

I formati per la conservazione digitale: criteri operativi per la selezione dei formati

In una strategia di conservazione (lungo periodo) digitale la scelta del formato è uno degli aspetti più importanti da considerare; in particolare sono da preferire formati che siano non proprietari, aperti, trasparenti e standard (possibilmente de jure)

STEFANO ALLEGREZZA

IL PROBLEMA DELL'OBSOLESCENZA DEI FORMATI

La scelta del formato rappresenta uno degli aspetti più importanti da prendere in considerazione in qualunque strategia di conservazione digitale. L'esperienza insegna infatti che molti dei formati che erano particolarmente in auge nel passato sono ormai pressoché scomparsi, e chi possiede ancora contenuti digitali codificati secondo quei formati incontra sicuramente serie difficoltà ad accedervi.

È tristemente noto il caso delle sonde Viking che furono lanciate su Marte nel 1975 per verificare se sul pianeta

rosso potessero esistere forme di vita¹. All'epoca, gli scienziati erano assolutamente certi che i dati scientifici raccolti sarebbero stati disponibili per le future generazioni di scienziati sui nastri magnetici sui quali vennero memorizzati. Tuttavia, solamente pochi decenni dopo, quando gli scienziati tentarono di riutilizzare alcuni di quei dati alla fine degli anni '90, si accorse che i nastri erano ormai talmente deteriorati da rendere quasi illeggibili i dati memorizzati; oltre a ciò, si resero conto che il problema più grande era costituito dall'impossibilità di decodificare il formato che era stato utilizzato solo una ventina di anni

prima; fortunatamente riuscirono a recuperare delle vecchie stampe su carta contenenti quei preziosi dati, ma furono costretti a ridigitare tutto.

Venendo ai nostri giorni, la situazione non si presenta molto diversa: enti pubblici ed aziende private possiedono archivi di documenti memorizzati su supporti ormai obsoleti e codificati secondo formati anch'essi ormai obsoleti. Si pensi che non più di venti anni fa il programma di videoscrittura più utilizzato era WordStar; non più di quindici anni fa molte amministrazioni pubbliche utilizzavano il programma IBM Writing Assistant; altre creavano i loro documenti con Lotus Word

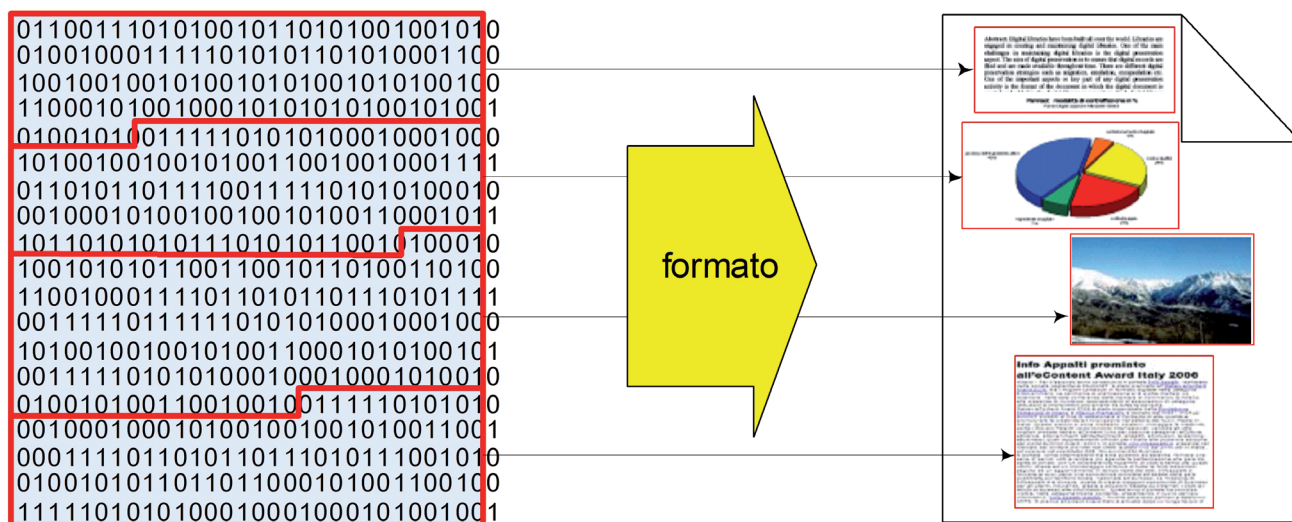


Figura 1 - Senza il relativo formato qualsiasi file altro non è che una sequenza di bit priva di significato



