

Formati di file aperti e standard ISO¹ per documenti informatici

Saverio Rubini²

Comunicazione, documenti di testo e Word

Nella comunicazione scritta, *Microsoft Word* è lo strumento informatico più diffuso e utilizzato nel mondo. Perché il contenuto di un documento creato con il programma possa essere trasferito a un altro utente, occorre che siano verificate alcune condizioni:

- che il documento venga salvato in un *file*;
- che il programma utilizzato dal destinatario sia in grado di aprirlo correttamente, ossia che ne riconosca il *formato*³;
- che la lingua utilizzata per il testo sia comune al mittente e al destinatario.

Dando per assunta l'ultima condizione, il salvataggio in un file è l'operazione preliminare a qualsiasi tipo di trasferimento del documento a chi ha interesse ad apprendere il contenuto: stampa, invio come allegato di posta elettronica, pubblicazione nel Web.

Per salvare un testo scritto con il programma *Microsoft* in un'unità di memoria di massa, si esegue la sequenza *File (Pulsante di Office in Word 2007), Salva* e gli si assegna un nome. Non accade di frequente che qualcuno si preoccupi di verificare quale sia l'estensione che viene assegnata al file, leggibile nella casella *Salva come*. Conoscerla, invece, corrisponde a sapere qual è il formato utilizzato.

Solo dopo aver eseguito l'azione, la maggior parte degli utenti scopre di aver creato un file con *estensione DOC*. Nomi del tipo "corso.doc", "relazione.doc" e altri nomi simili, quindi, sono identificati come: "i file di Word".

In termini più formali, *Microsoft Word* è un programma di gestione testi che, quando registra i propri documenti in un supporto di memoria, assegna al file un *nome* e un'*estensione*. Il nome viene fornito dall'utente. L'estensione viene assegnata automaticamente dal programma secondo impostazioni predefinite, che l'utente può cambiare. Per esempio, in *Word 2010* l'indicazione del formato predefinito di salvataggio viene eseguita con la sequenza: *File, Opzioni, Salvataggio*. La scelta viene operata nella casella a discesa *Formato* per il salvataggio dei file.

¹ *International Organization for Standardization*: <http://www.iso.org>

² Ingegnere elettronico, autore di libri e di articoli di informatica, docente in corsi di formazione professionale, funzionario dell'Agenzia delle Entrate (<http://www.srubini.it>)

³ Il *formato* di un file è l'insieme delle regole tecniche utilizzate da un programma per salvare i propri documenti in un supporto di memoria informatico. Solo i programmi che utilizzano lo stesso formato possono aprire quel file e gestirne correttamente il contenuto.

