

ODF和OOX文件格式架構及其展現之比較分析

歐陽崇榮

助理教授
淡江大學資訊與圖書館學系
E-mail : cjouyang@mail.tku.edu.tw

薛志峰

研究生
淡江大學資訊與圖書館學系
E-mail : 694070151@s94.tku.edu.tw

摘要

本研究主要在探討兩套新興的文件格式標準ODF和OOX；並且透過支援ODF和OOX的辦公室套裝軟體實際對文字檔、試算表和簡報檔進行實驗觀察及比較分析。發現其間有所差異並存在著相容性的問題。

關鍵詞：電子文件，辦公室文件，開放式文件格式，OfficeOpenXML

前 言

英國大英圖書館在2004年出版的「Publishing output to 2020」報告指出，未來大英圖書館的館藏資源都將以電子化出版或將紙本資源電子化，到2020年電子化的館藏將達到館藏量的40%（註1）。美國國家檔案及文件署（U.S National Archives and Records Administration，簡稱NARA）於2006年在其未來發展策略中提到由於現今的電腦科技逐漸進步，未來的電子文件將會呈現指數成長的趨勢（註2）。又由於電子文件逐漸成為目前使用者最常使用的文件類型，電子文件的長期保存也逐漸受到重視並成為政府機關、各種文化機關及私人企業或一般使用者之重要議題。電子文件保存所面臨的問題包括了電子文件的檔案大小、保存格式、保存媒介、保存環境等等。此外，在保存電子文件時也經常會碰到因為電腦系統升級而無法讀取的情況，或者在不同作業系統、軟體環境下

2007/01/31投稿；2007/04/10修訂；2007/04/11接受

導致文件無法正常開啟、閱讀等等風險。為能解決電子文件長期保存之問題，其中一個關鍵因素就是電子檔案的格式須要採用公開標準。

辦公室文件是電子文件的一種，主要指使用文書處理 (Word-Processing)、試算表 (Spreadsheet) 和簡報檔 (Presentation) 等辦公室軟體所產生的原生數位 (Born digital) 文件，也是最多人最常使用到的文件 (註3)。辦公室文件目前的主流是以微軟公司開發的「Office」套裝軟體為主流 (註4)，屬於一種封閉式軟體 (Closed software)，並採用二進制檔案格式 (Binary file) (註5) 來保護其軟體的著作權。因此，要成為長期保存之文件格式有其困難、窒礙之處。然而，在2006年出現了二套採用公開的延伸標示語言 (eXtensible Markup Language，簡稱XML) 來規範所有辦公室文件格式的標準：一為開放文件格式標準 (Open Document Format for Office Applications，簡稱ODF) (註6)，一為微軟公司的Office Open XML (簡稱OOX) (註7)。有了公開標準的文件格式，政府、企業和一般使用者在辦公室文件格式及其應用軟體採用之策略和方式上是否會有所影響和改變？是值得探討的課題。

本文從實務的角度著手，除了探討數位保存和辦公室文件的關係之外，並且分析ODF和OOX兩套開放標準之間的異同，利用實作的方式針對文書處理、試算表和簡報檔等辦公室文件檔案在採用ODF標準或採用OOX標準的辦公室套裝軟體之間進行轉換，試著找出轉換之間所產生的問題。

二、辦公室文件格式之探討

(一) 電子文件與數位保存

所謂電子文件，根據「電子簽章法」第二條指出：「電子文件：指文字、聲音、圖片、影像、符號或其他資料，以電子或其他以人之知覺無法直接認識之方式，所製成足以表示其用意之紀錄，而供電子處理之用者。」(註8)大陸「國家標準電子文件歸檔與管理規範」(GB/T18894-2002)對電子文件的定義是：「指在數位設備及環境中生成，以數碼形式存儲於磁帶、磁片、光碟等載體，依賴計算機等數位設備閱讀、處理，並可在通信網路上傳送的文件。」薛理桂教授綜合各國對於電子文件所下的定義，歸納出以下兩項重要的特質(註9)：

「1. 電子文件是一種透過電腦來儲存、讀取的數位格式文件資料。2. 電子文件主要的產生方式有三種，第一種為特定電子系統下所產生內涵機讀格式電子檔案；第二種則是不同類型的電腦應用程式或資料庫所產生之電子檔案；另一種則為文件數位化檔案形式，如將傳統文件透過掃描或數位翻拍轉換為電子文件的方式。」

在電子文件大量快速成長的同時(註10)，數位保存也變得相當重要。Ali-son Bullock認為數位保存或數位典藏 (Digital archiving) 的意義在於能夠制定策

略進行長期保存原生數位資料並存放在適當之媒體上(如CD、DVD等等)或將類比資料轉成數位型式並確保長期可以使用(註11)。目前電子文件的長期保存方法大致有：系統保存、更新、轉置、模擬、標準化、封裝、重複或分散儲存、印成紙張或其他瀏覽媒體……等(註12)。而在上述這些長期保存方式或策略當中，如封裝、轉置、模擬和標準化等多半都包含有關文件格式的標準制訂和文件交換轉換的相關格式。封裝策略當中所強調的就是封包內的資料欄位描述和交換封包的註記和標示；轉置策略也強調要選擇一定的格式作為轉置保存的依據；模擬策略除了選擇某個特定的電腦系統硬體作為模擬的依據外也很強調標準與文件格式；標準化的重點在於資料庫、字碼標準、詮釋資料、數位化格式、檢索等等都制定相關的統一標準(註13)。數位保存除了要制定相關策略和保存相關的軟硬體或載體之外，文件之間的格式標準也是長期保存策略所關心的議題。

