

數位化時代的古籍整理

New Solution for Tackling with Old Rare Books in the Digital Age

方廣鎬 **Fang Guang-chang**

上海師範大學哲學學院教授

Professor, College of philosophy, Shanghai Normal University

E-mail: f4g4c8@163.com

【摘要 Abstract】

現代數位化技術的突飛猛進，為中國古籍整理提供了前所未有的挑戰與機遇。本文介紹了筆者與電腦技術人員共同開發的人機互動古籍整理數位化新技術，以期為中華古籍整理開拓新的道路。

The nowadays fast advancement of new digitization technologies has provided us new solutions for tackling with old rare books. It could enable us to achieve what was not possible before but at the same time, serious challenges are awaiting for us to overcome. The author described how a new digitization technology for tackling with old rare books based on human-machine interaction was developed under the collaboration with computer technician. It is hoped that the new solution described would be a breakthrough methodology for the preservation of contents of Chinese invaluable old rare books.

【關鍵詞 Keywords】

數位化、古籍整理、古籍數位化

Digitization; Tackling with Old Rare Books; Digitization of Old Rare Books

一、前言

古籍是一個民族歷史與文化的重要載體，整理古籍也就成為各民族承襲歷史、弘揚文化的重要事業。中華民族是一個有著高度文明自覺的民族。筆者在這裡所謂的「高度文明自覺」，不僅指中華民族對自己創造的文明具有高度自信，而且指從古到今，中華民族始終努力採用各種方式力求將自己的文明繼承下來，發揚光大，傳承下去。由於典籍是文明傳承的主要方式，因此，皓首窮經、孜孜不倦地整理古籍的人士前仆後繼，代有人出。在這一過程中，中華古籍經歷了寫本、刻本、近現代印刷本等不同時期，現在已經踏入數位化的門檻。

數位化技術的出現為我們整理傳統古籍提供了新的機遇，能否抓住這一機遇，打開中華古籍整理的新局面，則在於中華古籍整理與研究工作者的努力。

二、傳統古籍整理模式及其弊病

百年來，古籍整理工作取得的巨大成績有目共睹，自然也存在若干有待研究與解決的問題。筆者認為，目前存在的最大問題是人類社會已經進入數位化時代，我們的古籍整理雖然也開始利用數字技術，但從總體看，還處在數位化的初級階段，沒有真正擺脫傳統古籍整理模式的束縛，沒有真正發揮數字技術的優勢以開創古籍整理的新局面。

所謂目前古籍整理還處在「數位化的初級階段」，主要體現在目前推出的古籍整理數位化成果，大多為紙本整理成果的載體轉換。亦即將古籍由以紙張為載體轉換為以數位化為表現載體。舉例而言，利用近現代印刷技術推出的古籍整理成果有點校排印本、影印本兩種形態。與此相應，初級階段的數位化古籍，也出現用文字錄入方式形成的電子排印本及用圖像掃描方式形成的電子掃描本。這些形態兩兩對應，祇是前者為紙張載體，表現為一本一本的實體書；後者為數字表現載體，並可以利用網絡傳播，在各類顯示器上閱覽。但無論哪一種古籍整理本，目前大抵屬於平面展現。

當然，嚴格地講，初級階段的數位化古籍已經與紙本古籍不可同日而語。比如，無論是電子排印本，還是電子掃描本，不少數位化古籍都可以實現全文檢索。有些數位化古籍開始採集檢索點，建成關聯資料庫，初步建立起相關的知識之網；有些數位

化古籍進而採用圖像技術，營建虛擬場景等等。特別應該指出的是，利用數字技術建立的古籍目錄資料庫，其強大的檢索功能使得同類的紙本目錄索引類著作瞠目難及。可以說，數位化技術已經在古籍整理領域顯示其無限的生命力與功能的擴展可能性，但是，由於目前古籍整理界還沒有真正擺脫傳統古籍整理模式的束縛，從而使上述數位化成果的品質也難以突破傳統古籍整理的水準，限制了數位化古籍各種功能的充分發揮。