Figura 2 - Classificazione dei requisiti dei formati

Pro; si tratta di tre software di videoscrittura piuttosto comuni all'epoca, ma oggi praticamente in disuso. Di conseguenza, a distanza di solamente un paio di decenni, i documenti creati con questi programmi risultano per lo più inaccessibili non tanto per il fatto che i supporti su cui sono stati memorizzati sono ormai deteriorati o illeggibili², quanto per il fatto che i moderni programmi di videoscrittura non riconoscono più questi formati obsoleti³.

Tutto ciò fa comprendere l'importanza di stabilire dei criteri operativi per la scelta dei formati che siano compatibili con un processo di conservazione a lungo termine ed assicurino la loro leggibilità a distanza di venti, cinquanta o più anni.

CHE COS'È UN FORMATO ELETTRONICO

È importante innanzitutto avere ben chiaro che cosa sia un formato elettronico. Si definisce formato elettronico, o anche formato di file (file format), l'insieme di codici e regole che, a partire dalla sequenza di bit che costituisce un oggetto digitale, permettono di riprodurre mediante un software (a video, a stampa o su altri dispositivi di output) il relativo oggetto informativo con lo stesso contenuto e nella stessa forma che gli ha dato l'autore del documento⁴. In altre parole, qualsiasi

contenuto digitale altro non è che una sequenza di bit "0" ed "1", che non avrebbe alcun significato se non fosse noto il formato secondo il quale tale sequenza di bit è stata codificata e con la quale deve essere decodificata. È grazie al formato che le sequenze di bit vengono decodificate in porzioni di testo, immagini, grafici, etc. (si veda la Figura 1). Le informazioni relative alle modalità secondo le quali tale sequenza di bit debba essere interpretata sono fornite da uno o più documenti tecnici, che nel loro complesso vengono denominati specifiche del formato.

I REQUISITI DEI FORMATI

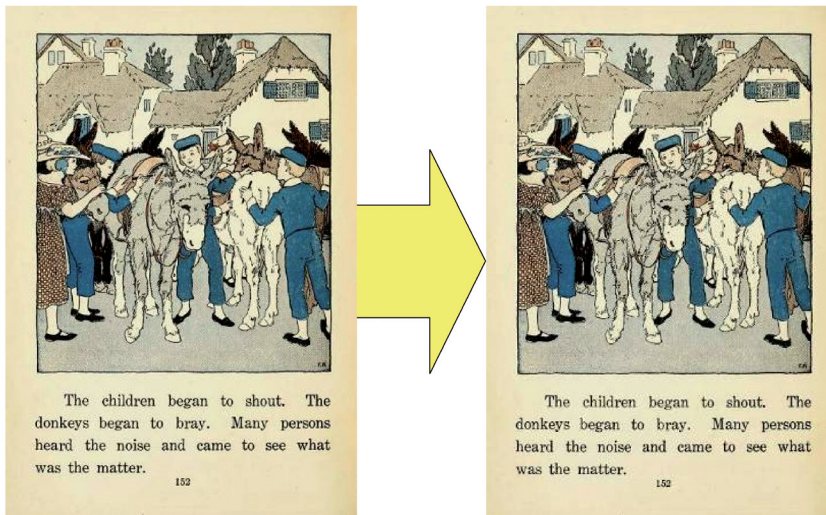
Uno studio molto approfondito sulle problematiche inerenti i formati elettronici è stato condotto presso il Laboratorio di Informatica Documentale dell'Università degli Studi di Macerata. Nell'ambito di questo studio sono stati identificati i requisiti che un formato deve possedere per essere considerato compatibile con un processo di conservazione digitale⁵. Tali requisiti (si veda la Figura 2) sono stati classificati in generali, ovvero applicabili a qualsiasi tipologia di formato, e specifici, ovvero relativi ad una particolare categoria di formati (ad esempio, i formati immagine, i formati video, i formati audio, etc.).

