

試論大學圖書館中的數位化典藏

館長 呂宗麟

一、前言

傳統大學圖書館的功能主要著重在蒐集圖書資料，系統地提供大學中教學與研究所需的服務，其基本過程是蒐集→組織→保存；提供服務內容畫分大致為：採訪、交贈、分類、編目、加值、典藏以及借還、參考諮詢、ILL、文獻資料複印等服務項目；自從網際網路以及相關資訊技術的逐漸成熟，現今的大學圖書館必須同時承擔傳統圖書館與數位圖書館(Digital Libraries)的功能與角色，國內學者顧敏稱之為「複合圖書館」(Hybrid Library)，並認為複合圖書館是一種存在的現實，同時也是一種提供服務的思考¹。

就典藏業務而言，現今大學圖書館的典藏內容而言，應同時包括傳統典藏與數位典藏(Digital Archives)兩種；因此，圖書館員必須同時透析大學圖書館的傳統典藏與數位典藏內涵，以及熟悉運用傳統服務與數位服務兩種性質的服務²；在本文中，並不對傳統典藏與傳統服務作描述，僅擬對數位典藏之基本概念，以及在大學圖書館運作中，必須思考的問題，作一初步的考量。

二、數位典藏的形成與發展

就數位典藏的發展歷程而言，基本上是從資料庫(database)開始，一九七二年美國 DIALOG 公司即已提供包含書目性(Bibliographical)、數目性(Numeric)與指南性(Directory)三種性質的線上資料庫(Online Database Access)業務；以初期數位典藏的內容而言，由於受到資訊技術的限制，只能以電腦檔案(file)的方式，簡單地典藏元素性的資料(data)，統計報表和書目資料是兩種主要的典藏內容，直至二十世紀八〇年代由於資料庫結構技術與資料庫應用語言的漸趨成熟，數位典藏始開始大幅增長；而一九八〇年《大英百科全書》及《美國學術百科全書》(American Academic)全文數位化典藏的製作成功，確實開啓了全文數位典藏的扉頁；一九九五年二進位系統成

¹ 顧敏，〈數位時代複合圖書館的服務及發展策略〉，《中國圖書館學會學報》第 63 期，1999 年，頁 91。

² 同前註。

為各種資訊包括文字、影像、聲音與動畫的共同基礎；³同時，帶動圖書館界必須思考如何在面對如此巨大的變化中，並不僅是由於 Internet 的因素，以及建構 Internet 知識的 content，而是整個大學圖書館經營中所有 contents 的改變，而此種改變的原因，係在於「數位典藏」的具體形成。

就「數位典藏」一詞的基本涵義而言，並非僅止於將文本式圖書資料數位化的簡單過程，換言之，不能單純地將其視為一個資料的轉換程序看待；正如對「數位圖書館」的理解，不應將其僅界定於「是一個電子化資源的收藏庫，讓人們可以隨時透過網際網路取得資訊並使用」，以及「以虛擬環境提供數位資訊，讓人們可以透過網際網路自由地瀏覽、使用並傳播數位資訊，……」等等較表象上概念的延伸；我們必須系統性地思考，在典藏數位化之後，這些資訊如何再運用的問題，未來的加值資訊生產者和使用者會是那些人？就讀者而言(並非僅限於現有讀者，實應包括潛藏讀者在內)會有那一些基本資訊使用的需求？如何以有限的經費建構典藏數位話的環境？以及如何能夠確信可以免除受到事件的發生？……等等問題的考量；因此，大學圖書館中的數位典藏的基本涵義，應該並非只是投入經費將文本式的圖書資料，轉成數位化資料在網路上流通而已，正如，今日「數位圖書館」的基本定義，是指提供數位典藏、服務和人員的整合環境，以支持資料、資訊和知識的生產、加值、使用與典藏資訊生命週期的數位資源⁴，而非僅止於網路化的圖書資料庫而已。

由上述對於「數位圖書館」的審慎思考，吾人大致可以得出對於「數位典藏」的理解，應指的是：「讓數位圖書館能夠輸出其藏品資源機制中的重要元件」，當然必須透過標準語言，方能達成其作用，正如 Attribute Model Proxy 將真實世界裏的 metadata 表示成物件(object)，亦如在概念中我們將各種的 metadata，如 US-MARC、Dublin Core、Z39.50、Bibl GILS 等，都以 Attribute Model Proxy 並提供 metadata 相關資訊的查詢，例如 metadata 的說明、屬性的形態(Type)與值域(Value)以及屬性間的關連等⁵。