什麼是所謂「傳統古籍整理模式」？簡單地講，就是東漢劉向總結的「校讎」。在劉向時代，所謂「校」，指某人對某一文本進行閱讀，依據上下文理，校正錯誤。所謂「讎」，則由兩人合作，一人執一本宣讀，一人對另一本逐字核對、校改。隨著時代的發展，後代的古籍整理逐漸演變為由某位整理者一人對校兩本或數本，最終定稿；少數亦有採用會讀乃至以讀書班方式從事古籍整理者。無論學養多麼高超，在一定條件下，人的水平總有局限；無論工作態度多麼精審，在某個時間段，人力終有窮盡。古籍整理還往往出現這樣的情況：舊的錯誤被糾正，新的錯誤又產生。所以，採用這種方式推出的成果，難免存在種種不足。以致自古流傳這樣的感慨：「校書如掃落葉，旋掃旋生。」任何一個古籍整理工作者，都不敢說自己的工作已經盡善盡美，自己完成的整理本不再存在錯誤。於是，我們可以看到這樣的景象：某類文獻、某種典籍，雖然已由多位整理者反復進行點校整理，但學界依然不滿意，重複的整理工作至今仍在進行。如敦煌本《壇經》的整理校注本，據筆者所知，已有二十種多種，雖則如此，至今尚未出現一個學界公認的「善本」。《祖堂集》至少有 10 種。敦煌變文的各種錄文與校註本，已經有 10 種以上，新的錄校本還在計劃推出。不僅佛教典籍，如司馬遷的《史記》，已經出版了二十多種。《太平廣記》這樣的大型類書，各種校訂本竟然也出版了十多種。整理者大量心血的付出，其間有多少屬於有效勞動，多少屬於無效勞動；有多少是真正的學術積累，有多少是無價值的學術浪費，實在難以計量。

問題還在於按照傳統的古籍整理模式進行工作，工作結束以後整理者雖然為讀者提供一個整理本、一份校勘記。讀者固然可以對照校勘記閱讀整理本，但如果整理者本人的校勘工作有疏漏或犯新的錯誤，則這種疏漏與錯誤自然不會反映在校勘記中，亦即此時校勘記不能真正反映底、校本的真實情況，讀者就有可能被整理本與校勘記所誤導。要想不受誤導，除非讀者自己依據原始資料逐一去核對整理本。然而，作為一個讀者，一般不易接觸到原始資料；且如讀者去逐一核對原始資料，則相當於讀者本人把該古籍再整理一遍。這都是不現實、不可能的要求。

那麼，面對一種新出的古籍整理本，讀者又如何鑒別其整理水平的高低、其錯誤的多少、其可信用的程度？目前一般的方法，無非是讀者依據本人積累的對某整理者學風的印象來做評價，或者查核整理本中的若干章節來做判斷，或聽取其他人的評價意見以為參考。坦率說，印象未必靠得住，同一個人的不同成果，水平可以有不同；查核若干章節，畢竟是以點代面，難以作出全面的評價；至於所謂「其他人」的評價，無非是另一個讀者而已。

總之，應該承認，目前對古籍整理成果的評價，實際具有一定的盲目性。但是，研究者依據古籍整理成果從事研究，卻又對該成果的整理水平難以做出正確的評價，那研究者又如何推進自己的研究呢？