I requisiti generali a loro volta sono

stati suddivisi in requisiti di primo livello (che consentono una prima selezione dei formati sulla base delle caratteristiche di non proprietà, apertura, standardizzazione e trasparenza) e requisiti di secondo livello (che vanno ad analizzare altre caratteristiche non meno importanti dei formati, quali la robustezza, la stabilità, l'auto-contenimento, etc.). Non è certamente questa la sede adatta per una trattazione approfondita dei vari requisiti; si cercherà tuttavia di fornire alcuni criteri operativi che possano indirizzare verso la scelta dei formati corretti.

I REQUISITI GENERALI DI PRIMO LIVELLO

Uno dei principali requisiti generali è quello della non proprietà: per garantire le migliori condizioni ai fini della conservazione nel tempo, è preferibile utilizzare formati che siano non proprietari, ovvero non legati all'esistenza di una specifica azienda che ne detiene i diritti di proprietà intellettuale e che potrebbe in qualsiasi momento decidere di modificarne le specifiche, renderle inaccessibili, o imporre restrizioni sul loro utilizzo. Ad esempio, sono non proprietari i formati dei documenti prodotti con le suite d'automazione d'ufficio Microsoft Office 2007 e OpenOffice.org, mentre sono proprietari (di proprietà Microsoft) i ben noti formati



TIFF non compresso prima e dopo essere stata sottoposta ad un processo di corruzione di alcuni dei bit che ne compongono il relativo file.

La Figura 4 rappresenta la stessa immagine non più in formato TIFF, ma in formato JPG, anch'essa prima e dopo essere stata sottoposta ad un processo di corruzione della stessa entità del caso precedente.

Come si può vedere, mentre l'immagine in formato TIFF non compresso "sopporta" bene il processo di corruzione (tanto che le differenze tra l'immagine originale e quella modificata non risultano immediatamente visibili), l'immagine in formato JPG ne viene irrimediabilmente compromessa¹². In altre parole, il formato TIFF non compresso è più robusto del formato JPG e rappresenta infatti uno dei formati più idonei per la conservazione delle immagini.

Altro requisito di fondamentale importanza è quello dell'auto-contenimento: un formato è auto-contenuto quando permette di includere all'interno dei file codificati secondo quel formato tutte le risorse necessarie (immagini, font, spazi di colore, etc.) per la loro rappresentazione. Ad esempio, non sono auto-contenuti tutti quei documenti che non incorporano le immagini all'interno del documento stesso, ma contengono al loro posto dei collegamenti (link) che vengono sostituiti con le immagini a cui si riferiscono nel momento in cui il documento viene visualizzato. Nel caso in cui il collegamento dovesse interrompersi¹³, le immagini non potranno essere visualizzate ed il documento non conserverà più l'aspetto originale. Analogamente tutti i font necessari per la corretta visualizzazione del documento devono essere incorporati all'interno del documento stesso per evitare il noto problema di "missing fonts"¹⁴.

Un altro requisito basilare è quello della stabilità, che è soddisfatto allorché il formato non è soggetto a continue e sostanziali modifiche nel corso del tempo. Ai fini della conservazione digitale, è preferibile che un formato sia stabile nel tempo e non sottoposto a continue trasformazioni e revisioni.

Il concetto di stabilità è strettamente collegato con quelli di compatibilità in

Figura 4 - Immagine in formato TIFF (non compresso) prima e dopo il processo di corruzione



Figura 5 - Immagine in formato JPG prima e dopo il processo di corruzione

aperto (le sue specifiche sono liberamente accessibili), non proprietario (anch'esso era di proprietà di Adobe Systems, ma dopo il riconoscimento come standard ISO 19005-1:2005 è diventato non proprietario), non trasparente (è binario) e standard de jure.

I REQUISITI GENERALI DI SECONDO LIVELLO

Oltre ai requisiti di primo livello appena visti, ne esistono diversi altri denominati di secondo livello, più per una questione di semplice classificazione che non perché siano meno importanti rispetto ai primi.