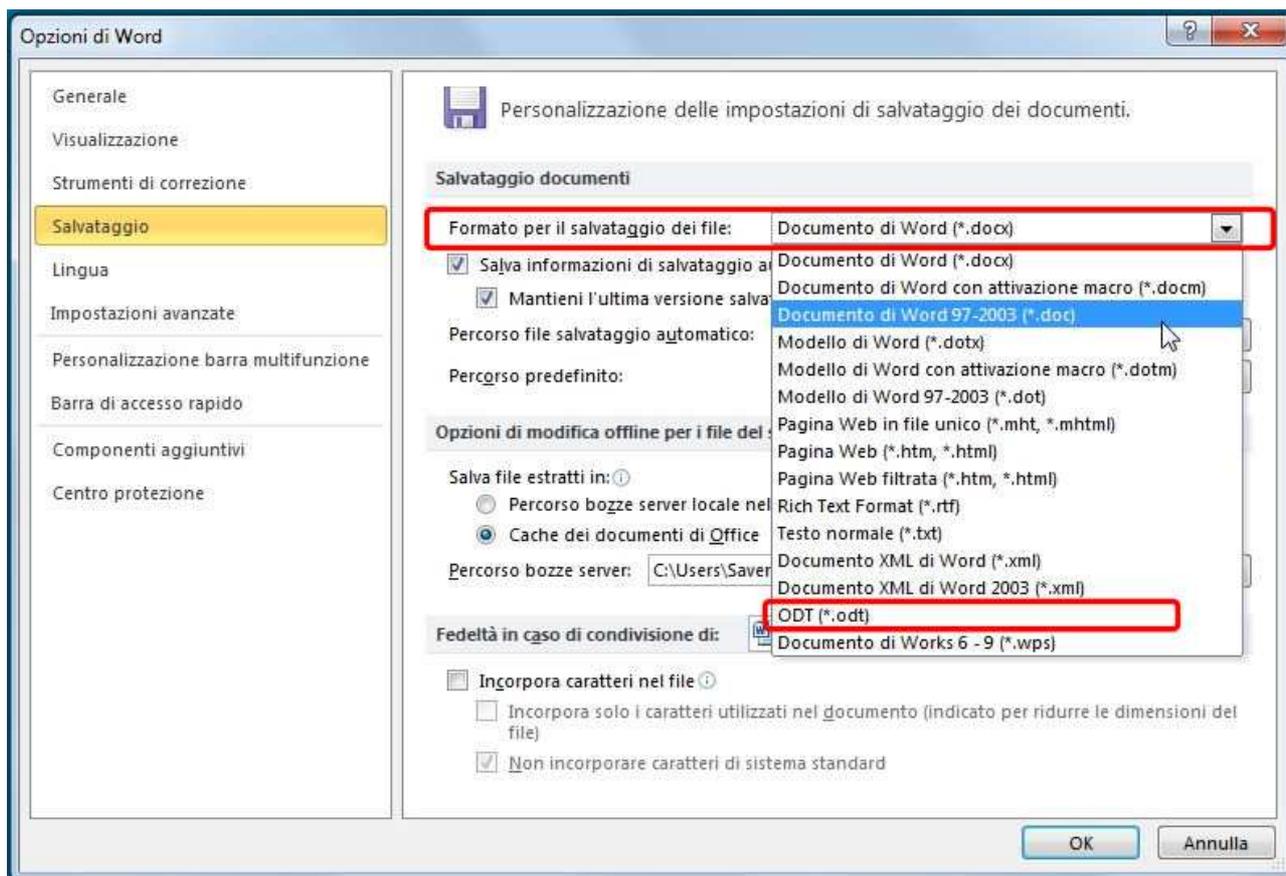


Figura 1. Come impostare il formato predefinito dei file in *Word 2010*.

DOC, DOCX & C.

Fino alla versione 2003, l'estensione predefinita dei file salvati da *Microsoft Word* è stata DOC. Nelle versioni successive del programma (2007 e 2010) è cambiata, diventando DOCX. Quando si prova a leggere un file DOCX con una versione di *Word* precedente alla 2007, però, non si riesce a farlo e nella finestra appare un errore. Il motivo è presto spiegato: il contenuto del file è scritto in un formato che le versioni più datate non possono riconoscere perché è stato creato dopo il loro rilascio. C'è la possibilità di installare gratuitamente un modulo software in Office 2003 per recuperare la compatibilità, fornito dalla stessa *Microsoft*. Bisogna saperlo, però, e installarlo in tutti i computer con versioni datate di *Office*.

Se vengono visualizzati messaggi di errore e generati problemi di compatibilità, allora, perché è stato creato il DOCX? Una risposta veloce potrebbe essere: "per avere un formato aperto, orientato all'interoperabilità tra i sistemi". Una risposta più articolata, invece, che sia anche più chiara per i non addetti ai lavori, deve partire necessariamente un po' lontano nel tempo.

ODT: chi era costui?

Intorno al 2000 ci si è posto con maggiore forza rispetto al passato il problema che i file DOC contengono dati *registrati con regole tecniche note solamente a Microsoft*. Il relativo formato, infatti, non è pubblico, ma la sua proprietà intellettuale appartiene alla società americana. Per esempio, se qualcuno provasse ad aprire un file DOC con *Blocco Note*, nella finestra vedrebbe una lunga sequenza di strani caratteri, normalmente non comprensibili e apparentemente non legati al testo presente nel documento.

