

Scripta manent, sed “digitale” permanent?

PAOLO TEMPESTINI

Gli strumenti di lavoro dello studioso moderno

Il problema della conservazione dei documenti e della loro condivisione è molto grave e profondo, probabilmente sentito già da molti. Ti è capitato di perdere qualcosa perché conservato in un vecchio supporto, magari un floppy disk ormai deteriorato? Hai un dispositivo in grado di leggerlo? Quanti ricordano i vecchi sistemi di archiviazione “a lungo termine” su nastro, oramai obsoleti? Quanti dati sarebbero andati persi se non si fossero trasferiti su altri supporti come i CD ed i DVD? Tra 10 anni esisteranno ancora i DVD? Ti è capitato di ricevere o di aprire un vecchio file senza riuscire a leggerlo? Magari risultava completamente spaginato o con simboli strani? Quante volte ti è capitato di salvare qualcosa in versione “compatibile con...”? Ecco, i nuovi problemi dello studioso moderno si racchiudono efficacemente in queste domande che rendono evidente un grave problema: perdita di dati e scarsa compatibilità, o meglio, scarsa interoperabilità.

In questo articolo non tratterò dei supporti nei quali conservi i dati, ma dei dati stessi: con quale programma li hai creati? Quante ore di lavoro trascorse davanti ad una delle 15 versioni di Office scarsamente compatibili tra loro che Microsoft ha prodotto negli ultimi 25 anni?¹

Gli Office 2003, 2007 e 2010 sono emblematici nel rappresentare problemi di compatibilità: a distanza di 7 anni mostrano numerose differenze, tanto che se viene prodotto un file in Office 2010 si deve conoscere la versione del programma del “ricevente” in modo da creare un file “compatibile” con la sua versione di Office. Detto ciò, se in 7 anni la tecnologia è cambiata così in fretta, come possiamo conservare la nostra documentazione a lungo termine e condividerla con gli altri? La risposta, per chi ne ha apprezzato gli indiscutibili pregi, sembra ovvia: “PDF!” il famoso Portable Document Format. Attenzione però, è pericoloso non distinguere tra i diversi formati dato che sono molti e non immediatamente distinguibili, alcuni dei quali NON adatti a conservare i documenti a lungo termine e NON adatti alla condivisione tra computer diversi².

Pensando alla storia è utile ricordare che sia gli strumenti – scalpelli, piume, penne a sfera, ecc. – che i supporti – ad esempio tavolette di argilla, papiro, carta – hanno subito un’evoluzione millenaria; nell’attuale tecnologia digitale, gli strumenti, i documenti digitali stessi (i file) e la loro conservazione, nel giro di soli 50 anni hanno fatto un percorso altrettanto lungo. Il calamaio è rimasto sulle scrivanie per centinaia di anni, l’inchiostro è tutt’ora utilizza-

to, la carta ha una storia altrettanto lunga. Dalla nascita dei computer, nati poche decine di anni fa, sono stati inventati, si sono diffusi e sono stati abbandonati supporti come le schede perforate, i nastri su bobina, i floppy disk esistenti in ben 3 dimensioni diverse, cassette a nastro magnetico, CD. Fatto meno visibile e probabilmente meno noto è che gli strumenti di scrittura digitale, i programmi che ci consentono di scrivere, sono nati e si sono evoluti ancor più velocemente dei supporti generando problemi ancor più importanti; dal semplice e primitivo strumento di scrittura senza opzioni per impaginazione e singolo formato testo, per passare al gran numero di strumenti più evoluti fino alle già citate 15 versioni di Microsoft Office. Tali velocissimi cambiamenti hanno creato negli utenti, aiutati dalle strategie delle multinazionali e dal monopolista del settore, Microsoft, una malsana confusione che è necessario eliminare essenzialmente per tre motivi:

- per difendere prima di tutto il proprio lavoro dalle insidie del tempo e dalla difficoltà nella condivisione – tra colleghi, tra computer diversi con sistemi operativi diversi, per rielaborare lo stesso file ad anni di distanza, ecc...
- per non legarsi a filo doppio ad un unico programma di scrittura e ad un unico produttore divenendo attore passivo di decisioni venute “dall’alto”. Coloro i quali hanno sofferto il passaggio da Office 2003 con le tendine (File, Modifica, Visualizza, ecc.) ad Office 2007 con il menù diviso a schede sa a cosa mi riferisco; con ciò non sostengo che è sbagliato cambiare, tuttavia è importante che la scelta ed il tipo di cambiamento venga fatta dallo studioso, non imposta dal mercato o, peggio, dal produttore di un singolo software. Il programma utilizzato è una cosa, il proprio documento scritto è un’altra: Immaginiamolo come una penna, una penna molto sofisticata, ed il documento, il file creato, come la carta. Concettualmente non solo un unico modello di penna può vergare un pezzo di carta, no? Lo studioso, come l’utente in genere, non può diventare dipendente dal programma che acquista.
- Per non essere costretti dal mercato o da un singolo produttore a cambiare le proprie abitudini, per non essere costretti dal mercato a cambiare i propri strumenti, il proprio computer, per seguire un’obsolescenza programmata basata sull’uso di sempre maggiori risorse per fare le stesse cose. Una lettera si scriveva 10 anni fa come si scrive oggi. Perché i programmi che mi consentono di scriverla sono sempre più complessi, sofisticati, pesanti da far girare?

Segue quindi la descrizione di alcuni strumenti che

l'utente in genere ed a fortiori lo studioso deve conoscere per proteggere il proprio lavoro.

I Documenti

Un documento di testo è un file che rappresenta in forma digitale la carta. Tutte le informazioni nel computer sono contenute in migliaia di file che hanno precise caratteristiche e si distinguono tra loro mediante un "nome del file" direttamente visibile all'utente ed una sorta di "cognome" che segnala al computer come trattare quello specifico file. Questo "cognome" solitamente non è visibile all'utente ma fa parte integrante del "nome complessivo" del file, poiché lo identifica come tipologia così come il nome ed il cognome identificano un individuo. Al computer non importa il "nome del file" (purché sia univoco), ma grazie ai diversi "cognomi", a lui comprensibili, riconosce i diversi tipi di file e può adottare dei comportamenti particolari per ogni tipo.

È importante per l'utente, quindi per lo studioso, riconoscere questo "cognome" utile alla macchina che d'ora in poi chiameremo **formato** poiché proprio nei formati diversi si annidano pericolose insidie.



Il formato definisce proprio la "forma" di un file. Se si vuole modificare un file, ad esempio un documento di testo, è necessario avere un programma in grado di gestire quella "forma" di file. Immaginiamo il gioco da bambini nel quale si devono far passare degli oggetti di legno con diverse forme attraverso un pannello forato in modo da riprodurre tali forme. Gli oggetti sono i nostri file, il pannello forato sarà il programma. Se abbiamo un oggetto di forma quadrata avremo grandi difficoltà nell'inserirlo in un pannello con un foro a forma di stella o di cerchio, quindi avremo una **incompatibilità**. Se il programma non gestisce un formato non ci consentirà di visualizzarlo e di modificarlo; se non esiste più un programma adatto a gestire un formato, potremmo aver perso i dati in esso contenuti poiché non più visualizzabili, eventualità detta **obsolescenza digitale**³, sufficientemente rappresentata da Vint Cerf⁴, uno dei padri di internet, che ha evocato il rischio di una "epoca nera digitale" o "medioevo digitale"⁵.

Il formato indica le modalità con cui i dati vengono rappresentati elettronicamente in modo che i programmi possano elaborarli; per elaborare un file, un programma deve sapere come i dati vi sono "memorizzati" cioè in che forma, ovvero formato, sono rappresentati⁶.

Solo alcuni formati sono riconosciuti come standard internazionali, quelli cui lo studioso dovrebbe attenersi, altri hanno caratteristiche tutte loro. Sorprendente è scoprire

che i formati più comuni sono quelli NON standard, divenuti nel tempo "*standard de facto*" grazie alla diffusione del programma che li crea, ad esempio Microsoft Office. Utilizzare un formato non standard vuol dire consegnare il proprio lavoro alla volontà ed alle scelte aziendali di un produttore. Quando vorrà cambiare, quindi abbandonare o modificare il vecchio formato oppure imporre delle *royalties* per il suo uso, nessuno potrà impedirlo⁷.

Negli ultimi anni il sistema operativo evita di mostrare i formati agli utenti facendo vedere solamente il nome del file ed assegnando ad ogni formato un'icona diversa.

Versione	Nome del file	Formato	Icona
Microsoft Word 2003	Nome-del-file	.doc	
Microsoft Word 2013	Nome-del-file	.doc e .docx	
LibreOffice Writer	Nome-del-file	.odt	

Il "nome-del-file" è sempre visibile, il formato no.

Quindi i formati possono essere:

- Standard o NON standard;
- Proprietari o Aperti.