2014年11月參加日本佛教大學舉行的關於佛教大藏經的國際研討會，會上日本人文情報學研究所永崎研宣提到：日本《大正藏》雖以《再刻高麗藏》為底本，但工作中實際將《頻伽藏》用為工作底本。永崎研宣（2014）亦提到，早在1928年，山崎精華先生就在日本《現代佛學》上撰文提到《大正藏》用《頻伽藏》作工作底本這一問題。不知其他學者以前是否關注過這一點，起碼筆者以前沒有注意過。永崎研宣先生的論述使筆者大為吃驚。因為《頻伽藏》印刷於上世紀初，錯誤很多，從來被認為是不可信用的本子。而《大正藏》編成以後的八十年中，從來被認為是質量最高、學術性最強的本子，乃至風行全世界，既成為全世界佛教研究者從事學術研究的依據，也成為目前許多電子版大藏經的基礎。筆者以前也曾發現過《大正藏》的若干錯誤，但總以為那是《大正藏》編輯者的偶爾疏漏—這種疏漏，對任何一部大藏經都是難以避免的。現在按照永崎研宣先生的提示進行複查，發現很多問題果然出在《頻伽藏》。

比如，隋費長房《歷代三藏記》卷十五論及自己編纂《歷代三寶記》，有這樣一段話：

臣幸有遇，屬此休時。忝預譯經，稟受佛語。執筆暇隙，寢食敢忘。十餘年來，詢訪舊老，搜討方獲。雖粗緝綴，猶慮未周。廣[21]究博尋求敬俟來俊。（費長房，無年代）

末句之「廣[21]究博尋求敬俟來俊」，文意不通。《大正藏》對此有校記，作：「[21]〔究〕—【宋】【元】【明】【宮】。」這條校記說明宋、元、明、宮本中無「究」，亦即按照《大正藏》的體例，《大正藏》所用底本—《再刻高麗藏》—的文字即為「廣究博尋求敬俟來俊」。但查《再刻高麗藏》，原文明明作「廣究博尋，敬俟來俊」。按

照永崎研宜先生的提示，再查《頻伽藏》。果然，《頻伽藏》的文字為「廣究博尋求敬俟來俊」。也就是說，在這裡，《大正藏》明明以《頻伽藏》為底本，多出一個「求」字，以致行文錯誤。但未作交代，依然自稱底本是《再刻高麗藏》，從而誤導了研究者。

《大正藏》更離奇的錯誤，是筆者 1980 年撰寫碩士論文時發現的。筆者的碩士論文為「《那先比丘經》研究」，研究中意外發現，《大正藏》所收的二卷本《那先比丘經》出現錯版，並因此擅加文字，且不出校記。

該二卷本《那先比丘經》在《大正藏》中編為第 1670 號，載第 32 卷。其中第 702 頁中第 27 行末至 702 頁下第 9 行有一段關於智者與愚者作惡後得殃是否相同的問答。為了說明問題起見，筆者把三卷本《那先比丘經》的同一段問答（載第 32 卷，第 718 頁上）也抄錄如下，以作比較。

表 1

《那先比丘經》二卷本、三卷本經文比較一

二卷本	三卷本
<p>王復問那先：「智者作惡，愚人作惡，此兩人殃咎，誰得多者？」那先言：「愚人作惡得殃大；智人作惡得殃小。」王言不知那先言。王言：「我國治法，大臣有過則罪之重，小民有過罪之□。¹是故我知智者作過惡得殃大，愚者作惡得殃小。」那先問王：「譬如燒鐵在地。一人知為燒鐵，一人不知。兩人俱前取燒鐵。誰爛手大者耶？」王言：「不知者手爛。<u>不制其身口者，不能持經戒，如此曹人亦不樂其身。</u>」那先言：「<u>其學道人者，能制其身，能制口，能持經戒。能一其心得四禪，便能不復喘息耳。</u>」王言：「善哉！善哉！」</p>	<p>王復問那先：「智者作惡，愚人作惡，此兩人殃咎，誰得多者？」那先言：「愚人作惡得殃大；智人作惡得殃小。」王言不如那先言。王言：「我國治法，大臣有過則罪之重，愚民有過則罪之輕。是故智者作惡得殃大，愚者作惡得殃小。」那先問王：「譬如燒鐵在地。一人知為燒鐵，一人不知。兩人俱前取燒鐵。誰爛手大者耶？」王言：「不知者爛手大。」<u>那先言：「愚者作惡，不能自悔，故其殃大。智者作惡，知不當所為，日自悔過，故其殃少。」</u>王言：「善哉！」</p>

