底發行「CBETA 電子佛典集成」內含「校勘版」與《卍續藏》的電子資料，2004年7月完成《卍續藏》禪宗部文獻的數位化作業。簡言之，CBETA 的《大正藏》電子化作業資料量龐大而齊全，都是以 XML/TEI 的標準為作業基礎。

四、論——佛教文獻數位化的技術與研發

上文簡單敘述 CBETA 《大正藏》近億字的輸入與標記等數位化作業內容，如電腦缺字處理、標記作業與 XML, Unicode 的採用。以及，資訊技術與佛學知識庫的建構、標準規範 TEI, Dublin Core, Topic Maps 的應用與實施等，接著討論有關的技術與研發。

(一) 現代佛學數位資源的開發

1. 中華電子佛典協會

中華電子佛典協會自 1998 年成立後在藏經電子化方面有各項成果，²¹ 如 1998 年發表「《大正新脩大藏經》第 5 冊至第 10 冊」，1999 年及 2000 年發表「《大正新脩大藏經》第 1 冊至第 32 冊」，2001 年發表「《大正新脩大藏經》第 1 冊至第 55 冊暨第 85 冊」。

2. 佛學數位圖書館暨博物館

「佛學數位圖書館暨博物館」的前身為臺大釋恆清教授於 1994 年所創建的「佛學網路資料庫」，²² 進而於 1999 年 10 月至 2003 年 3 月由中華佛學研究所與臺灣大學合作進行「佛學數位圖書館暨博物館」的建構與

發展。該資料庫自 1994 年起至今已經完成的佛學書目資料約計 10 萬筆，建置中英文學術論文全文資料庫有 3000 篇左右，以及「佛學網路梵文、巴利文、藏文佛典語言教學」、「佛學網站資源」、「佛教工具」等資料庫提供學界檢索應用。此項計畫目前由臺大圖書館負責，正蓬勃發展中。

3. 國家數位典藏等佛教數位資料庫

除上述佛學研究資源的數化作業之外，近年在許多學術團隊合作下，並完成多項大型計畫，所建構的資料庫有：數位博位館玄奘西域行（國科會計畫：國立臺灣大學、國立臺北藝術大學、中華佛學研究所合作執行 1999-2001）、臺灣佛教數位博物館——蓬萊淨土遊（國家數位典藏計畫：國立臺灣大學、國立臺北藝術大學、中華佛學研究所合作執行 2001-2002）、臺灣佛教文獻數位資料庫（蔣經國國際學術計畫：國立臺灣大學、國立臺北藝術大學、The Catholic University of America、中華佛學研究所合作執行 2001-2003）、法鼓全集電子版光碟專案（中華佛學研究所與法鼓文化事業合作執行 2001-2003）、《瑜伽師地論》資料庫——電子佛典製作與運用之研究（行政院國科會計畫案：1999 - 2002）、《法華經》多種語文版本數位資料庫的建構與研究（蔣經國國際學術計畫：日本立正大學、日本德島大學以及中華佛學研究所合作執行 2003-2004）等計畫。同時，中華佛學研究所也建構有緬甸聖典寫本簡明目錄電子資料、發行穆克紀教授訪臺講學紀念專輯，以及各項學術資料文獻資料庫等。

以上諸計畫的研究報告，歷年來也都在國際會議發表，如 EBTI, ECAI 及 PNC 等會議中與眾分享。²³

21 參照杜正民，〈以 CBETA 為例談大量文獻之建立——漢文藏經電子化作業簡說〉，《中央研究院計算中心通訊》第 15 卷第 13 期（1999.6），頁 117-123；及〈漢文電子大藏經的製作緣起與作業流程：以「中華電子佛典協會」為例〉，《佛學研究中心學報》第 4 期（臺灣大學文學院佛學研究中心，1999.7），頁 347-369；〈中華電子佛典協會 1999 年度成果發表會——《大正藏》印度撰述部經、律、論電子佛典發行〉，《現代佛教學會會訊》第 5 期（2000.1），頁 10-13；〈漢文佛典電子化記事——回顧 CBETA 1998 至 2000 年的活動與成果〉，《佛教圖書館館訊》第 24 期（2000.12），頁 60-83。