Dello standard internazionale ISO parleremo più avanti con maggior dettaglio in quanto lo STANDARD è sicuramente uno degli strumenti dello studioso, ora distinguiamo tra proprietari e liberi:

I formati proprietari:

«Diremo che un formato è proprietario se il modo di rappresentazione dei suoi dati è opaco e la sua specifica non è pubblica. Si tratta in genere di un formato sviluppato da un'azienda di software per codificare i dati di una specifica applicazione che essa produce: solo i prodotti di questa azienda potranno leggere correttamente e completamente i dati contenuti in un file a formato proprietario. I formati proprietari possono inoltre essere protetti da un brevetto e possono imporre il versamento di royalty a chi ne fa uso»⁸.

Microsoft Office, ad esempio, è un software proprietario ed i programmi in esso contenuti, Word, Excel, Powerpoint salvano per impostazione predefinita in formato chiuso e NON standard.

I formati aperti:

«Diremo che un formato è aperto se il modo di rappresentazione dei suoi dati è trasparente e/o la sua specifica è di pubblico dominio. Si tratta generalmente (ma non esclusivamente) di standard fissati da autorità pubbliche e/o istituzioni internazionali il cui scopo è quello di fissare norme che assicurino l'interoperabilità – ovvero la capacità di la-

vorare insieme, di cooperare – *tra software. Non mancano tuttavia casi di formati aperti promossi da aziende, che hanno deciso di rendere pubblica la specifica dei propri formati*»⁹.

LibreOffice, ad esempio, è un software libero ed i programmi in esso contenuti salvano per impostazione predefinita in un formato aperto e descritto nello Standard ISO/IEC 26300.

Fatta la distinzione tra i due tipi di formati, aperti e chiusi, invito chi voglia conservare nel tempo il proprio lavoro a farlo esclusivamente con formati aperti, poiché le caratteristiche di questi formati sono di pubblico dominio, sono proprietà di tutti, quindi non sono soggetti a modifiche di orientamento aziendale o di mercato, assicurano la sana concorrenza e ci rendono gli unici proprietari del nostro lavoro.

Standard ISO / IEC

La produzione di documenti di testo digitali soffre, come in molti altri casi, dell'antagonismo tra due tipi di standard nati da gestazioni e producenti effetti molto differenti, sono: lo standard *de jure*, emesso da un ente certificatore (ISO, W3C, ANSI, ecc.); e lo standard *de facto*, imposto di fatto (industry standard) e non garantito poiché il proprietario ha la piena libertà di modificarlo.

ISO è un'organizzazione internazionale atta alla standardizzazione in senso generale, è quindi un ente che stabilisce gli standard *de jure*. Con circa 160 paesi aderenti è la più importante organizzazione a livello mondiale per le norme tecniche¹⁰. Lo standard è sicuramente uno strumento dello studioso, non si potrebbero certo scrivere documenti scientifici esprimendo misure in cubiti o stadi od il volume in cotili, per ottenere un documento di qualità è necessario utilizzare gli standard internazionali. Oltre allo standard nel testo, nel caso di un documento digitale, è necessario seguire strettamente uno **standard del formato** per evitare i problemi relativi alla obsolescenza digitale.

Esistono due formati standard certificati da ISO/IEC:

- **ODF – Open Document Format** – Formato aperto definito dallo standard ISO / IEC 26300:2006 definisce uno *“schema per applicazioni di ufficio e sua semantica. Lo schema è adatto per documenti d'ufficio, includendo documenti di testo, fogli di calcolo, tabelle e documenti grafici come disegni o presentazioni, ma non è ristretto a questo tipo di documenti. [...]”*¹¹. Recepto in Italia dall'UNI, l'ente italiano di normazione con il codice UNI CEI ISO/IEC 26300 nel 2007 con un significativo comunicato stampa: *“La norma tecnica UNI CEI ISO/IEC 26300, [...] rappresenta il capitolo finale di un lungo percorso che ha portato per la prima volta alla definizione, in ambito normativo, di uno standard universale per i documenti elettronici, svincolandoli dalla applicazione che*

*li ha generati [...] Ad oggi la gran parte dei documenti che siamo soliti gestire con il computer sono generati da software commerciali, ognuno dei quali possiede un proprio formato proprietario. Ciò comporta che, per poter leggere o modificare tali file, l'utente debba avere a disposizione lo stesso programma (ed in alcuni casi anche la stessa versione) utilizzato dall'estensore del documento oppure un programma che faccia da filtro per la visualizzazione. Il formato OpenDocument [...] permette di superare questi vincoli. Esso è infatti ciò che viene definito uno “standard aperto”, libero pertanto da restrizioni tecniche e da diritti d'autore del produttore”*¹².

