

行政院國家科學委員會專題研究計畫成果報告

網路化電子出版品之文件與格式探析：應用於學術出版之實作研究 **An Exploration and Experiment in the Networked Etexts and Formats for Scholarly Electronic Publishing**

計畫編號：NSC87-2415-H-032-025

執行期限：86年8月1日至87年7月31日

主持人：邱炯友 淡江大學教育資料科學學系

協同主持人：宋雪芳 淡江大學教育資料科學學系

Email: joyo@mail.tku.edu.tw

tinasong@mail.tku.edu.tw

一、中文摘要

電子文件檔案格式乃決定網際網路上電子出版品儲存與輸出之形式要項，同時也關係著電子出版品流通和閱讀效益。在眾多的網路化電子文件計畫裡，有著各種不同的檔案格式，不論針對文字、圖案、影像、聲音等，皆有其對應適合採行之檔案格式，單就出版品主體之文字檔部份而言，便存在著多種可應用的文件檔案格式。而檔案格式以家族型式存在或者可分成專屬與通用兩種格式。在眾多的國外計畫應用選擇中，由簡易解讀的ASCII文件格式到Postscript、Word、WordPerfect、PDF、LaTeX、XML、HTML和SGML等等格式，可接受的檔案格式或單一、或混合不一而足，其輸出呈現的文件效果也各具特色。事實上，單一格式或同時多項格式並行採用，在諸多現行網路計畫上皆可互見。瞭解現有網路化電子出版品所採行之文件格式景況，分析其間優點與困難而評估之，將可做為國內學術電子出版環境應用之參佐。

關鍵字：電子文件、檔案格式、電子出版、學術出版

Abstract

The essential element of storage, output,

circulation and reading in the networked electronic publishing rests upon the file format of electronic text (Etext) itself. The presence of networked Etext projects has a variety of file formats, which could include any format suitable for text, image, video, audio and sound. In regard to text, the main body of publication, many kinds of format have now been available in a way of family. They are ASCII text, Word, WordPerfect, Postscript, PDF, LaTeX, HTML, XML, SGML and so on. File format the current networked projects accepted could be either the only one or a multiple choice. To realize the circumstances of file formats of the networked publishing, as well as to analyze and evaluate those advantages and disadvantages, it is of great help to the need of scholarly electronic publishing.

Keywords: Electronic Text, File Format, Electronic Publishing, Scholarly Publishing

二、緣由與目的

二十世紀末，網際網路上活躍著許多以整理與儲存電子出版品為職志的相關計畫和機構，從美國1970年代早期的「古騰堡計畫」，乃至現今許多國外學術單位也在類似理念下，建立起具有自己風格與內涵

之各種主題的電子文件資料庫，例如：「線上圖書計畫(OBI)」¹、「人文電子文件中心(CETH)」²、「信息資源圖書館(Cause IRL)」、「期刊儲存計畫(Journal Storage Project, JSTORE)」、「英國的超期刊計畫(SuperJournal)」、「英國電子圖書館計畫(eLib)」、以及「國家數位化碩博士論文圖書館(NDLTD)」等等。這些計畫的執行說明了：網路化電子出版能替傳統的紙本出版品利用提供可行的解決方案；也提供圖書資料一種新的典藏環境。資訊的「擁有」與「取用」兩大觀念在這些電子文件計畫裡，同時得到了良好的典範與實踐。本研究計畫之目的與貢獻亦在利用整理出之理想文件檔案格式，就淡江大學教育資料科學研究所碩士論文，以及相關刊物做後續之實作。藉由該教資系所之網路伺服器展現應用成果，以做為私人出版社、圖書館、政府出版機制、大學出版社或其它大學與學術機構等之學術出版界，共同邁向網路化電子出版時代之整備工作。

本研究就網路上若干電子文件相關計畫做彙整，對於文獻資料電子化的處理及電子文件提交之過程中，各計畫的相對要求為何？以及電子化處理後所要提供的服務，及其處理資料所遭遇的困難與限制予以綜述。再就其採用之文件格式做探討和分析，瞭解箇中優劣與適性，以提供學界瞭解網路化電子文件相關計畫之背景與發展趨勢，並做為後續教學或研究者之參考。而文件檔案格式之比較和探析，為國內外學界新的研究課題做一結論暨建議，則更足以補充國內學界完整之網路化電子出版或電子文件之製作技術，增進電子文件輸出品質與選擇性。對有興趣建置類似電子文件計畫者，不論是大學、研究機構或出版實業界，都能明確提供選擇方案及參佐之架構。在學術論文電子化探討實作上，更有助於國內新大學院校之時代性思