(二)電子文件格式

電子文件格式指的是用來辨識每一筆位元資料的檔案名稱和用途，一個電子文件格式說明了所有電腦程式讀取和撰寫檔案所必需的資訊(註14)。電子文件格式在電子出版系統中，設計者經常要將處理過的文件依其不同特性而儲存成不同的檔案格式；而由於電子文件經常會出現表格、圖像、多媒體元件、特殊效果……等，因此便有各種使用方便且功能完善的電子文件格式。例如：文字導向型文件常見的TXT、DOC(Microsoft Word的格式)；影像導向型文件的BMP、GIF、TIFF、JPEG、PNG；多媒體影音導向型文件的WAV、MIDI、MP3、AVI、MPEG等等，可謂琳瑯滿目，種類眾多(註15)。一般而言，電子文件格式都會具有至少一種檔案格式，甚至在同一份檔案之中，都有可能包含眾多不同的檔案(如超連結或嵌入圖片等等)(註16)，根據邱炯友的研究大致可分成：文字型、圖形影像檔案、多媒體檔案或稱動態影像檔、標準描述語言檔案(註17)。黃國斌則認為主要的電子文件資料可分為三種類型：1.文字處理型文件(Text-base documents)，2.多維文件形式(Multi-dimension documents)，3.多媒體文件形式(Multi-media documents)等(註18)。加拿大圖書館及檔案館(Library and archives Canada)對於電子檔案之格式規範成：數位聲音(Digital audio)、數位靜態影像(Digital still imagery)、數位影像(Digital video)、文件-本文(Documents-textual)、電子郵件(Email)、地理空間資料(Geospatial data)、結構化資料-資料庫及電子試算表(Structured data - Databases and spreadsheets)及技術繪圖(Technical drawings)等八種類型(註19)。

然而，根據Green在2005年針對英國赫爾大學(University of Hull)研究人員所作的調查報告顯示，使用者最常使用的電子檔案格式前三名分別為文書處理檔案格式、簡報檔檔案格式和試算表檔案格式，都屬於一般所稱辦公室文件

(註20)。歐盟IDABC (Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens) 在2003年針對辦公室套裝軟體的市場佔有率調查顯示微軟公司的辦公室套裝軟體佔有率達94% (註21)。在顯示出辦公室文件已經成為使用者最常使用的檔案，而且大都使用微軟公司的產品。

Marco Fioretti認為電子文件應用軟體逐漸成為使用者處理、儲存和呈現的工具，而應用程式的呈現和檔案格式是一體的，文件的格式如果可以讓每一個使用者都容易了解，如：XML格式，而不使用二進制檔案格式則使用者只要了解XML欄位的意義即可自行編寫應用。未來的文件格式將會以XML格式作為主流，因為XML格式不同於以往的二進制格式，是採用純文字檔描寫標示語言，容易閱讀和資料交換，也有助於文件格式的開放和標準化的產生，像是ODF和OOX格式的誕生(註22)。所以電子文件的保存不在於應用軟體的演進或升級而在於文件本身的格式是否為開放格式。由於XML文件格式是一套開放式的格式，所以可作為數位保存文件格式時所採用。Gartner公司即預估2006年XML-enabled的電子表單使用率至少會成長兩倍，2007年將會有40%的知識工作者採用XML-aware內容創作工具(註23)。Forrester研究公司則預估到2008年，XML將成為檔案管理的主流文件格式之一(註24)。

(三) Open Document Format for Office Applications(ODF)

ODF的誕生，最早是由Sun Microsystems公司在其產品「StarOffice」所推動的一個計畫，2000年將其原始碼交給GNU(GNU's Not Unix)的LGPL(Lesser General Public License)使用許可授權下被應用(註25)。2002年時，Sun Microsystems公司宣稱ODF是一種公開的，利用XML語法的辦公室應用軟體標準規格書，並決定交由OASIS(Organization for the Advancement of Structured Information Standards)制訂相關的標準(註26)，OASIS是一個國際性非營利組織，主要目的在發展一些電子商務標準，提供資料和資訊內容交換技術。OASIS於2002年11月成立技術委員會，研究發展一套以XML語法架構為主的文件格式說明書，其目的是建立一套可以使辦公室軟體進行資料互操作的標準(standard data interoperability)(註27)。歐盟技術委員會TAC(Telematics between Administrations Committee)在2004/05的報告書，建議各國採用以XML為基礎的公開檔案格式標準，同時建議OASIS將ODF提報到ISO成為ISO的標準(註28)。於是ODF在2005年5月正式公布為OASIS的標準，並於2006年5月正式成為ISO標準「ISO/IEC26300」(註29)，且在2006年10月推出了1.1版(註30)(至2007年1月止之最新版，預計2007年8月會推出1.2版)。

根據OASIS推出的「open document format for office applications (opendocu-

ment) v1.1」報告書指出(註31)，這份文件格式是利用XML schema架構所制定出的一套辦公室軟體和語意描述。這個schema很適合用在辦公室文件處理上，包括文字檔、試算表、圖表等等，並不侷限於任何一種文件形式。這個schema提供了高水準的資訊幫助編輯文件。它定義了適當的辦公室文件並提供了友善的使用XSLT作轉換或其他的XML基礎工具。

(四) OfficeOpenXML(OOX)

1999年，微軟(Microsoft)公司在其產品「Microsoft Office Excel 2002」中採用XML格式，稱為SpreadsheetML；之後又在「Microsoft Office Word 2003」產品採用更為完善的XML格式，稱為WordProcessingML。但這兩種XML格式仍然有許多限制，不能算是完整地開放格式(註32)。OOX標準是由微軟公司在2006年推出的文件格式標準，於2006年12月通過ECMA標準組織認證，成為ECMA Standard 376，並同時申請ISO標準認證中(註33)。