- **OOXML – Office Open XML** – Dopo un primo rifiuto di ISO nel 2007 questo standard, promosso da Microsoft e basato sui suoi formati chiusi già esistenti, è stato certificato a seguito di numerose modifiche nel 2008 con codice ISO / IEC 29500. Veri e propri giganti del settore hanno cercato di far desistere ISO dal certificare un secondo standard sullo stesso “argomento”, tra questi Google, Oracle, Red Hat, IBM, Sun Microsystems, Free Software Foundation Europe. Le argomentazioni erano molteplici, ad esempio il fatto che non ci fosse alcuna garanzia certa ma solo una “promessa” sul fatto che si possa scrivere software che utilizzi completamente o in parte la specifica OOXML senza essere perseguibili per infrazione di brevetto o senza dover pagare licenze di brevetto alla Microsoft¹³. La *Open Specification Promise* (OSP), la promessa di cui sopra, è in sostanza un impegno unilaterale (pubblicato su un'apposita pagina del sito di Microsoft) a non intraprendere azioni legali basate sulla proprietà intellettuale contro chi implementerà quello standard. Posizione giuridica molto labile per un colosso monopolista ed influente¹⁴.

Esistono quindi due standard per i documenti di testo: ODF ed OOXML. Quale utilizzare? Formalmente sono ambedue standard *de jure*, tuttavia è importante considerare da chi e perché sono stati promossi prima di affidare i propri documenti ad uno od all'altro formato. Mentre l'ODF veniva utilizzato già molto tempo prima del riconoscimento come standard ISO per le sue caratteristiche native, ovvero per il fatto di essere un formato aperto, l'evidenza mostra che l'OOXML di Microsoft è nato con l'intento di dare caratteristiche di standard ai formati nativi e proprietari di Microsoft Office. La stessa Microsoft, non utilizza di default lo standard OOXML per il salvataggio dei documenti nei suoi programmi.

Visto quanto detto sopra, lo studioso che vuole preservare il suo lavoro deve utilizzare, per qualsiasi tipo di documento (immagine, testo, presentazione, eccetera) il *formato aperto standard*, come l'ODF, per testi, presentazioni, fogli di calcolo, eccetera...

Detto tipo di formati consente di avere quattro importanti benefici:

- i. **Indipendenza** – La documentazione pubblica e completa del formato consente l'indipendenza da uno specifico prodotto e fornitore;
- ii. **interoperabilità** – Usando formati aperti sistemi diversi (diverse versioni di programmi, diverse macchine, ecc.) possono condividere gli stessi dati;
- iii. **Neutralità** – I formati aperti non obbligano ad usare uno specifico prodotto (ad es. programma di scrittura) lasciando libero l'utente di scegliere sulla base del rapporto qualità/prezzo;
- iv. **Persistenza** – Tutela del patrimonio informativo nel tempo a fronte del mutamento tecnologico.

Programmi di scrittura

Spesso si indica col **nome di un programma** un **tipo documento**: "Preparo un PowerPoint" oppure "Scrivo un paio di pagine di Word". Quante volte lo abbiamo detto/sentito? La confusione è indotta dal monopolio e dallo standard de facto di Microsoft. Bisogna tuttavia scindere quello che è il **programma** che si usa dal **documento** che si crea, poiché si possono usare diversi programmi per creare un documento di testo.

Mentre ODF viene utilizzato come formato di salvataggio di default, il formato OOXML per Microsoft è, alla prova dei fatti, una mossa commerciale. La dimostrazione è evidente quando si salva un documento senza toccare alcuna impostazione. Programmi come LibreOffice che rispettano lo standard ODF, salvano in maniera nativa in quel formato; Microsoft Office, anche nella versione 2013, NON salva per default in formato standard OOXML, ma serve una selezione manuale utilizzando il tasto "salva come..." per poi scegliere il formato corretto, forse per pura casualità, il penultimo di-

spicabile. Oltre a ciò il programma Word **non consente all'utente di distinguere** tra "DOCX standard OOXML" e "DOCX non standard".