³ 同前註文，頁 92。

⁴ 參閱，何建明等，〈典藏數位化資訊環境之探討〉，台北《圖書館學與資訊科學》第 26 卷第 2 期，2000 年 10 月，頁 40。

⁵ 柯皓仁等，〈詮釋資料與數位圖書館系統互通性之探討〉，國立台灣大學《大學圖書館》第 5 卷第 1 期，2001 年 3 月，頁 58。

我們在思考數位典藏時，應該就會考慮如何將各種的文體、各式的出版媒介，以及如何才能將人類可以瞭解的知識表達，很明顯地用一些標誌將其標示出來，這就是發展建構數位典藏環境中的 Text Encoding Initiative(簡稱 TEI)標準，TEI 所用的 meta 語言，一樣是用 SGML 與 XML (Extensible Markup Language)；誠如國內學者劉清俊教授所言：「metadata 是做數位典藏的背景訊息，TEI 做文件本身內容的註釋，這兩個合起來，才變成一個比較典型的計算機中知識的表達」⁶，也才能成為形成數位典藏環境的重要基礎。

三、數位典藏化環境建立的問題

大體而言，資訊系統的設計，通常必須先考量使用的環境，以及影響環境的可能變項(variable)，實不能僅就技術層面的可行性與否，作為判斷的唯一思考，筆者僅就數位典藏環境建立的資料庫標準化問題--即是否有一個公共資訊系統可以建立起來，使讀者均能分享這個機制，再則，如何公平、合理地解決智慧財產權(數位產權)兩個面向，思考之。

(一)資料庫標準化—公共資訊系統的建立

在不同資訊系統領域所建立的數位資料形式，會產生頗大的差異，如國內圖書館自動化中傳技系統(TOTALS II)伺服器的 Big 5 碼就與 INNOPAC 系統伺服器的 unicode 碼不同，為了考量不同的使用者需求，以及資料庫存取資料的效率等因素，再加上不同資料供應商提供資料庫功能與格式上的差異，資料庫的標準化確實是一個相當困難解決的問題⁷。

因此，近年來圖書館界比較著重的是資料 metadata 的標準化建立，metadata 標準包括了 Dublin Core、XML (Extensible Markup Language)、FGDC (Federal Geographic Data Committee)、CDWA(Categories for the Description of Works of Art)、CIMI(Consortium for Computer Interchange of Museum Information)、TEI(Text Encoding Initiative)、EAD(Encoded Archival Description)、與傳統上圖書館使用的 MARC(Machine Readable Cataloging)等⁸。

再則，在建構數位典藏資料庫的時候，除了資料庫標準化和檢索功能

⁶ 劉清俊〈數位典藏及資訊共享的遠景〉，引自《中央研究院計算中心通訊》第 17 卷第 3 期，2001 年，頁 21。

⁷ 同前註 4 文，頁 41。

⁸ 同前註。

的考量之外，實也必須考量資料如何分類典藏，以便日後閱讀使用，若直接將未經整理的資料放於網路或資料庫之內，除將會減弱資訊化的原始目標⁹，亦會形成過多而無用的資訊垃圾(電腦其實有一樣基本特徵即是 garbage in and garbage out)。

(二)智慧財產權--數位產權(Digital Rights)

運用網際網路傳遞使用的數位資料庫，或許可以以公鑰加密系統(public-key crypto system)來確保資料之完整與機密性，但是無法防止使用者將取得的資料轉手、販賣提供給他人，而造成資料庫所有者在權益上的損失，因此，目前電子期刊(包括電子書)管控的方式，大致係以鎖 IP 或提供使用者帳戶密碼的方式，作為是否准許使用資料庫的重要依據，但是仍無法解決資料庫被轉手複製與傳輸等問題¹⁰。

再則，著作權的所有者，在法律的規範下，甚難對著作物主張「絕對的權力」(absolute power)，例如 amazon.com 網站上的書目資料(包括售價、書評、銷售情形)是否受著作權上的保障？還是只是事實的編輯，若它不是著作，那麼 amazon.com 可對它主張何種權力？在現行法律上的保護為何？是否該立法保護？對於是否該立法保護由事實編輯而成的資料庫，在美國曾引起諸多討論¹¹，似乎也尚未有一明確之答案。

四、結語

自二十世紀的六〇年代開始，由於資訊科技的快速進展，促使圖書館的經營方式產生了重大的轉變，特別是由於數位革命之後，使得「數位典藏」從傳統的「電子典藏」中脫穎而出；當然作為在數位時代的大學圖書館而言，必然應該努力建構各種數位典藏的使用權以及所有權，以滿足大學師生教學與研究所需的數位資源；但是筆者更關切的是，大學圖書館中的專業館員是否能夠真正瞭解各種典藏媒體、技術以及各種服務方式之間的互動性、互補性與關聯性¹²？誠盼吾等大學圖書館館員均能熟稔數位典藏的義涵及其運作方式，以祈成為建構與發展當代大學數位圖書館的重要礎石。

⁹ 參閱，同前註。

¹⁰ 同前註文，頁 43。

¹¹ 同前註。

¹² 參閱，同註 1 文，頁 92-93。