根據ECMA在2006年10月技術報告(White paper)(註34)指出：OOX是一套開放標準，可以使得文字檔、簡報檔及試算表等文件可以自由地在不同的應用程式及/或不同的電腦系統平台作業。讓組織可以創造使用該格式標準的軟體，也可以幫助企業和政府更容易了解文件的本質和格式，教育者或創造者可以學習該格式。最後，所有的使用者都可以從XML標準中得到好處，包括穩定性(Stability)、保存(Preservation)、互操作性/互通性(Interoperability)和持續發展。OOX標準化由ECMA Technical Committee45委員會(簡稱TC45)所推動與維護，委員會的名單包括：Apple、Barclays Capital、BP、大英圖書館、Essilor、Intel、微軟、NextPage、Novell、Statoil、Toshiba和美國國會圖書館等。OOX的目的就是要忠實地呈現以往採用微軟公司所制訂的二進制格式底下所產生的文字檔、簡報檔和試算表，將過去的格式用XML語法的方式表現出來並加以延伸，提供詳細的使用說明和增加互操作性(註35)。

(五)文件的呈現

ODF和OOX都提供了一套開放式的文件格式讓辦公室文件之間能夠利用XML的方式進行文件的描述，並用於文件之間的交換、溝通和傳遞。兩者也都採用了XML Schema和RELAX NG Schema等XML標示語言標準來讓辦公室文件能夠完善地描述該文件本身，同時也能夠藉由這些公開標準讓辦公室文件內容得以持續長久保存。利用這兩套標準作為文件格式的模板，也可以減少文件之間無法交換的衝突。

除了開放或封閉格式之外，對於政府機關及企業乃至使用者而言，透過軟體呈現文件的外觀、真實性也會影響到辦公室文件保存的選擇。不論ODF或OOX都是開放式的文件格式標準，但在不同的辦公室文件軟體，其呈現上是否

會有差異也是值得注意的。產生差異的原因除了軟體本身的功能不同之外，兩套標準之間的相容性、互操作性也有可能影響文件產生呈現上的差異。本研究即針對兩套分別支援ODF和OOX的軟體觀察其產生的差異並試著找出問題所在。

三、研究方法與實施

本研究透過實際辦公室文件進行實驗觀察及內容分析法。主要利用「Open Office.org2.0」和「Office2007BETA版」這兩套分別支援ODF和OOX標準的辦公室應用軟體作為實驗的平台，並於95年11月28日到12月18日以淡江大學教學支援平台網站(<http://www.tku.edu.tw/list/courselist.htm>)上隨機蒐集原生微軟Office2003檔案，總計文字檔120份、試算表120份、簡報檔120份，接著在兩套軟體中的「Writer」、「Calc」、「Impress」和「Word」、「Excel」、「PowerPoint」開啟比較其差異性。並採用漸進式觀察法，由空白文件、單一頁面文件、兩頁頁面到五頁頁面文件以及分別用OpenOffice.org2.0和Office2007BETA各建立三個範例檔(文書處理、試算表、簡報檔)並將先前統計常用的各種屬性置入，如：文字、圖片、表格、註腳、文字方塊等等，同時分析其封包架構。

四、結果與分析

針對研究方法設計的實驗進行分析，分別從軟體呈現和XML的封包分析兩方面探討不同標準對應在不同軟體上的差異。

(一)軟體呈現分析

1. 文書處理

本研究將觀察到的差異分為：段落(行距、間距)、字體屬性、表格、圖像、項目符號、標點符號、引文目次、其他等八個類型，分別敘述如後：

(1)段落(行距、間距)：

本類型包括段落中之行距、字元間距，以及文字在段落中位置的呈現等差異。發現的差異包含了有部分行距、字元間距的不同，段落前空兩格在Writer 2.0上變成段落前空一格，以及部分字句的位置不同等6項差異。

(2)字體屬性：

本類型包括文字的體例、外觀、粗細、斜體等差異。發現的差異包含有字體(羅馬數字I)的形狀不同、文字的解析度不同、不支援某些字型，以及粗體字變成非粗體字等4項差異。

(3)表格：

本類型包括表格的屬性與表格內容的呈現等差異。發現的差異包含有部分

表格內文字的位置不同、部分儲存格的大小不同、表格中的斜線不見、頁面內容被切割等4項差異。

(4)圖像：

本類型包括圖像的位置、大小、解析度等差異。發現的差異包含有圖像的解析度不同、圖案所在位置不同、圖案形狀不同等3項差異。

(5)項目符號：

本類型包括項目符號的外觀、支援情形、位置等差異。發現的差異包含有部分項目符號圖形大小不同、項目符號出現中文數字，如一會變成阿拉伯數字1、項目符號與文字的距離不同等4項差異。

(6)標點符號：

本類型包括標點符號的大小、位置、形狀等差異。發現的差異包含有標點符號「一」變成「-」、中文標點符號位置有異、標點符號的形狀不同等3項差異。

(7)引文、目/表/圖次：

本類型包括目次超連結部分的差異，包括位置、外觀、目次文字內容、超連結格式等。發現的差異包含有目次超連結格式錯亂、表次後面的數字不見等2項差異。

(8)其他：

不屬於以上七類之差異者，歸入本類。發現的差異包含有註腳的分隔線不同、線段消失、冒出多出來的文字、標題前面的點消失、出現不知名紅色小點等6項差異。

例如在文書處理的目次超連結方面會有格式竄亂的情形產生，參見圖1所示(左為OpenOffice.org2.0、右為Office2007BETA)：

目次	目次
第1章緒論.....	第一章 緒論.....1
第1節研究目的.....	第一節 研究目的.....2
第2節研究問題.....	第二節 研究問題.....2
第3節研究方法.....	第三節 研究方法.....3
個案研究法 (Case Study) 是對單一的人或事進行深入具體	第四節 研究範圍與限制.....5
的研究，研究的人或事可能是典型的，也可能不是典型的	第五節 名詞解釋.....5