¹ 參見 施純福，「Internet 上的 OBI 電子文件計畫」台北市立圖書館館訊，11 卷 4 期(民 82 年)，頁 42-49。

² 參見 王喜沙，「CETH 人文科學電子文件中心簡介」台北市立圖書館館訊，12 卷 1 期(民 83 年)，頁 82-88。

考與作為。

三、研究方法與步驟

因本研究計畫之內容旨趣之故，研究方法主以網際網路上之文獻蒐集分析與個案探討行之；並就若干電子檔案格式做實際模擬測試。其後續則據所研究結果與發現，期為淡江教資研究所之學位論文電子化提供建置之解決方案。文獻探討的目的在於為實作鋪路，與實作還有一段差距，唯有結合相關領域的專家發揮所長共同參與實際的建置，讓整個出版計畫在有目標、有組織的架構下，建置並提供使用，才是最有價值的。其實國外電子出版計畫之所以能提供寶貴經驗的原因，正因有實作的基礎，期待國內的圖書館界可以在電子出版的行列中找到可以發揮所長的一席之地。

四、研究發現與討論

就「電子文件計畫(Etext project)」之涵意而論，其文件建置規模與內容差異性頗大，規模既可涵蓋純屬個人製做的小網站；也可以歸屬於網路數位化圖書館式的大型資料庫之建置。在實際探討電子文件計畫之整體內涵時，是不能單就文字內容部分予以探討；也不能將文件中所包含圖表之展現與圖表檔案格式略去不談，兩者在電子文件的內容特性中，實互為表裡。綜觀而論，探討電子文件的檔案格式之目的是在於解決文件字型、字體與圖表的電子化標準，同時是試圖將文件中所有的視覺圖樣與符號(包含文字圖表)轉化成數位型式儲存，而且考慮到所儲存之資料在未來依舊可被取用。

對許多不同的文件建置者而言，電子文件之意義各有其程度上的解讀差異，例如：身為學術出版領域的大學出版社面對出版品電子化時，除了必須考慮到一般的電子文件需求，也必得考量諸如文件之數理公式符號與表格的呈現標準等問題，而對影像與動畫之需求就非當務之急。將正文部分與字型符號圖表等加以統整，應是

傳統平面出版邁向電子出版發展中最為迫切與棘手的問題之一。電子文件一般乃見於 Internet 與 CD-ROM 之媒體上。值得注意的是以網際網路上的 FTP、Gopher 與 WWW 等為載體的資料，才是電子文件的主要來源。狹義上，電子文件也排除了網際網路上其它的資訊傳遞與服務形式(例如 Email、BBS 或 UseNet 等)。時至今日，絕大多數的電子文件計畫也積極拓展文件的內涵與意義，邁向多媒體資料庫的應用領域。

除了在電子文件計畫中，有不同檔案格式採行背景介紹，在電子文件計畫的製作過程中，也發現圖書館或多或少參與其中，參考表 1 將各計畫圖書館涉入與否進行分析，圖書館主要是角色扮演是進行電子文件的規畫，所謂的規畫 (Planning) 是一個計畫在選擇方向 (Direction)，擬定目標 (Goals) 及目標 (Objectives)，和決定如何在行動進行時的主要作業程序，圖書館涉入電子文件計畫在規畫、與資料提供為主要角色扮演，再者由計畫中可以發現圖書館對於使用者利用資料的一些使用行為，也可為設計電子文件計畫提供重要的參考。

檔案格式與文件出版有著密不可分的關係，不同的檔案格式是因文件在出版流程中的需求不同而形成，由調查顯示，這些需求的考量包括：學科領域的差異、原始作者的習慣或喜好甚至是出版單位為統一出版文獻而訂定的限制等等。從出版流程與檔案格式的關係中可以發現，現存的檔案格式多以「家族」的型式並存，在一個資訊化的時代，資訊分享是一件非常重要的事情，而資訊是否能夠真正達到分享的境界，標準是最大的關鍵所在，如果沒有標準的建立，則資訊分享無異於空談。先就資訊標準來說，目前存在許多不同檔案格式的標準包括 TeX 家族、PDF 家族、SGML 格式家族、多媒體檔案格式家族，從目前幾個發展的體系來看，仍屬於標準與非標準的紛爭時期，標準化格式訂定的最大困難是在不同的利益團體及派系中取得一致的看法。因此最後大家必須妥協，

而其結果通常是紙上談兵，由於標準的太複雜或太抽象加上沒有適當的輔助編輯工具，使得沒有人願意去實行它。建立一些有效的標準目的在於可以實際的解決問題、滿足使用者的需求、足夠一般化作為將來使用及交換之考量，一旦由市場決議的格式為標準組織所接受，即形成一種以市場為基礎的發展策略，如 Adobe 公司建立 PostScript 的標準。