上述兩段文字，前半部分相同，後面劃線的部分大異。根據行文內容，很顯然，三卷本的文字正確，二卷本的文字錯誤。對照巴利語《彌蘭陀王之間》，結論也完全一樣。在此再將巴利語《彌蘭陀王之間》的相關段落翻譯如下：

王問：「那伽先那尊者！知者行惡與不知者行惡，誰的禍大？」

¹ 原文此處空一字，詳見內文。

長老回答：「大王！不知者行惡，所得禍大。」

「原來如此。尊者那伽先那！我們的王子、大官如果作惡，要比不知者作惡，予以加倍的處罰。」

「大王！您（對下述情況）是怎麼想的呢？如果有一個灼熱、燃燒著的鐵球。一個人知道而去握它；另一個人不知道也去握它。那麼，誰被燒傷得厲害呢？」

「尊者！不知道而去握它的人被燒傷得厲害。」

「大王！與此相同，不知者行惡，所得禍大。」

「善哉！尊者那伽先那！」（中村元、早島鏡正，1972）²

那麼，二卷本有無上述三卷本錄文中劃線的「大。那先言：愚者作惡，不能自悔，故其殃大。智者作惡，知不當所為，日自悔過，故其殃少。王言：善哉！」這一段文字呢？有！就在 702 頁下第 25 行至第 27 行。全文一字不差，只是最後一句彌蘭陀王的讚歎語中多說了一個「善哉」而已。進而仔細檢查，發現從二卷本 702 頁下第 6 行「不制其身口者」起，到同欄第 24 至 25 行「和所為得人者」止的 295 個字都與原文行文不協，肯定是從其他地方脫落後竄入此處的。

那麼，這 295 個字是從哪裡脫落的呢？仔細研究，這 295 個字包括了四個問題：關於止息喘息的問答的後部分；關於大海的問答；關於得道思維深奧眾事的問答；關於人神、智、自然異同問答的前部分。經查，原來它們應該位於第 703 頁上欄第 16 行的「不能」與「那先問王」之間。「不能」之前，正是關於止息喘息問答的前部分；而「那先問王」之後，正是關於人神、智、自然異同問答的後部分。二卷本此處本來語義也不通。但把脫落的文字加入後，意義就連貫通順了。與三卷本的相同部分的文字也正好對應。

但新的問題又出來了。為了便於說明這個問題，在此把將脫落文字插回原處之後的二卷本有關段落與三卷本的有關段落抄錄比較如表 2。

² 據中村元、早島鏡正日譯本轉譯。

表 2

《那先比丘經》二卷本、三卷本經文比較二

二卷本	三卷本
<p>王復問那先：「卿曹諸沙門說言：『我能斷喘息之事。』」王言：「奈何可斷喘息氣耶？」那先問王：「寧曾聞志不？」王言：「我聞之。」那先言：「王以為志在人身中耶？」王言：「我以為志在人身中。」那先言：「王以為愚人不能不制其身口者，不能持經戒，如此曹人亦不樂其身。」那先言：「其學道人者，能制其身，能制口，能持經戒。能一其心得四禪，便能不復喘息耳。」王言：「善哉！善哉！」</p>	<p>王復問那先：「卿曹諸沙門說言：『我能斷喘息之事。』」王言：「奈何可斷喘息氣耶？」那先問王：「寧曾聞志不？」王言：「我聞之。」那先言：「王以為志在人身中耶？」王言：「我以為志在人身中。」那先言：「王以為愚人不能制其身口者，不能持經戒者，如此曹人亦不樂其身。」那先言：「其學道人，能制身口，能持經戒。能一其心，得四禪，便能不復喘息耳。」王言：「善哉！」</p>