Uno di questi è la robustezza: un formato è robusto quando in caso di corruzione¹⁰ del file codificato secondo quel formato consente il recupero, totale o parziale dei suoi contenuti. A seconda della possibilità di recuperare totalmente o parzialmente il contenuto dell'oggetto digitale si distinguono vari gradi di robustezza. Ad esempio, i formati binari e quelli compressi sono i meno robusti, dal momento che è sufficiente a volte la sola corruzione di un solo bit per rendere illeggibile tutto il file¹¹. Invece i formati non binari e non compressi sono in generale i più robusti. Si veda il seguente esempio: la Figura 4 rappresenta un'immagine in formato

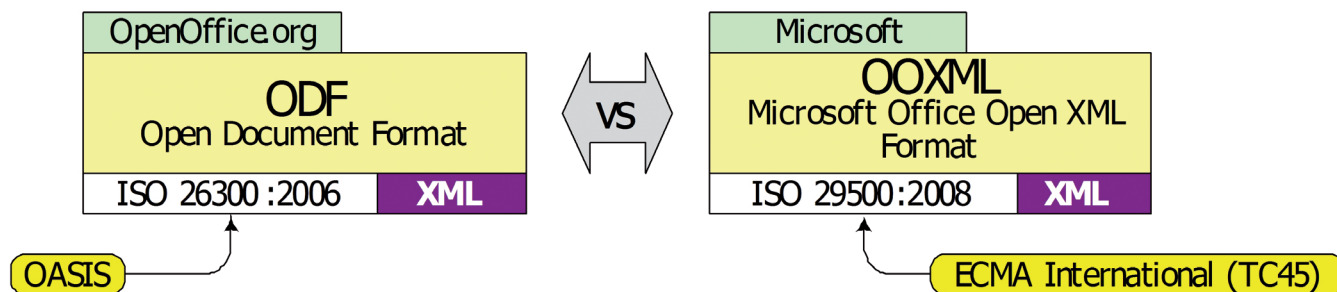


Figura 6 - La guerra dei formati per l'office automation

avanti e compatibilità all'indietro. Normalmente tutti i formati assicurano la compatibilità in avanti (anche se spesso solo per un limitato numero di versioni precedenti), mentre non sempre viene assicurata la compatibilità all'indietro¹⁵, obbligando gli utenti all'acquisto di nuove versioni del software e creando una non auspicabile dipendenza da una determinata software house¹⁶. Ai fini della conservazione nel tempo dei contenuti digitali, è di fondamentale importanza la compatibilità in avanti, ma anche la compatibilità all'indietro ha un suo peso: purtroppo, è abbastanza raro incontrare formati che la garantiscano, ed è ancora più raro incontrare formati che garantiscano¹⁷ in maniera corretta sia la compatibilità in avanti che all'indietro. Fin qui sono stati esaminati alcuni dei requisiti generali di secondo livello, ma ne esistono numerosi altri, non meno importanti, ai quali si farà un brevissimo accenno:

- **L'indipendenza dal dispositivo (o portabilità)** - fa riferimento alla capacità di un file, codificato in un determinato formato, di essere rappresentabile in maniera attendibile e allo stesso modo indipendentemente dalla piattaforma hardware e dal software;
- **Passenza di meccanismi tecnici di protezione** - per conservare i documenti informatici e garantirne la fruibilità nel lungo periodo, si deve avere la possibilità di replicarne il contenuto su nuovi supporti, effettuare migrazioni e normalizzazioni, renderli disponibili per la diffusione; ciò non è possibile in presenza di meccanismi di protezione (ad esempio, la crittografia o la protezione mediante password);

- **Passenza di limitazioni sull'utilizzo** - l'esistenza di brevetti (patent) su un formato digitale, o la richiesta del pagamento di royalty per il suo utilizzo, incide negativamente sulla possibilità di conservare i file codificati secondo quel formato;
- **L'auto-documentazione** - i formati che permettono l'inclusione dei metadati nei file codificati secondo quel formato sono da preferire;
- **L'accessibilità** - un formato deve consentire una facile fruizione anche da parte di persone diversamente abili;
- **la sicurezza** - un formato adatto alla conservazione digitale deve essere sicuro, ovvero non passibile di attacchi da parte di virus ed altre forme di codice maligno, che potrebbero modificare il contenuto di un file all'insaputa dell'utente pur lasciandolo leggibile, corrompere alcune sue parti rendendolo inutilizzabile o, nei casi estremi, eliminarlo completamente;
- **L'efficienza** - dal momento che la quantità di informazioni che vengono archiviate è in continuo e costante aumento, è importante considerare i formati anche dal punto di vista della dimensione dei file che vengono prodotti; sotto questo profilo esistono formati che, a parità di contenuto informativo, riescono a contenere maggiormente le dimensioni del file risultante e quindi sono più efficienti.