圖1 文書處理檔差異呈現圖

2. 試算表

本研究將觀察到的差異可分為文字、座標軸、圖例、圖表等四個類型，分別敘述如下：

(1)文字(字向、符號、位置)：

圖表座標軸上的座標軸名稱文字方向會自動從直向變成橫向；貨幣符號會變更或消失不見；文字與百分比的顯示位置會有所不同；此外，圖表中的數學運算式會消失不見等4項差異。

(2)座標軸

圖表中各項顯示資料的順序會左右或上下顛倒；圖表中座標軸顯示的組距會變更(可能因預設的不同)，進而導致圖形變形；圖表中座標軸欄位辨識錯誤等10項差異。

(3)圖例

圖表中顯示的圖例有變大變小情形；圖表中的圖例辨識錯誤，無法讀取到正確欄位。此外，圖表中標示的圖案亦有所不同；計有4項差異。

(4)圖表

圖表中沒有原生資料的顯示圖案，會自動呈現相似的圖案；原生資料的比例顯示方式會有所不同；資料顯示項目之順序會產生相反現象；內容顯示方面有錯誤產生；其圖表背景顯示顏色也會改變；完全無法顯示股票圖；會顯示與原生資料圖表完全不同的圖表格式；圖表內容會較原生資料中的圖表顯示不清；計有15項差異。

例如在圖形圖呈現會有截然不同的情形產生，參見圖2(左為OpenOffice.org2.0、右為Office2007BETA)：

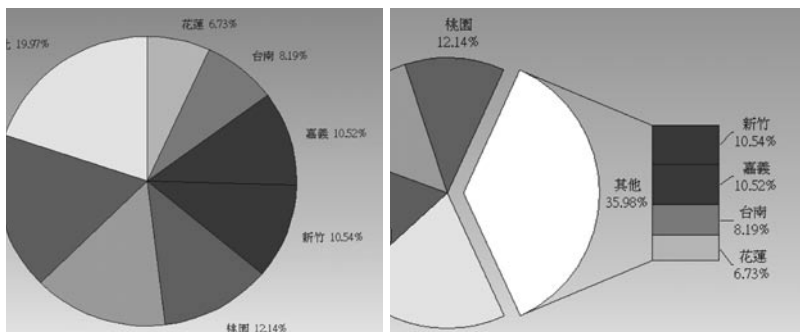


圖2 試算表呈現差異圖

3. 簡報檔

本研究將觀察到的差異分為：段落(行距、間距)、字體屬性、表格、圖像、項目符號、標點符號、動畫、其他等八個類型，分別敘述如下：

(1)段落(行距、間距)：

本類型包括投影片中文字的位置、間距、行距等差異。發現的差異包含有文字行距差異、字句位置不同、段落開頭位置不同、文字內容超出畫面、字元間距不同等6項差異。

(2)字體屬性：

本類型包括文字的體例、外觀、粗細、斜體等差異。發現的差異包含有粗體字變成非粗體字、字型的差異(Arial變成新細明體、標楷體變成新細明體)、文字陰影不同、字體大小不同、斜體字變成非斜體字等9項差異。

(3)表格：

本類型包括表格的屬性與表格內容的呈現等差異。發現的差異包含了表格的框線粗細不同、表格的底色不見、文字內容與表格重疊等3項差異。

(4)圖像：

本類型包括圖像的位置、大小、解析度等差異。發現的差異包含了圖案解析度不同、圖片內容差異、圖片色彩不同、圖片位置不同、部分圖案消失、背景解析度不同等7項差異。

(5)項目符號：

本類型包括項目符號的外觀、支援情形、位置等差異。發現的差異包含了項目符號不見、某些項目符號不支援、項目符號位置不同等4項差異。

(6)標點符號：

本類型包括標點符號的大小、位置、形狀等2項差異。發現的差異包含了中文標點符號與其他文字間距變寬、某些標點符號不見等差異。

(7)動畫：

本類型包括動態圖片、插入的動作、換頁時的動作等差異。發現的差異包含了GIF動畫的差異、動畫方式的不同、動畫造成的內容遺漏、動畫沒有顯現等4項差異。

(8)其他：

不屬於以上七類差異者，歸入本類。發現的差異包含了超連結的顯示不同、日期呈現方式不同、背景不同等4項差異。

例如在文字陰影不同，參見圖3(左為OpenOffice.org2.0、右為Office-2007BETA)：



圖3 簡報檔呈現差異圖

(二)XML 封包架構分析

1. ODF 和 OOX 空白封包架構比較

這兩套格式標準都是採用 Zip 技術將文件檔案進行壓縮。壓縮的好處不但可讓檔案大小變小，也可藉由封裝的方式將文件所含的各種資訊包含在裡面。研究者實際利用 OpenOffice.org2.0 和 Office2007 BETA 軟體建立空白文件（文書處理、試算表、簡報檔各一份）並利用 Zip 軟體開啟。