Come si può notare nella figura 1, la finestra "salva con nome" di Word mostra numerosi formati di salvataggio. Quello utilizzato dalla maggior parte degli utenti, corrispondente al bottone "Salva" viene attivato per default ed è evidenziato in figura dal primo quadrato rosso; il file creato in questo modo NON è un file standard; qualora volessimo salvare secondo lo standard OOXML in Word dovremmo selezionare la voce "Documento Strict Open XML (*.docx)", corrispondente al secondo quadrato rosso. Una procedura tutt'altro che agevole che ci porta al risultato in figura 2.

I due file sembrano identici, neanche tra le proprietà è possibile trovare differenze. La confusione indotta da que-

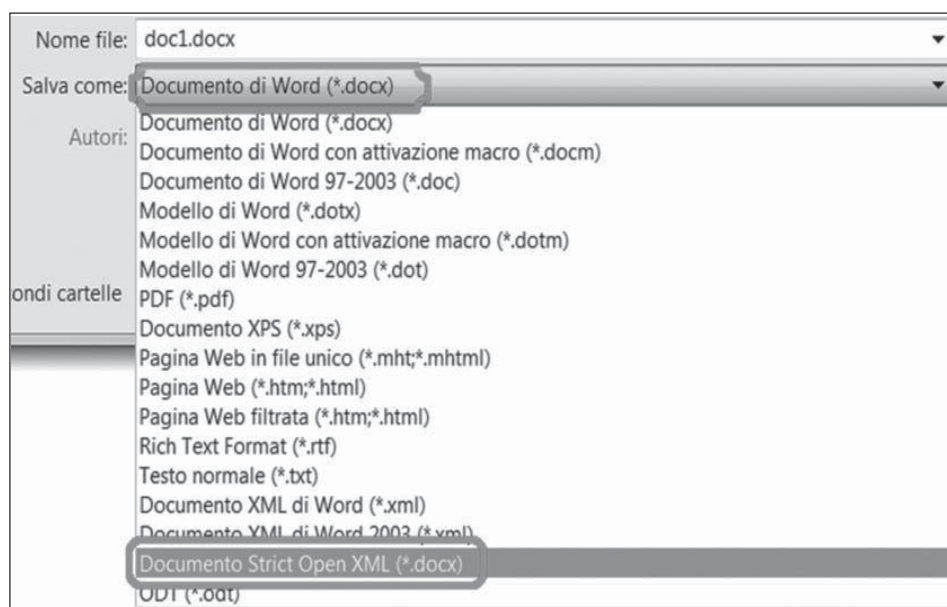
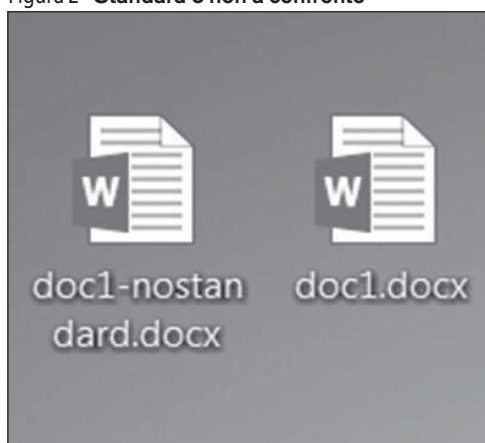


Figura 1 - Finestra "salva come..." di Word 2013

Figura 2 - Standard e non a confronto



sta scelta è pericolosa per chi vuole preservare i propri documenti dall'obsolescenza digitale. Salvare in formato non standard e chiuso come Microsoft Word salva per default è sicuramente da evitare.

Per completezza descrivo brevemente i due programmi più volte citati nel testo:

- **Microsoft Office** – programma proprietario protetto da Copyright che salva per impostazione predefinita i suoi documenti in formato NON

standard – DOC o DOCX. Dalla versione 2013, scegliendo manualmente, può salvare in formato standard ISO, mantenendo il formato DOCX, creando delle incertezze nell'utente circa l'aderenza o meno allo standard ISO¹⁶ anch'esso nato sotto numerose critiche, dubbi e perplessità^{17 18}. Nasce oltre 25 anni fa, evoluto 15 versioni diverse, vede nella versione 2007 una svolta nella veste grafica caratterizzata da schede chiamate "ribbon". È il software d'ufficio più conosciuto;

- **LibreOffice** – programma libero con trent'anni di storia alle spalle¹⁹. È stato protagonista di grandi opere di migrazione da programmi proprietari ed adottato come software di ufficio da interi ministeri ed uffici pubblici di stati come la Francia, Spagna, Taiwan, la pubblica amministrazione di Monaco ed ultimo solo in ordine di tempo dal Ministero della Difesa Italiana²⁰ che oltre a scegliere il programma sceglie specificatamente il formato ODF²¹. Il suo codice sorgente è liberamente disponibile, liberamente scaricabile ed utilizzabile poiché è protetto da una licenza di libero uso e distribuzione²².