Accanto ai requisiti di tipo generale fin qui analizzati devono essere presi in considerazione i requisiti specifici del particolare formato o della particolare categoria a cui quel formato appartiene (ad esempio: formati di testo, formati

immagine, formati audio, etc.); essi prendono in esame le proprietà che si ritengono importanti per quei formati ai fini della loro migliore conservazione nel tempo. A titolo di esempio, riferendoci ai formati immagine è possibile prendere in considerazione requisiti relativi alla risoluzione, alla profondità di colore, alla funzionalità di ingrandimento e riduzione, etc. Per i formati audio i requisiti possono riguardare il grado di fedeltà, il numero di canali, il bitrate, e così via.

LA "GUERRA DEI FORMATI"

Per garantire la conservazione nel tempo dei contenuti digitali, occorre partire con il "piede giusto" iniziando dalla scelta del formato corretto. In particolare, è importante preferire formati che siano non proprietari, aperti, trasparenti e standard (possibilmente de jure). Questi requisiti sono talmente importanti che tutte le più grandi aziende produttrici di software (che sono anche i principali produttori di formati) si stanno indirizzando verso la creazione di formati che soddisfino tali requisiti e ciò ha scatenato una vera e propria "guerra dei formati".

Ad esempio, se si fa riferimento ai formati per la produzione di documenti di office automation, la situazione vede ormai uno scontro aperto tra la comunità open source OpenOffice.org, che ha sviluppato il formato ODF utilizzato dall'omonima suite di automazione d'ufficio OpenOffice.org, e Microsoft, che ha creato il formato OOXML utilizzato dall'ultima versione della sua ben nota suite Microsoft Office 2007 (si veda la Figura 6)¹⁸.

Il formato Open Document Format (ODF)¹⁹ soddisfa in pieno tutti i re-

quisiti precedentemente analizzati: è non proprietario (la sua gestione appartiene alla comunità open source di OpenOffice.org e non è controllato da alcuna azienda privata), aperto (le sue specifiche sono liberamente accessibili sul sito dell'OASIS e, a pagamento, su quello dell'ISO) e completamente documentato (la sua specifica è costituita da un documento di circa 700 pagine in cui il formato viene descritto in maniera esauriente e dettagliata), trasparente (è basato sull'XML ed i tag utilizzati sono stati disegnati specificamente per consentire la massima leggibilità, anche in maniera diretta da parte dell'uomo utilizzando un semplice editor di testo; risulta quindi non binario, eccezion fatta per i contenuti multimediali in essi eventualmente presenti); è standard de jure (il formato è stato riconosciuto come standard prima dall'OASIS e poi dall'ISO con la denominazione di ISO/IEC 26300:2006) ed ampiamente adottato (è il formato di default per molte applicazioni²⁰ ed inoltre molti Stati lo hanno scelto come formato obbligatorio per l'accettazione dei file nei depositi digitali e per la pubblicazione sul web).

Nei confronti dei requisiti generali di secondo livello, l'ODF si presenta robusto (è stato concepito per essere più robusto dei formati binari)²¹, auto-contenuto (tutti gli elementi che compongono il file possono essere inclusi all'interno del file in maniera da evitare qualsiasi tipo di riferimento verso l'esterno), auto-documentato (i metadati sono inclusi al suo interno), indipendente dal dispositivo (essendo stato progettato proprio per essere indipendente dalla piattaforma tecnologica), privo di meccanismi tecnici di protezione (per impostazione predefinita i file ODF non vengono cifrati), privo di limitazioni sull'utilizzo (non è soggetto a restrizioni di alcun tipo, dovute a brevetti o a licenze, non richiede il pagamento di diritti per il suo utilizzo del formato e può essere adottato liberamente da qualsiasi produttore di software), accessibile (essendo esaustivamente e pubblicamente documentato, chiunque può creare il software necessario per la fruizione da parte di utenti diversamente abili), stabile (è uno standard internazionale ISO e nessun

produttore di software ha la possibilità di apportare autonomamente modifiche alla specifica del formato; inoltre, ogni successiva versione del formato deve essere il frutto di un accordo unanime e sottoposto a revisione da parte della comunità open source), sicuro (se si fa attenzione all'utilizzo delle macro che, in quanto codice eseguibile, rendono il formato vulnerabile all'attacco di virus), efficiente (la compressione, opzionale impiega un valido algoritmo di compressione standard e di tipo lossless).