從 ODF 的封包中可看出文書處理、試算表和簡報檔的封包架構都是相同的。主要有 Configurations2、Pictures、Thumbnails、META-INF 四個資料夾，以及 mimetype、layout-cache、content.xml、settings.xml、meta.xml、styles.xml 等六個檔案。其中 Configurations2 資料夾是組態；Pictures 資料夾是用來存放圖片檔；Thumbnails 資料夾是縮圖；META-INF 是用來描述整個封包的架構，mimetype 檔案是用來描述文件的類型；layout-cache 是快取檔案；content.xml、settings.xml、meta.xml、styles.xml 四個檔案是描述文件的資訊、規格、設定和內容等。

OOX 文書處理、試算表和簡報檔的封包架構則有所不同。文書處理的封包分成三個資料夾、一個檔案；三個資料夾分別是 rels、docProps、word 和一個檔案 [Content Types].xml。試算表的封包分成三個資料夾、一個檔案；三個資料夾分別是 rels、docProps、xl 和一個檔案 [Content Types].xml。簡報檔的封包分成三個資料夾、一個檔案；三個資料夾分別是 rels、docProps、ppt 和一個檔案 [Content Types].xml。如圖 4 和圖 5 所示：

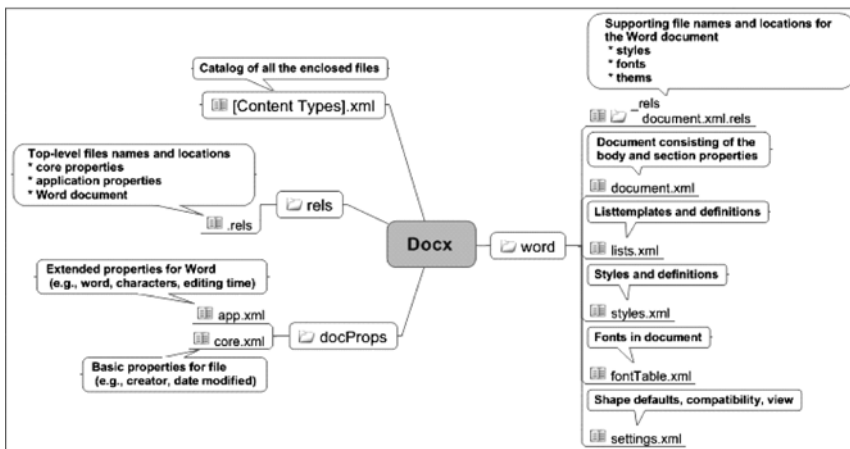
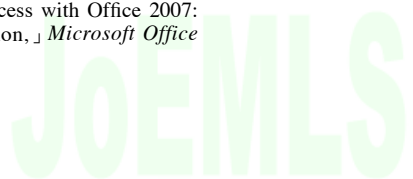


圖4 OOX文書處理封包架構圖

資料來源：Sherry Kappel, 「Document Success with Office 2007: An ILTA '06 Encore Presentation,」 *Microsoft Office* (2006): 16.



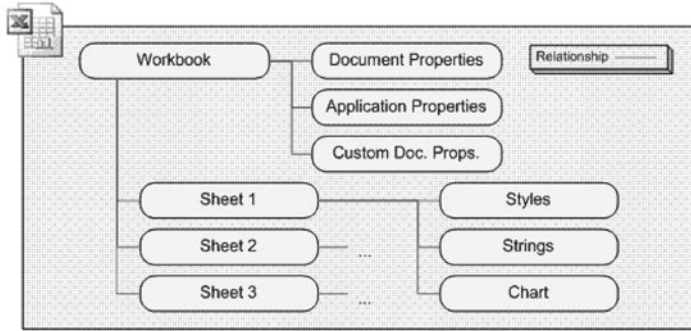


圖5 OOX試算表封包架構圖

資料來源：Frank Rice, 「Introducing the Office (2007) Open XML File Formats.」<http://msdn2.microsoft.com/en-us/library/ms406049.aspx> (accessed January 28, 2007).

從兩者的封包結構可以看出，ODF封包是分成四個資料夾和六個檔案，而OOX封包則是三個資料夾和一個檔案，OOX封包本身在不同類型文件中也不盡相同。ODF採用MIME的方式描述檔案的類型，而OOX則採用XML格式當成檔案類型的描述：即rels.xml。詮釋資料部分，ODF是用單一的XML檔案作描述：meta.xml，OOX則用兩個XML檔案作描述：app.xml和core.xml。ODF利用manifest.xml描述整個封包的架構，而OOX用[Content Types].xml作為整個封

表1 ODF與OOX空白文件封包比較表

項目	ODF	OOX
封包結構	四個資料夾和六個檔案	三個資料夾和一個檔案
檔案類型描述	採用MIME的方式	採用XML
詮釋資料	以meta.xml之XML檔案描述	以app.xml和core.xml兩個XML檔案作描述
封包架構描述	以manifest.xml之XML檔案描述	以[Content Types].xml之XML檔案描述
文件主體	以content.xml之XML檔案描述	以word、xl、ppt檔案夾描述
設定及型式	以style.xml和seeting.xml之XML檔案描述	以word、xl、ppt檔案夾描述
圖片檔位置	在Picture及Thumbnails兩個資料夾下	在Media資料夾下
附註、註腳、工作簿、表格型式、呈現等特色	無	有endnote.xml fontTable.xml、footnotes.xml、workbook.xml、presProps.xml、tableStyles、viewProps.xml、theme1和webSetting