分久必合，合久必分是不斷循環的現象，資訊界亦不例外，從前面格式的介紹中可發現各類型檔案以家族的型式結合存在，每一個家族有其各自的標準，軟體廠商可依既定的模式要求不斷研發更進步的軟體以滿足資訊時代中使用者的多項需求，而廠商間市場競爭造成軟體也帶來技術的進步。而看似獨立的各家族如今似乎走向合併，在以標準格式為主力的發展下漸漸形成模糊的界線，以目前 XML 的發展為例，它是依循 SGML 的標準加以修改建立的，之下研發的 MathML 格式則是一套描述數學文件的機讀格式，所提供的多項功能均結合了 HTML 與 TeX，彌補了 TeX 無法滿足電子文件與 SGML 標準無法滿足數學特殊符號的雙重缺點，換言之，這說明了 TeX 家族與 SGML 家族的已朝向合併發展；此外的例子還包括 Web TeX 轉換成 MathML 的轉換程式；IBM techexplorer 可瀏覽科學技術方面的文獻並可做為 LaTeX 與 MathML 的轉換器等等。由此可見檔案格式各家族體系發展中，也開始兼顧其他檔案格式家族的發展，而走向單一模式的整合趨勢。

例如多媒體家族就包括 GIF、JPEG、MPEG 與 TIFF 等格式，同一家族遵循著一定的標準，對使用者而言，這似乎是一個好現象，他們可以在該家族檔案群中選擇所需要的軟體來製作各樣的文件，由於軟體的選擇取決於使用者，於是這間接促成了軟體公司的競爭，為了使所研發的軟體，能滿足多數使用者的需求、獲得多數使用者的支持與使用，可想而知的是新的軟體將不斷因競爭而產生，造成軟體開發上的一種進步的動力；對使用者而言，又

成為製作文件上的福音，整個研究生態在這股良性循環中，新起的研究人員，一代比一代花更少的時間用在製作文件、排版上，而能將時間與人力都用於所研究的領域中，讓科技不斷的進步。

上述的情形有時會造成檔案格式的過多，對出版社是一種負擔，畢竟對讀者而言，好用的軟體並不一定是當時使用的趨勢，對出版社在進行排版前的轉檔工作而言，是一個很大的負擔，這也是為什麼出版者會針對繳交的文件給予格式上的部分限制。轉檔並不是一件簡單的工作，從本研究之實作經驗中得知，轉檔過程中的問題有時很難掌握，系統的不穩定性使得同樣的工作環境和測試文件會產生不同的測試結果，卻無法找出原因；就算發現了問題，也因為對整套軟體設計架構的不了解以及缺乏電腦方面的專業知識而無法排除，更重要的是測試與障礙的排除相當耗費時間與人力，對只有圖書館專業背景的人員而言，更是如此。於是，當圖書館界加入電子出版的行列時，是否有能力全部包辦？究竟又該如何使原本的專業得到最有價值的發揮？可能是一個值得重新思考的問題。

合作是國外成功的電子出版計畫共同的特徵，合作可以結合多樣的專業，例如：出版社、大學、圖書館、軟體設計者與軟體供應商的力量共同規畫完成，達到單一組織所無法滿足的目標；同時也分攤了所需的人力、時間、金錢等資源，對任一參與的組織而言，無非是一種精簡的組合。在科際整合的時代下，資訊界的合作是必然的，身為資訊界一員的圖書館也不例外，倘若能在電子出版計畫中，體認合作的重要性，邀請相關領域協助並參與規畫，其成果將是大家所樂見的。

五、結論與建議

一個電子文件計畫的建立都有其動機存在。這些計畫按其產生之背景、目的、以及欲解決之問題，而賦予計畫所具之特質與風貌，並藉以描繪出計畫建置的理念。透過對於電子文件計畫和建置之相關

模式的瞭解，將有助於更明確地反映網路化電子出版的主要活動與價值。在國外實例中，電子文件計畫常由各種學術機構所經營，處於電子出版資訊時代的「大學圖書館」與「大學出版社」兩者對於出版活動的參與已多所交疊，本文除提供出版界、圖書館、學術社群、政府單位之整體角色與其經營之省思外，茲綜結以下若干議題暨現象予以廓清：

(一) 電子文件計畫之永續發展

有前瞻性的電子文件計畫之發展是永續性的。然而，將出版當做是市場的供給與需求關係來看，到底出版商是否會完全投注於電子出版的發展？則答案主要還是要看使用者是否有這樣的需求。姑且就目前資訊傳播之發展而論，保持紙本資料、電子出版文件，抑或兩項平行發展？同樣的資料是否有傳統紙本版、電子版？答案便仍歸結於市場需求與業者之企圖心。反觀學術界所建置的電子出版文件，其動機多賴於典藏或檢索之所需，市場與利潤考量卻在其次。目前電子文件計畫，有其收錄範圍，所以如何就各電子文件之間加以整合、連結，以達到使用者最方便的利用，這才是最佳發展之指導原則。