Per tutta risposta Microsoft, già sviluppatrice dei vecchi formati proprietari DOC, XLS e PPT utilizzati dalla sua suite di office automation Microsoft Office fino alla versione 2003, con l'ultima versione della sua suite Office 2007 ha introdotto un nuovo formato, l'Office Open XML Format (OOXML)²². Si tratta di un formato non proprietario (è stato sviluppato da Microsoft ed era quindi proprietario fino a quando, all'inizio del 2008, ha ottenuto il riconoscimento come standard ISO), aperto (le sue specifiche sono liberamente accessibili sul sito Microsoft, dell'Ecma International e, a pagamento, su quello dell'ISO), completamente documentato (il formato è esaustivamente descritto in una specifica di oltre 6000 pagine), trasparente (non binari) e standard de jure (è stato riconosciuto come standard prima dall'Ecma International con la norma ECMA-376 e poi, con due anni di ritardo rispetto all'ODF, dall'ISO con la norma ISO/IEC 29500:2008), ampiamente adottato (essendo il formato predefinito della suite Microsoft Office 2007, si prevede una sua ampia diffusione), trasparente (è basato sull'utilizzo di un file "contenitore" in formato ZIP al cui interno trovano posto i contenuti testuali in formato XML ed eventuali contenuti binari, come le immagini).

È inoltre robusto (a differenza dei formati delle precedenti release di Microsoft Office è di tipo non binario, eccezion fatta, ovviamente, per la presenza di eventuali contenuti grafici, audio e video), auto-contenuto (tutti gli oggetti necessari alla rappresentazione di un file in formato OOXML, possono essere inclusi all'interno del file "contenitore"), auto-documentato (i metadati

sono inclusi anch'essi nel file "contenitore"), indipendente dal dispositivo (essendo basato sull'XML, è possibile sviluppare applicazioni che siano in grado di gestire il formato indipendentemente dalla piattaforma tecnologica utilizzata), privo di meccanismi tecnici di protezione (per impostazione predefinita, i file OOXML non vengono cifrati), privo di limitazioni sull'utilizzo (il formato è esente da diritti sull'utilizzo), accessibile (essendo basato sull'XML è possibile produrre file che, facendo un uso corretto dei tag, risultano accessibili), stabile (il fatto che sia stato riconosciuto come standard ISO fornisce ampie garanzie sul fatto che non subirà trasformazioni dettate da interessi privati), sicuro (come impostazione predefinita, i nuovi formati DOCX, XLSX e PPTX non possono contenere codice eseguibile – e quindi risultano immuni agli attacchi da parte dei virus da macro), efficiente (un file in formato OOXML è un contenitore di file XML e di altri oggetti digitali come immagini, audio, etc., compresso con la tecnologia ZIP; poiché l'XML è un formato basato sul normale testo e può essere compresso con molta efficacia, è possibile ottenere significative riduzioni nella dimensione del file rispetto ad un analogo file binario con lo stesso contenuto).

CONCLUSIONI

Allo stato attuale non è assolutamente semplice fare previsioni su quali saranno i formati che riusciranno ad imporsi e che ritroveremo tra dieci, venti o cinquanta anni. Appare tuttavia piuttosto chiara la direzione verso la quale si stanno dirigendo le grandi aziende e le comunità produttrici di software e di formati: gli esempi prima citati dei formati ODF e OOXML, così come quelli di numerosi altri formati di grande adozione, dimostrano che si sta andando verso la creazione di formati che soddisfino i requisiti generali di primo livello (non proprietà, apertura, trasparenza e standardizzazione) e il maggior numero possibile di requisiti generali di secondo livello, condizioni queste che forniscono le maggiori garanzie per il futuro. È importante quindi orientarsi verso la scelta di quei formati che soddisfano al massimo

livello questi requisiti, in maniera da fondare su basi solide il processo di conservazione digitale.

NOTE

¹ Si veda MARTIN WALLER, ROBERT SHARPE, Mind the gap. Assessing digital preservation needs in the UK, Digital Preservation Coalition, The Digital Preservation Coalition Innovation Centre, York Science Park, Heslington, YORK YO10 5D, 2006, disponibile all'indirizzo <<http://www.dpconline.org/docs/reports/uknamindthegap.pdf>>.