Salva per impostazione predefinita in formato ODT, standard ISO²³ dal 2006.

Una volta descritte le caratteristiche degli standard, dei documenti di testo elettronici e dei programmi di scrittura, strumenti essenziali forieri di nuove difficoltà, l'utente ha a disposizione gli strumenti base per l'estensione di un documento di testo che abbia le anzidette importanti caratteristiche di indipendenza, interoperabilità, neutralità e persistenza.

Ovviamente il mondo in cui ci siamo affacciati è molto più vasto e non basta un articolo per descrivere tutto. Rimangono da considerare i formati delle immagini, delle presentazioni e dei fogli di calcolo (anche se ne è stato fatto breve accenno), i Font con i quali scriviamo, i supporti fisici nei quali conserviamo i dati stessi e molto altro. Esiste tuttavia una norma che ritengo di buon senso che mi permetto di racchiudere in una regola aurea:

"Standard" è importante, ma "Aperto, Libero e Standard" è meglio.

NOTE

1) Edizioni e relative versioni di Microsoft Office https://it.wikipedia.org/wiki/Microsoft_Office

2) Lo standard Portable Document Format https://it.wikipedia.org/wiki/Portable_Document_Format

3) Obsolescenza Digitale – https://it.wikipedia.org/wiki/Preservazione_digitale#Obsolescenza_digitale

4) VINTON G., Cerf is vice president and Chief Internet Evangelist for Google. He contributes to global policy development and continued spread of the Internet. Widely known as one of the "Fathers of the Internet". <http://research.google.com/pubs/author32412.html>

5) VINTON G., Cerf - 20th international conference on Electronic Publishing, Göttingen, June 6-9, 2016. – http://www.bibliothek-oechslin.org/bibliothek-oechslin.ch/forschung/digital_humanities/electronic-publishing-2016

6) Ministero per l'innovazione e le Tecnologie – *Indagine conoscitiva sul software a codice sorgente aperto nella Pubblica Amministrazione* - 2003 - pag. 17.

7) Si veda il caso dei proprietari del formato per le immagini "GIF".

8) SIMONE ALIPRANDI - *Apriti standard! Interoperabilità e formati aperti per l'innovazione tecnologica* - Ed. Ledizioni 2010.

9) SIMONE ALIPRANDI - stessa fonte nota 11.

10) ISO standard – <http://www.iso.org/iso/home/about.htm>

11) ISO 26300:2006 – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43485

12) SIMONE ALIPRANDI - Apriti Standard 2010.

13) SIMONE ALIPRANDI - Apriti Standard 2010 - Capitolo 4.3.

14) SIMONE ALIPRANDI - Apriti Standard 2010 - Capitolo 4.4.

15) Ministero per l'innovazione e le Tecnologie - *Indagine conoscitiva sul software a codice sorgente aperto nella Pubblica Amministrazione* - 2003 - pag. 18.

16) ISO/IEC 29500-1:2012 - defines a set of XML vocabularies for representing word-processing documents, spreadsheets and presentations, based on the Microsoft Office 2008 applications. It specifies requirements for Office Open XML consumers and producers that comply to the strict conformance category.

17) FreeSoftware Foundation – divisione Europe – sei domande che rappresentano i dubbi della comunità del software libero. <https://fsfe.org/activities/os/msooxml-questions.en.html>

18) WikiPedia - Google, FSF e IBM hanno chiesto ad ISO di non accettare il nuovo standard. https://it.wikipedia.org/wiki/Office_Open_XML

19) LibreOffice - <https://it.wikipedia.org/wiki/LibreOffice>

20) Libreatalia - Accordo tra Libreatalia onlus e Difesa per l'adozione del prodotto LibreOffice. - <http://www.libreatalia.it/accordo-di-collaborazione-tra-associazione-libreatalia-onlus-e-difesa-per-ladozione-del-prodotto-libreoffice-quale-pacchetto-di-produttivita-open-source-per-loffice-automation/>

21) Stato Maggiore Difesa – SMD-I-002 – Standardizzazione dei formati dei documenti elettronici della Difesa - 2016.

22) LGPL – Lesser General Public Licence.

23) Standard ISO/IEC 26300 – http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43485.