資料來源：研究者整理

包的內容描述。ODF的文件主體是用content.xml描述，而OOX則是對應不同的文件類型而分成不同的資料夾作描述，如：word資料夾、xl資料夾、ppt資料夾等等。style.xml和seeting.xml以及圖片檔的位置設定是兩者皆有的部分，只是封包的位置不同。endnote.xml、fontTable.xml、footnotes.xml、workbook.xml、presProps.xml、tableStyles、viewProps.xml、theme1和webSetting是OOX特有的(如表1所示)。

2. 逐頁分析

在分析兩者的空白文件封包架構之後，研究者利用一頁文件、兩頁文件和五頁文件進行文書處理和簡報檔逐頁封包架構比較，試算表則增加一個工作表、兩個工作表和五個工作表(OpenOffice.2.0空白試算表的預設值為一個工作表，Office2007BETA空白試算表的預設值為三個工作表)，查看是否有所差異。結果如下：

(1)一頁文件和增加一個工作表

ODF的文書處理、試算表、簡報檔和OOX文書處理封包架構不變。OOX試算表多了xl\worksheets路徑下的sheet4.xml一個XML檔案。OOX簡報檔多了ppt\slides路徑下的slide1.xml；ppt\slides_rels路徑下的slide1.xml.rels兩個XML檔案。

(2)兩頁文件和增加兩個工作表

ODF的試算表和簡報檔和OOX文書處理封包架構不變。ODF的文書處理多了layout-cache快取檔案。OOX試算表多了xl\worksheets路徑下的sheet4.xml和sheet5.xml兩個XML檔案。OOX簡報檔多了ppt\slides路徑下的slide1.xml、slide2.xml；ppt\slides_rels路徑下的slide1.xml.rels、slide2.xml.rels等四個XML檔案。

(3)五頁文件和增加五個工作表

ODF的試算表和簡報檔和OOX文書處理封包架構不變。ODF的文書處理多了layout-cache快取檔案。OOX試算表多了xl\worksheets路徑下的sheet4.xml、sheet5.xml、sheet6.xml、sheet7.xml和sheet8.xml等五個XML檔案。OOX簡報檔多了ppt\slides路徑下的slide1.xml、slide2.xml、slide3.xml、slide4.xml和slide5.xml；ppt\slides_rels路徑下的slide1.xml.rels、slide2.xml.rels、slide3.xml.rels、slide4.xml.rels和slide5.xml.rels等10個XML檔案。

從逐頁分析可以看出封包架構有其規則可循，文書處理檔案大致不變，ODF文書處理只增加了快取檔案。OOX試算表會增加sheet的數量，而OOX簡

報檔則會增加slide和slide.xml.rel的數量(如表2所示)。

表2 逐頁分析封包架構表

辦公室文件類型	一頁差異	二頁差異	五頁差異
ODF 文書處理	無	Layout-cache	Layout-cache
ODF 試算表	無	無	無
ODF 簡報檔	無	無	無
OOX 文書處理	無	無	無
OOX 試算表	sheet4	sheet4、sheet5	sheet4~sheet8
OOX 簡報檔	slide1.xml、 slide1.xml.rels	slide1.xml、slide2.xml、 slide1.xml.rels、slide2.xml.rels	slide1~5.xml、 slide1~5.xml.rels

資料來源：研究者整理

3. 內容分析

除了進行空白封包和逐頁分析之外，本研究分別用OpenOffice.org2.0和Office2007BETA各建立三個範例檔(文書處理、試算表、簡報檔)，並將先前統計常用的各種功能置入，如：文字、圖片、表格、註腳、文字方塊等等，同時進行封包架構上的觀察，結果如下：

(1)文書處理：

ODF文書處理多了Pictures路徑存放圖片，圖片描述有tif、gif、png、jpg、wmf等五種類型、ObjectReplacements路徑下的Object1、Object1兩個檔案。OOX文書處理多了word\media路徑存放圖片，圖片描述有tif、gif、png、jpg、wmf、emf等六種類型。多了footer1.xml、footer2.xml、header1.xml、word_rels\路徑下的footnote.xml.rels、numbering.xml等XML檔案。word\embeddings\路徑下的oleObject1.bin以及CustomXML資料夾存放自訂的XML。

(2)試算表：

ODF試算表多了Object1~Object11路徑存放content.xml和style.xml，共計22個XML檔案。多了ObjectReplacements路徑存放Object1~11，共計11個檔案。OOX試算表多了charts路徑存放chart1~chart11共計11個XML檔案，多了drawings路徑下的drawing1.xml和drawings_rels路徑下的drawing.xml.rels兩個XML檔案。

(3)簡報檔：

ODF簡報檔多了Pictures路徑存放圖片，圖片描述有png和jpg兩種。OOX簡報檔多了DocProps路徑的custom.xml，ppt\media路徑存放圖片，圖片描述有png和jpg兩種。多了ppt\noteMaster路徑下的noteMaster1.xml和ppt\noteMaster_rels路徑下的noteMaster1.xml.rels。theme路徑也變成theme1.xml和theme2.xml。ppt\slideLayout路徑下多了slideLayout12.xml、ppt\slideLayout_rels路徑下

的slideLayout12.xml.rels兩個XML檔案。

從內容呈現分析可以發現兩套軟體之間相對應的文字檔、試算表或簡報檔在資料呈現時仍然有許多差異存在(參考表3),有些差異甚至造成資料的錯亂。兩套軟體之間呈現存在的差異也讓文件的呈現外貌產生落差,對於數位保存來說,除了文件內容資料之外,文件的呈現也是很重要的一部分,若不能夠有效保證文件呈現的真實性,就會降低選用該保存格式的可行性。