(二) 多媒體技術之研發

電子文件一向有其典藏、儲存、編排、查檢與輸出的多重使用特性，若可以搭配多媒體效果來呈現不同主題特色的表現，則可增強文本的說服力，也可以使作者與讀者間進行互動性的交流。但有些文獻對多媒體的表現方式需求不殷，僅需要文字的表達，或簡單的圖表展示即可，而這些主題就不需要多媒體的配合。目前以多媒體技術為主的電子出版仍在持續發展中，多媒體技術與產品是未來電子文件的發展趨勢，不過多媒體出版技術的研發，最佳之解決方案應是整合出版者、軟體開發者、圖書館(數位典藏單位)之力量，共謀多媒體技術的合作發展，如此才能在出版、技術、資訊使用等，多方面考量之下才能建置完美的人文科技整合之系統。³

³ “Is it possible to write for E-text and Print at the

(三) 多種檔案格式之並存

綜合來看，不論現況是檔案格式以家族型式存在或者可分成專屬與通用兩種格式，不論是專屬格式或通用格式，並無統一的必要性。使用者針對自己的學科領域撰寫文獻、製作文件，最方便的方式是根據文件的需求來使用其熟悉的工作環境與軟硬體，如果發展一套統一的、並整合文理學科的強大檔案格式，對人文科學的文字檔而言，同樣迫使使用者必需具有等級較佳的電腦設備，來應付該軟體的龐大體積以擁有較快的執行速度，造成無形中的浪費。從檢索的角度來看，即便今日已是一科際整合的時代，但文理學科領域的界限仍存在，一套整合性的檔案格式的確可以節省檢索不同資料庫，而需學習操作不同系統的時間，但從另一個角度觀之，有多少人在檢索資料時，對跨文理學科界線資料的需求是一致的？在整合學科領域的

一個電子文件(資料庫)計畫之首務在於文件檔案格式之選擇與確立。如何尊重讀者的選擇性與便利性，而提供多種文件檔案格式並存之選取，亦不失為值得嘗試的作法。而身為一位電子文件計畫之建置者而言，應有責任來隨時平衡這些顧慮與需求，同時也不妨藉此教育使用者，讓他們日後得以自己做判斷與選擇；或至少也懂得欣賞每個文件計畫背後的複雜性。本研究後續的實作正可掌握上述諸原則與精神，建立一致而適切的例證。

六、參考文獻

Wusteman, Judith. "Formats for the Electronic Library.", available from <http://www.ariadne.ac.uk/issue8/electronic-formats/>; Internet.

資料庫中，可想而知的是檢索時間的增加，而且這樣的投資是沒有意義的。建置

same time?" see Elements of Electronic Text Style;
available from
<http://www.ind.net/Internet/guides/estyle10.txt>;
Internet.

表 1 圖書館涉入電子文件計畫分析一覽表

圖書館是否涉入 計畫名稱	規畫	參與資料 提供	技術支援	說明
ARIADNE	✓	✓	✓	出版圖書館學與資訊科學的期刊 The UK Office for Library and Information Networking(UKOLN)以及 Abertay 大學合作，圖書館以出版社的角度參與。
CAUSE	✗	✓	✗	高等教育資料儲存及利用計畫
CETH	✓	✓	✗	由 STAFF 分析主要計畫主導人背景專長為英國文學。管理電子文件人員為擅長寫 SGML、XML、JAVA 及相關工具者。
eLib	✓	✓	✓	英國電子圖書館計畫
IJEJ	✓	✓	✓	IJEJ 是一合作型計畫，由 University of Birmingham、University of Leeds、University of Manchester、University of Oxford 四所大學圖書館合作完成，他們同時也是 CURL(Consortium of University Research Libraries)、RLG(Research Libraries Group) 的會員。
JSTOR	✗	✓	✗	期刊儲存計畫
Project Open Book	✓	✓	✓	耶魯大學企圖探討館內資料微卷轉換成數位化影像的可能性，包括：金錢、評估、選擇的考量及資料的加強檢索及傳播之探討。
SuperJournal	✓	✓	✗	出版社、大學、圖書館、軟體設計者共同參與規畫與技術提供。技術支援則由軟體供應商設計提供。
TULIP	✓	✓	✓	由出版社 Elsevier Science 公司和美國九所大學合作完成。
VWWP	✓	✓	✓	由一名發起學生、一名館員與一位資深的維多利亞文學專家，合作完成。
古騰堡計畫	✗	✓	✗	
維吉尼亞大學 早期小說電子化 計畫	✓	✓	✓	由維吉尼亞大學圖書館電子文件中心負責處理資料電子化及資源的檢索。