上述二卷本錄文中的劃線部分為移入的脫落文字。與三卷本錄文對照，最大的差別在於三卷本說：「愚人不能制其身口」；而二卷本的文字卻是「愚人不能^不制其身口」，多了一個「不」字，以致文意完全相反。很顯然，三卷本的文字是正確的。證之巴利語《彌蘭陀王之間》，結論也相同。而二卷本的那個「不」字是整理者擅加的。整理古籍，擅加文字而不出校記，是古籍整理的大忌。《大正藏》中出現了這種低級錯誤。

1980年，筆者還沒有條件查對《再刻高麗藏》原文，由於錯亂的字數大體相當於《高麗藏》一版，當時推測上述文字的錯亂乃至那個擅加的「不」字都是由於原本《再刻高麗藏》錯版所致。十多年以後有條件查對《再刻高麗藏》原文，纔知道《再刻高麗藏》並沒有錯，那麼，錯誤自然出在《大正藏》了。所以，筆者當時的結論是：「或者由於《大正藏》所利用的那部《高麗藏》的印本此處裝裱顛倒所致？但《高麗藏》每版均有版片號，按道理能夠發現這種顛倒。不管怎樣，錯版發生了，但沒有被發現糾正。不僅如此，《大正藏》的校對者竟然擅自又加上一個『不』字」（方廣錫，2011）。

此次按照永崎研宣先生的提示，筆者查核《頻伽藏》，發現《大正藏》這一的錯誤的出處依然在《頻伽藏》，原來是《頻伽藏》行文錯亂，並且擅自加了那個「不」。《大正藏》不過是照抄而已。

值得注意的還有，表 1 顯示《大正藏》本有一句話作「小民有過罪之□」，《大正藏》本此句「之」下空一格，表示底本缺漏一字。但查《再刻高麗藏》，此處明明有字，作「輕」。這樣，該句應為「小民有過罪之輕」，上下文意通順。再查《頻伽藏》，

也有這個「輕」字。也就是說，《再刻高麗藏》、《頻伽藏》都有的字，《大正藏》竟然稱之為缺漏。令人奇怪的還在於，《大正藏》整理者既然知道此處有缺漏，用空一字的方式表示該缺漏，且《大正藏》所用的底本《再刻高麗藏》、所用的工作底本《頻伽藏》均有此「輕」字，《大正藏》卻沒有用來補正自己的缺漏。作為一個曾經實際從事藏經編輯的工作者，筆者以為這裡體現的是《大正藏》在 workflow 方面的疏失。

我們無法一一查核《大正藏》、《再刻高麗藏》與《頻伽藏》，不知道類似的錯誤到底有多少。學術研究依靠文獻資料，如果所依靠文獻本身的準確性無法保證，那學術研究又如何保證自己的水平，保證不出錯誤？

筆者是研究佛教文獻的，上述舉例也都以佛教文獻為主。其實，即使中華書局出版的二十四史，向來被視為古籍整理的翹楚，但至今依然有人不斷發現問題，提出商榷，進行修訂。學術研究依靠文獻資料，如果所依靠文獻本身的準確性無法保證，那學術研究又如何保證自己的水準，保證不出錯誤？

問題的關鍵，還是上文所說：傳統的古籍整理雖然為讀者提供一個整理本、一份校勘記。讀者可以對照校勘記閱讀整理本，但由於整理者沒有提供他所依據的原始資料，因此，如校勘工作本身有疏漏，整理本的錯誤沒有反映在校勘記中，或校勘記不能反映底、校本的真實情況，那讀者就頗為困惑了。因為讀者既然不知道自己所依據的整理本竟然還存在諸如此類的問題，祇能依據這種整理本來開拓自己的研究。相反，如果在提供整理本的同時，能夠一併提供相應的原始資料，並讓讀者方便地查核與利用它們，讀者就可以依據原始資料作出自己的抉擇，不至於被前人的整理本中的錯誤所誤導。