從XML封包架構可以看出,兩者的封包架構都具有一定的規則,ODF的封包比較簡單,OOX的則結構較佳並且對常用的功能與特殊的設定都有較詳盡的描述,如:標題、註腳等等。但從ODF或OOX的封包中都可以看出採用XML格式之下的文件內容,利用這些XML檔案就可以清楚地描述該文件內容,也更容易用來創造、儲存、交換資料。

表3 內容封包表

辦公室文件類型	增加的資料夾或路徑	增加的檔案
ODF 文書處理	Pictures、 ObjectReplacements	存放圖片tif、gif、png、jpg、wmf等五種類型、Object1、Object1
ODF 試算表	Object1~Object11、 ObjectReplacements	存放content.xml和style.xml、Object1~11
ODF 簡報檔	Pictures	存放圖片png、jpg兩種類型
OOX 文書處理	word\media、word_rels、 word\embeddings\ CustomXML資料夾	存放圖片tif、gif、png、jpg、wmf、emf等六種類型、footnote.xml.rels、oleObject1.bin、footer1.xml、footer2.xml、header1.xml、numbering.xml
OOX 試算表	charts、drawings 和drawings_rels	chart1~chart11、drawing1.xml和drawing.xml.rels
OOX 簡報檔	DocProps、ppt\media、 ppt\noteMaster、ppt\ noteMaster_rels、theme、 ppt\slideLayout、ppt\ slideLayout_rels	Custom.xml、存放圖片png、jpg兩種類型、noteMaster1.xml、noteMaster1.xml.rels、theme1和theme2、slideLayout12.xml、slideLayout12.xml.rels、

資料來源：研究者整理

五、結論與建議

辦公室文件格式的標準化對於使用者、企業,甚至政府而言,都具有正面意義。使用者可以自由選擇喜愛的應用程式處理文件,而不用擔心無法和其他人交換資料;企業可以藉由開放式標準來達到文件之間的互通、交換和交流;典藏機關在進行數位保存時也可以避免因封閉式的格式而使得文件流失或無法開啟。在有關公開標準的問題上,ODF已是ISO標準(ISO/IEC26300),OOX亦為ECMA組織通過之標準並在ISO組織審議中;且Baker & McKenzie公司針對OOX就其標準過程、ECMA組織及法律層面上之探討,認定OOX為公開標

準(註36)。又IDC公司針對公開過程(open process)、公開文件(open documentation)及公開智慧財產權(open IPR)等三個面向分析ODF及OOX,亦認可同樣為開放文件標準(註37)。換句話說,ODF及OOX都是開放標準;也就是說,這二種格式都符合開放文件標準,可以作為政府、機構、企業等組織之正式文件並適合作為長期保存之用。然而,選擇ODF或選擇OOX作為辦公室文件格式還是須要市場的驗證(註38);像是美國麻州政府和比利時政府雖然大力推動ODF,卻也在OOX也有可能成為ISO標準的情況下改採取觀望的態度。對於使用者來說,開放式文件格式之軟體若不能有強大之功能、友善之介面,及後續的強力技術支援,就會大大降低使用的意願,畢竟多數使用者都會根據市場主流來選擇使用的文件格式。

辦公室文件格式之間的相容性和互操作性也是十分重要的。本研究藉由兩套分別支援ODF和OOX標準的應用軟體針對文字檔、試算表及簡報檔進行實驗後發現,其實兩者的差異性仍然存在,兩套開放標準目前也沒有相容的方式。除了要選擇辦公室文件標準以及對應的應用軟體之外,辦公室文件格式的開放也導致文件本身容易被拆解和解讀,所以文件的安全性和真實性也須要發展一套相對應的機制來配合。有鑒於辦公室文件保存格式的目的是要使得文件可以長久保存並加以應用,所以在選擇哪一套標準作為辦公室文件保存格式上,除了要審慎評估兩套標準的優缺點之外,也應該注意哪一套標準在應用系統上的互操作性是較理想的。

註 釋

註1 The British Library, "Publishing Output to 2020," <http://www.epsltd.com> (accessed December 14, 2006).

註2 The Strategic Plan of the National Archives and Records Administration (NARA), "Preserving the Past to Protect the Future: The Strategic Plan of the National Archives and Records Administration," http://www.archives.gov/about_us/strategic_planning_and_reporting/2000_strategic_plan.html (accessed October 18, 2006).

註3 Green R., "Report on Research User Requirements On-Line Survey," http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research_survey_data_10.pdf (accessed March 1, 2007).

註4 IDABC, "Comparative Assessment of Open Documents Formats Market Overview," <http://ec.europa.eu/idabc/servlets/Doc?id=17982> (accessed March 1, 2007).

註5 二進制檔,由8位元資料或可執行程式碼依序組成的檔案,它和由人類易於讀取的ASCII碼所組成之文字檔有所區別。二進制檔通常是一種僅可由程式讀取的格式,經常以某方法被壓縮或重組以讓特定程式能易於讀取, <http://cls.admin.yzu.edu.tw/cai-edit/edit1-1-2.htm> (檢索於2006年12月15日)。

註6 ODF在2006年5月正式通過ISO認證。

註7 OOX於2006年12月通過ECMA標準組織認證成為ECMA標準並且申請ISO

標準認證。

註8 全國法規資料庫,「電子簽章法第二條」, <http://law.moj.gov.tw/Scripts/Query4A.asp?FullDoc=all&Fcode=J0080037> (檢索於2006年10月19日)。

註9 薛理桂、黃國斌,「淺論電子文件鑑定」, 檔案季刊 1卷, 2期(2002): 22。

註10 The British Library, “Publishing Output to 2020,” <http://www.epsltd.com> (accessed December 14, 2006); The Strategic Plan of the National Archives and Records Administration (NARA), “Preserving the Past to Protect the Future: The Strategic Plan of the National Archives and Records Administration,” http://www.archives.gov/about_us/strategic_planning_and_reporting/2000_strategic_plan.html (accessed October 18, 2006).