² Dal momento che i relativi drive non sono più disponibili: si pensi ai lettori di floppy da 5" 1/4 che sono stati dismessi ormai da più di 15 anni, o ai lettori di floppy da 3" 1/2 che stanno scomparendo, rendendo impossibile la lettura dei relativi supporti.

³ La soluzione di ricorrere a software house per la produzione di appositi software di conversione risulta normalmente assai costosa.

⁴ Cfr. STEFANO PIGLIAPOCO, La memoria digitale delle amministrazioni pubbliche. Requisiti, metodi e sistemi per la produzione, archiviazione e conservazione dei documenti informatici, Maggioli editore, Sant'Arcangelo di Romagna (RN), 2005.

⁵ I risultati di questo studio sono stati pubblicati nel volume STEFANO PIGLIAPOCO, STEFANO ALLEGREZZA, Produzione e conservazione del documento digitale. Requisiti e standard per i formati elettronici. Volume I, Edizioni EUM, Macerata, 2008.

⁶ Le versioni della suite Microsoft Office che si sono succedute nel tempo sono le seguenti: Office 3 (la versione su floppy disk e CD-ROM comprendente Word 2.0c, Excel 4.0a, PowerPoint 3.0 e Mail), rilasciata il 30 agosto 1992 e successivamente ri-pacchettizzata come Office 92; Office 4 (Word 6.0, Excel 4.0, PowerPoint 3.0), rilasciata il 17 gennaio 1994; Office 4.3 (l'ultima versione a 16 bit, che comprendeva Word 6.0, Excel 5.0, PowerPoint 4.0 e, nella versione Professional, anche Access 2.0), rilasciata il 2 giugno 1994; Office 95 (versione 7, comprendente Word 95, Excel 95, etc.), rilasciata il 30 agosto 1995; Office 97 (versione 8, comprendente Word 97, etc.), rilasciata il 30 dicembre 1996 e disponibile sia su CD-ROM che su un set di 45 floppy disk da 3,5 pollici; Office 2000 (versione 9, comprendente Word 2000, etc.) rilasciata il 27 gennaio 1999; Office 2002 o Office XP (versione 10, comprendente Word 2002, etc.) rilasciata il 31 maggio 2001; Office 2003 o Office System (versione 11, comprendente Word 2003, etc.) rilasciata il 17 novembre 2003; Office 2007 (versione 12), rilasciata contemporaneamente alla commercializzazione del nuovo sistema operativo Windows Vista.

⁷ Ad esempio l'ISO, l'ANSI, l'ECMA internazionale, il W3C, etc.

⁸ Esempi di standard de facto sono il DOC, l'XLS, l'RTF; esempi di standard de jure sono l'HTML (W3C Recommendation e ISO 15445), l'XML (W3C Recommendation), l'ODF (ISO 26300), il TIFF (ISO 12639), il JPG (ISO 10918-1), il JPEG2000 (ISO 15444), il PNG (ISO

15948), il PDF (ISO 32000) e alcuni suoi subset quali il PDF/A, il PDF/E e il PDF/X.

⁹ Sono non binari quei formati in cui la sequenza di bit che costituisce il file (o la maggior parte di essa) è interpretabile direttamente come una sequenza di caratteri ordinari (lettere, cifre, segni di interpunzione, etc.) o di controllo (ritorni a capo, tabulazioni, interruzioni di linea, etc.), associando ad ogni byte il carattere corrispondente (solitamente nella codifica ASCII o UNICODE).

¹⁰ Per corruzione si intende l'alterazione o la cancellazione, solitamente accidentale e non voluta, di uno o più bit tra quelli che compongono il file.

¹¹ Nel caso in esame si è provveduto a porre a zero lo 0,006% dei byte che compongono i file delle due immagini, il che corrisponde a circa 6 byte corrotti ogni 100.000 byte.

¹² Ciò è dovuto essenzialmente al fatto che il JPG è un formato compresso.

¹³ Ad esempio, perché il file dell'immagine è stato eliminato o ridenominato, oppure è stata spostata la cartella che lo conteneva, etc.