三、數位化古籍整理的初步構想

數位化技術為我們走出上述困境提供了機遇，我們可以按照如下四個基本原則，利用數位化技術，開創古籍整理的新局面。第一，起於最底層：古籍整理，要從最基礎的原始資料的圖形文字、書寫符號的切割開始；亦即將所有用於校勘的古籍原本上的全部文字與符號統統切割下來，存入資料庫。第二，資訊全覆蓋：資訊採集要覆蓋原本上的全部原始資料。亦即保留原本中一切可供研究的資訊。第三，過程可追溯：利用數位化技術，通過人機互動的切字、認字、定字、校勘、標點等工作環節，最終

完成古籍整理。全部工作流程的每一個步驟，包括每一環節所用資料、所做工作、所得結果都作記錄，事後都可以追溯與覆覈。第四，功能可擴展：界面友好、開放，可與讀者互動。並可隨時根據新的要求，擴展新的功能。我們設計的工作流程如圖 1。

首先，從古籍原本最基礎的文字與書寫符號的切割、辨認開始，即把古籍原本上的每一個文字、每一個符號都切割下來，並將它們全部轉換成電腦可以識別的具有電腦內碼的文字與符號，由此形成基礎工作文本與基礎字形庫。基礎工作文本是一個與該古籍原本行文完全一致的數位化文本，以供校勘之用。在這一過程中，傳統的底本、校本概念將被顛覆，所有的原始文本在新的古籍整理工作中將被一視同仁，處於平等地位。基礎字形庫存儲該古籍原本的所有文字與符號，並保留其原始圖像形態，既供追溯檢查所用，又可以成為文字研究者的研究資料。其次，系統排比基礎工作文本，提示整理者辨析異本。古籍在流傳過程中時有異本產生，異本可以參校，不宜混同。故正式校勘之前，必先區別異本。

再次，校勘與標點。有經驗的古籍整理者都知道，一般來說，用於對校的兩個文本的差異不會超過 90%。但校勘者必須耐下性子，一個字一個字去校勘，唯恐有所疏漏。因此，校勘中付出的勞動，幾乎有 90% 左右實際屬於無用的功夫。而由電腦系統自動比對不同文本，如果對應的文字相同，系統自動忽略，僅將不同的文字用色標顯示，提示研究者進行勘校。這樣，研究者固然還是需要通讀全文，但就校勘環節而言，工作量可以減輕 90% 左右。不僅如此，針對不同情況，系統設計了規範的校勘記表述格式並按要求自動生成規範校勘記。研究者進而可利用該系統，在同一界面對文本進行分段、標點。