註11 Alison Bullock, “Preservation of Digital Information: Issues and Current Status,” <http://www.collectionscanada.ca/9/1/p1-259-e.html> (accessed December 29, 2006).

註12 歐陽崇榮,「數位資訊保存之探討」, 檔案季刊 1卷, 2期(2002): 38。

註13 林巧敏,「電子檔案長期保存方法與策略初探」, 圖書與資訊學刊 55期(2005年11月): 81-84。

註14 Marco Fioretti, “Everybody's Guide to OpenDocument,” <http://www.linuxjournal.com/article/8616> (accessed January 2, 2007).

註15 林信成、龔裕民,「XML 與電子文件展示技術之探討」, 圖書與資訊學刊 37期(2001年5月): 61。

註16 邱炯友,「數位化時代的出版:文字與圖形檔案格式探析」, 研考雙月刊 24卷, 3期(2000年6月): 14。

註17 同上註。

註18 黃國斌,「我國政府機關電子文件鑑定制度之研究」(碩士論文, 政治大學, 2003), 18。

註19 歐陽崇榮, 各國政府機關電子文件管理之研究(台北市: 檔案管理局, 2005), 74。

註20 Green R., “Report on Research User Requirements On-Line Survey,” http://www.hull.ac.uk/esig/repomman/downloads/R-D3-research_survey_data_10.pdf (accessed March 1, 2007).

註21 IDABC, “Comparative Assessment of Open Documents Formats Market Overview,” <http://ec.europa.eu/idabc/servlets/Doc?id=17982> (accessed March 1, 2007).

註22 Marco Fioretti, “Format Wars, File Formats: The Past, the Present and a Possible Future,” *Free Software magazine* 1 (February 2005): 1-4.

註23 Bernadette Hearne, “Gartner and Forrester: Use of Web services skyrocketing,” <http://www.utilitycomputing.com/news/404.asp> (accessed January 29, 2007).

註24 Robert Markham, “XML And PDF: A Will Reign as Future Document Archive Formats,” <http://www.forrester.com/Research/Document/Excerpt/0,7211,36355,00.html> (accessed January 29, 2007).

註25 Organization for the Advancement of Structured Information Standards, <http://www.oasis-open.org/who/> (accessed October 19, 2006).

註26 Eben Moglen, “Software Freedom Law Center,” <http://www.softwarefreedom.org/publications/OpenDocument.html> (accessed October 31, 2006).

註27 OASIS, “OASIS NEWS,” http://www.oasis-open.org/news/oasis_news_11_20_02.

php (accessed October 31, 2006); OASIS, “OASIS NEWS,” http://www.oasis-open.org/news/oasis_news_05_23_05.php (accessed October 31, 2006).

註28 IDABC, “IDABC REPORT 22,” <http://ec.europa.eu/idabc/servlets/Doc?id=17085> (accessed October 31, 2006).

註29 ISO, “ISO and IEC Approve OpenDocument OASIS Standard for Data Interoperability of Office Applications,” <http://www.iso.org/iso/en/commcentre/pressreleases/2006/Ref1004.html> (accessed October 1, 2006).

註30 OASIS, “Open Document Format for Office Applications OpenDocument v1.1,” <http://www.oasis-open.org/committees/download.php/20847/OpenDocument-v1.1-cs1.pdf>

註31 同上註。

註32 Erika Ehrli, “Walkthrough: Word 2007 XML Format,” <http://msdn2.microsoft.com/en-us/library/ms771890.aspx> (accessed January 28, 2007).

註33 Ecma, “Ecma International Approves Office Open XML Standard,” http://www.ecma-international.org/news/PressReleases/PR_TC45_Dec2006.htm (accessed December 21, 2006).

註34 Ecma, “Ecma International Technical Committee TC45,” http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf (accessed December 18, 2006).

註35 同上註。

註36 Baker and McKenzie, “Standardisation and Licensing of Microsoft’s Office Open XML Reference Schema,” <http://www.bakernet.com/NR/rdonlyres/CC54A6B6-79E8-4E0D-B290-C836D5F70867/0/OpenXML.pdf>

註37 Per Andersen, “Technology Assessment: The Road To Open Document Standards,” <http://openxmldeveloper.org/attachment/868.ashx>

註38 ODF可以在許多電腦系統平台上執行如IBM、SUN……等；OOX只能在微軟公司的Windows系統之Office軟體上執行較方便而有其侷限。其實OOX亦可在其他公司之應用軟體上執行，但就目前而言(2007年)是比較少的。

The Study on Office e-Documents Format of ODF and OOX

James C. Ouyang

Assistant Professor
Department of Information and Library Science, Tamkang University
E-mail: cjouyang@mail.tku.edu.tw

Jr Feng Shiue

Graduate Student
Department of Information and Library Science, Tamkang University
Taipei, Taiwan, R.O.C.
E-mail: 694070151@s94.tku.edu.tw

Abstract

This research investigates two standards of office e-Documents—Open Document Format (ODF) and the Office OpenXML (OOX). Researchers also apply two office applications that one supports ODF and the other supports OOX to open and compare e-Documents of word-processing, spreadsheet and presentation. The result shows that there are some differences between ODF and OOX.

Keywords: *Electronic documents; Office documents; Open document format; OfficeOpenXML*