¹⁴ Tale problema si presenta quando i font necessari per rappresentare un determinato contenuto non sono stati incorporati all'interno del relativo file né sono installati nel sistema operativo dell'elaboratore dove viene rappresentato. In questo caso tali font vengono sostituiti da altri "simili", con la conseguenza che si ottiene una rappresentazione non conforme all'originale. Vale la pena ricordare che su qualsiasi computer che utilizzi i sistemi operativi Windows Me/2000/XP sono presenti di default i seguenti tipi di carattere: Courier New, Arial e Times New Roman (tutti TrueType, devono essere presenti anche le varianti Bold, Italic e Bold Italic), Symbol e Wingdings (TrueType), MS Serif e MS Sans Serif (standard).

¹⁵ Un formato si dice compatibile in avanti, quando è possibile utilizzarlo con una delle versioni del software successive a quella con cui è stato prodotto; si dice, invece, compatibile all'indietro se è possibile utilizzarlo con una delle versioni precedenti dell'applicativo che lo ha prodotto. Un formato è infine compatibile all'indietro e in avanti quando è possibile utilizzarlo sia con le versioni precedenti dell'applicativo che lo ha prodotto, sia con quelle successive.

¹⁶ Ad esempio, se si dispone di Microsoft Word 2003 e si ricevono documenti di testo nel nuovo formato DOCX (di Microsoft Word 2007), si è "costretti" ad acquistare la nuova versione del software di videoscrittura (o quantomeno ad utilizzare il Microsoft Word Viewer 2007, un visualizzatore di documenti di testo nel formato DOCX scaricabile gratuitamente dal sito Microsoft).

¹⁷ Alcuni formati, come il PDF; assicurano una compatibilità all'indietro limitata: consentono alle versioni precedenti del software di utilizzare i file generati nei formati gestiti dalle versioni successive, ma non permettono di gestire le nuove caratteristiche introdotte dai formati più recenti, avvisando di questo fatto l'utente mediante opportuni alert box.

¹⁸ Nel momento in cui scriviamo è imminente il rilascio della successiva versione, Microsoft Office 2010.

¹⁹ Si tratta più precisamente di una famiglia dei formati alla quale appartengono l'ODT per i documenti di testo, l'ODS per i fogli di calcolo, l'ODP per le presentazioni e l'ODG per i grafici, solo per citare i più conosciuti.

²⁰ L'ODF è il formato predefinito per le suite di office automation OpenOffice.org 2.0, KOffice 1.5, StarOffice 8, IBM Lotus Symphony e per altre applicazioni. Altre suite di ampia diffusione, compreso Microsoft Office 2007, consentono di salvare nel formato ODF mediante appositi add-in.

²¹ Dal momento che esso memorizza separatamente i singoli componenti del file, il rischio di perdita di dati a causa di file danneggiati o corrotti viene ridotto e migliorano le possibilità di recupero. Infatti, anche se una delle parti che compongono il file fosse corrotta, è possibile "ignorarla" e estrarre le informazioni dalle rimanenti parti non corrotte. Inoltre, poiché il formato è ben documentato ed aperto, chiunque può creare strumenti per recuperare parti del file, per correggere le parti XML che non sono ben formate o per aggiungere eventuali elementi necessari che non sono presenti. Nel formato ODF la compressione, fattore che riduce la trasparenza, è opzionale. Nel caso si scelga di utilizzarla, l'algoritmo di compressione adottato è lo ZIP, di tipo lossless, aperto, completamente documentato, non soggetto a licenze ed ampiamente adottato.

²² Si tratta, più precisamente, di una famiglia di formati composta dai noti DOCX per i documenti di testo, XLSX per i fogli di calcolo e PPTX per le presentazioni, oltre che da diversi altri formati.

²³ Inoltre, Microsoft Office 2007 garantisce la compatibilità all'indietro con alcune delle precedenti versioni della suite: Microsoft Office 2003, Microsoft Office XP e Microsoft Office 2000. Gli utenti di queste versioni possono pertanto continuare ad utilizzare i vecchi formati binari DOC, XLS e PPT, che rimangono completamente compatibili con Microsoft Office 2007. Se invece intendono passare al nuovo formato non devono far altro che scaricare un add-in gratuito che permette loro di aprire e modificare i file nel formato OOXML pur continuando ad utilizzare le vecchie versioni della suite.

²⁴ Se si desidera abilitare le macro, occorre esplicitamente utilizzare i formati macro-enabled, che vengono identificati da una particolare estensione e possono così essere facilmente individuati.



STEFANO ALLEGREZZA

Professore a contratto di Informatica e sistemi di elaborazione delle informazioni presso l'Università degli Studi di Macerata