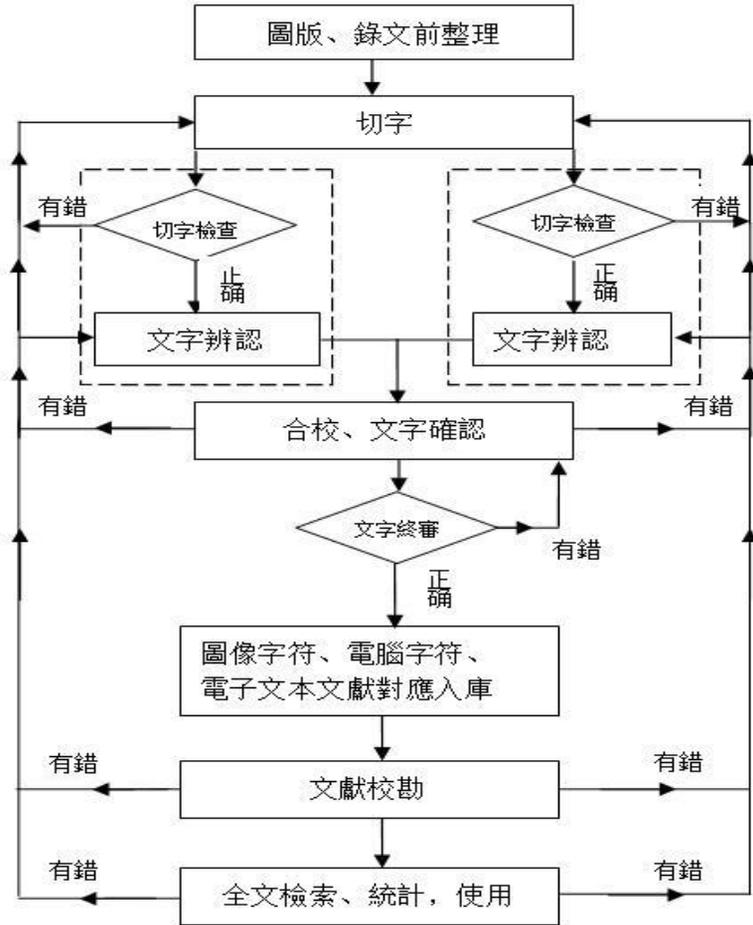


圖 1 數位化古籍整理流程示意圖

由此完成的數位化整理本，將充分利用數位化技術的多層次、多功能的縱深優勢，將被整理古籍的文字、文物、文獻、研究史等各種資訊鏈結為一個綜合性的關聯資料庫，並予以多層次呈現。這樣，每個古籍文獻都可以建成一個資料庫，既相對獨立，又與其他古籍文獻資料庫相互關聯，且具有開放性、互動性，為讀者提供交流平臺。讀者可以在該平臺追溯、檢查整理者的全部工作及所用各校本諸文字、符號的原始圖版，評點整理者的工作，提出自己的修訂意見，從而使整理本得以不斷修訂錯誤，逐漸升級，最終臻於至善。

按照上述思路，新的數位化互動整理模式將徹底改變目前古籍整理中大量出現的重複勞動，使每個整理者的工作、每位讀者的修訂意見都成為對該被整理文獻的有效學術積累或不同的參考意見。這種整理本也將給知識點的採集、知識網的構建等各種後續工作賦予更加堅實的基礎。

如上所述，採用上述方式，傳統底本、校本的觀念將被顛覆，無論什麼本子，在上述校勘過程中地位平等、作用相同。校勘工作者將對所有的工作本一視同仁，逢異必校，擇優而從，並保留各個本子的所有資訊，以備複查。

四、結論

數位化時代的古籍整理，將充分依託飛速發展的數位技術及資料庫技術，將每個古籍、文獻的全部資訊，採集、歸納、整理建設為一個資料庫，並將諸多不同古籍、文獻的資料庫建設為關聯資料庫，從而將平面的古籍文獻拓展為立體的資訊資源庫，以充分發揮保護古籍、利用古籍的效用。

從 2013 年夏天開始，我們已經按照上述思路，編製相關軟體，漸次開展工作，並已經取得階段性成果。目前正在進一步探索中。

附記：本文改寫自拙作「數位化時代的古籍整理」，刊登於《鑑藏：兩岸古籍整理與維護研討會論文集》（2015 年 8 月），31-38 頁。

參考文獻

（傳統文獻）

東晉·作者不詳（無日期）。那先比丘經（二卷本，《大正藏》本）。在《大正新脩大藏經》第三十二冊 No. 1670A。檢自 CBETA 電子佛典 2014。光碟版。

東晉·作者不詳（無日期）。那先比丘經（三卷本，《大正藏》本）。在《大正新脩大藏經》第三十二冊 No. 1670B。檢自 CBETA 電子佛典 2014。光碟版。

隋·費長房（無日期）。歷代三寶記，《大正藏》本。在《大正新脩大藏經》第四十九冊 No. 2034。檢自 CBETA 電子佛典 2014。光碟版。

（近代文獻）

中村元、早島鏡正（1972）。彌蘭陀王之問。平凡社，8 月，第 246 頁。

方廣錫（2011）。《大正新修大藏經》評述。在《隨緣做去，直道行之》（頁 77）。北京市：國家圖書館出版社。

永崎研宣(2014)。デジタル化の現場から見えてくる大正新修大藏經。在平成26年度秋期特別展關連シンポジウム——縮刷藏經から大正藏經へ。日本佛教大學宗教文化ミュージアム，京都。