22 相關資料，請參照杜正民，〈「佛學網路資料庫」的建構過程與內容簡介〉，《國立臺灣大學哲學論評》第 23 期（2000.1），頁 313-335；〈佛學數位圖書館暨博物館——臺大與中華佛研所共同開啓新的里程碑〉，《佛學研究中心學報》第 5 期（臺灣大學文學院佛學研究中心，2000.7），頁 341-351。

23 杜正民，〈佛學數位博物館的建構——從文獻數位化跨向文物數位化的一小步〉，《佛學研究中心學報》第 6 期（臺灣大學文學院佛學研究中心，2001.7），頁 347-366。

(二) 立基前人智慧 古今結合實例

依上述佛學數位資源為基礎，所建立的引用索引資料庫及引用分析等進一步功能發揮，未來如何將上述資源統籌建置完整的 e-Library 提供服務，如何運用這些資源作 e-Learning 教學與推廣，這些計畫都在持續進行。

中華電子佛典協會的《大正藏》電子化作業，²⁴由於作業流程²⁵相當繁複緊密，故本文僅以作業流程示意圖表示如下：

資料輸入及比對經文搜集 → 缺字及格式整理
 → 檔案比對 → 看圖校對 → 人工校對 →
 標記處理 → 電子檔上網 → 全文檢索等應用
 → 使用者之建議與指導 → 修正並更新

由上表的作業流程可看出，利用標記語言來進行經文資料的瀏覽、管理和維護，是 CBETA 具特色的作業方案。由於近年來許多高品質及多功能的瀏覽器相繼推出，再加以網路使用的普及，使用標記語言進行經文資料的整理，實具有許多的便利與價值。因此在純文字經文處理的階段，CBETA 採用簡單標記作業，進而利用這些簡單的行標記方式，再經過程式的轉換和處理，便能產生較精密的 TEI/XML 標記版本。

五、結語：佛教文獻數位化的挑戰與省思

身處日新月異的資訊時代，如何掌握數位資源的特性，運用所建構的佛教數位文獻，以開展契機的學習方式，為當代佛教的重要議題。本文嘗試以新時代的資訊方法，賦予古文獻新的生命，一方面藉此活化傳統佛學豐碩資源的應用，另一方面作為建構新時代佛學研究方法學的論題及人才培養規劃的議題。²⁶

佛教文獻數位化不僅是在於確立相關的標準與規範，針對網路資訊日新月異的特性，不可諱言的面臨了許多的困境，比如培養兼具佛學與資訊知能的人才，運用資訊科技的媒體與工具，有效管理資訊文獻，改進佛教教學、研究等各個不同的問題。

對應以上在文獻數位化過程及未來將面對的挑戰，透過佛學與資訊課程的規畫及設計，希望培養未來能夠兼具佛學與資訊知能的人才，以及考量如何推動發展佛學數位學習，以建立資訊時代佛學終身學習環境，促進佛教教學、研究之創新與變革，建構教學、研究結合之知識網絡，讓佛學資訊之相關研究及教育工作更符合時代脈動。

24 杜正民，〈漢文佛典目錄暨全文之數位化作業與規範〉，「古籍聯合目錄資料庫合作建置研討會——古籍的名稱規範」，(2001.04.19-20) 臺北：國家圖書館。

25 有關中華電子佛典的作業流程與簡介，請參閱杜正民，〈漢文電子大藏經的製作緣起與作業流程：『中華電子佛典協會』簡介〉，《佛學研究中心學報》第4期（臺灣大學文學院佛學研究中心，1999.7），頁347-369。

26 杜正民，〈佛學與資訊的整合〉，「兩岸佛教學術研究現況與教育發展研討會」（臺北：中華佛學研究所，