Non conoscendo le regole tecniche con cui è formato il contenuto dei DOC, la difficoltà di comprendere cosa sia effettivamente registrato in un file da parte di un essere umano si riverbera anche in quella di gestirne automaticamente i dati con altri programmi. Le conseguenze sono di due tipi:

- ipotetici rischi per la sicurezza degli utenti;
- impossibilità di recuperare automaticamente i contenuti del file con applicazioni informatiche.

La questione “*sicurezza*” deriva dal fatto che, proprio perché incomprensibili, nessuno può garantire che tra gli strani caratteri potrebbero esserci alcuni con dati relativi al nome utente, al computer, ad altri elementi sensibili o privati legati a chi lavora su quella macchina. L'impossibilità di estrarre dati tramite programma, invece, impedisce lo scambio automatico di dati tra computer e, più in generale, tra diversi sistemi di elaborazione, altrimenti detta *interoperabilità*.



Figura 2. Scambio di dati tra computer.

Sempre intorno al 2000, si è cominciato a fornire soluzioni concrete per risolvere questi problemi e superare le diffidenze create in seno alle strutture governative e alle organizzazioni internazionali sull'utilizzo dei file DOC. A questo scopo, per salvare i propri file i programmi della *suite* per l'automazione delle attività di ufficio *OpenOffice* hanno cominciato a utilizzare i formati ODF⁴, creati sotto l'egida dell'organizzazione internazionale OASIS che si occupa di standard aperti in informatica. ODF⁵ sta per *Open Document Format*, “formato aperto per documenti” ed è basato su XML⁶ (*eXtensible Markup Language*).

⁴ http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

⁵ Un'introduzione ai formati *OpenDocument* in inglese con riferimenti e risorse aggiuntive:
<http://www-03.ibm.com/able/resources/odfintro.html>

⁶ <http://www.w3.org/XML/>



Figura 3. Home page del sito *Openoffice.org*.

L'XML è il linguaggio di marcatura di dati e strutture di documenti, aperto per definizione. Il suo utilizzo garantisce libero accesso ai contenuti dei documenti e interoperabilità tra i sistemi di elaborazione dati che trasmettono e ricevono i file con cui sono stati creati.

I formati gestiti da ODF riguardano diversi tipi di documenti, che tra l'altro comprendono:

- documenti di testo con estensione ODT (*Open Document Text*);
- fogli di calcolo con estensione ODS (*Open Document Spreadsheet*);
- grafica di presentazione con estensione ODP (*Open Document Presentation*).

In ODF possono essere creati anche file di *data base* (ODB), di disegni (ODG), di grafici (ODC), di formule (ODF) e di altri tipi ancora.

I file ODT non hanno alcun legame tecnico quelli DOC. Un programma che riconosce entrambi i formati, però, è in grado di leggere l'uno e di salvare nell'altro e viceversa. Che si utilizzi l'uno o l'altro dei formati per memorizzare un file, l'aspetto e la struttura del documento rimangono le stesse.

Standard e legislazioni nazionali

La famiglia di formati ODF è stata ritenuta talmente affidabile da essere proposta e accettata come standard ISO-26300⁷. Anche per questo, tra la metà e a fine del 2007 è stato scelto per l'utilizzo negli uffici delle amministrazioni pubbliche in Belgio, in Sudafrica e in Giappone.

The screenshot shows the ISO website interface. The main header includes the ISO logo and the text 'International Organization for Standardization' and 'International Standards for Business, Government and Society'. A navigation menu at the top includes 'Home', 'Products', 'Standards development', 'News and media', and 'About ISO'. The breadcrumb trail reads: 'Products > ISO Standards > By TC > JTC 1 Information technology > SC 34'. The left sidebar contains a list of links such as 'ISO Store', 'ISO Standards', 'By ICS', 'By TC', 'How to use the ISO Catalogue', 'Standards in action', 'Management and leadership standards', 'The ISO portfolio', 'FAQs', 'Country codes (ISO 3166/MA)', 'Directory of aerospace standards', 'Publications and e-products', 'ISO Concept Database (ISO/CDB)', and 'Copyright'. The main content area features the title 'ISO/IEC 26300:2006' with a small icon, followed by the subtitle 'Information technology – Open Document Format for Office Applications (OpenDocument) v1.0'. Below this is a 'Media and price' section with a table:

Language	Format	Add to basket
English	PDF (14 198 kB) (includes electronic inserts)	CHF 352,00
English	PDF on CD (includes electronic inserts)	CHF 352,00

A note below the table states: 'Note: The electronic version of this International Standard can be downloaded from the ISO/IEC Information Technology Task Force (ITTF) web site'. Below the table is a 'General information' section with the following details:

Number of Pages: 722	
Edition: 1 (Monolingual)	ICS: 35.240.30
Status: Published	Stage: 60.60 (2006-11-30)

Figura 4. Pagina Web del sito *iso.org* con i riferimenti dello standard *OpenDocument*.

In Italia il formato è stato ratificato come standard UNI CEI ISO/IEC 26300⁸. A seguito dei risultati delle attività del gruppo TAC (*Telematics between Administrations Committee*), l'Unione Europea ha approvato la raccomandazione a utilizzare i formati aperti basati su XML. In Italia, il comune di Arezzo ha deliberato l'utilizzo di ODF come formato da utilizzare dal 30 giugno 2008.

Nelle pagina web <http://linguistico.sourceforge.net/pages/opendocumentformat.html> e in quelle collegate, viene riportato un corposo gruppo di riferimenti e di collegamenti ipertestuali a documenti in Internet sull'attività amministrativa relativa alle disposizioni ufficiali emanate da nazioni e da organizzazioni internazionali, come l'ONU e la NATO, sull'adozione dell'ODF.

⁷ Pagina web del sito ISO (*International Organization for Standardization*) con i riferimenti allo standard *Open Document Format*: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485

⁸ <http://webstore.uni.com/unistore/public/productdetails?productId=UNICI2630000!EEN>

Microsoft crea i DOCX

Dato l'interesse suscitato dai formati "aperti", con *Office 2007* Microsoft rilascia i formati *OpenXML*⁹: DOCX, XLSX e PPTX. La "X" aggiunta alla fine delle estensioni utilizzate per anni dai programmi della suite *Office* indica che anche questi formati sono basati su XML.

Il fatto che Microsoft si riservi la proprietà intellettuale del formato e la possibilità esclusiva di gestirlo, però, ne blocca il processo di standardizzazione. A questo si aggiunge la sua scarsa diffusione perché chi lavora ancora con le versioni precedenti del pacchetto *software* non riesce a leggere i file con i nuovi formati.

Dopo una serie di informative tecniche inviate da Microsoft, comunque, il 15 agosto 2008 *OpenXML* viene definito uno standard con la norma ISO/IEC DIS 29500¹⁰, parti da 1 a 4.

Per i testi, allora: DOC, DOCX oppure ODT?

Dal punto di vista formale, chi lavora nella pubblica amministrazione non dovrebbe avere dubbi. I vantaggi che derivano dalla maggiore sicurezza e standardizzazione degli *Open Document Format*, poi, dovrebbero favorire anche un maggiore utilizzo dei formati aperti nel mondo del lavoro e tra i privati. Il problema vero, però, sta nell'indolenza di chi lavora con il computer, spesso restio a cambiare le proprie abitudini, nonostante il farlo comporti l'ottenimento di migliorie e di benefici a vari livelli.

Anche se può sembrare paradossale, chi può aiutare i formati ODF a espandersi è proprio *Microsoft*. Infatti nella versione di *Office* più recente, la 2010 (disponibile da più di un anno), i formati ODF sono gestiti normalmente in lettura e in scrittura. Li si trova nella casella citata nella fase di salvataggio dei documenti all'inizio di questo articolo: *Salva come*. Il nome del formato appare anche in chiaro nella colonna dei tipi di file gestiti dall'applicazione. Per chi svolge attività di redazione di documenti di testo, quindi, non ci sono più alibi per utilizzare formati aperti e standard come ODT per i propri file.

⁹ <http://www.microsoft.com/interop/openup/openxml/default.aspx>

¹⁰ La prima parte dello standard con le definizioni e i riferimenti del tipo di marcatura è disponibile nel sito Internet dell'ISO nella pagina web:

http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59575.

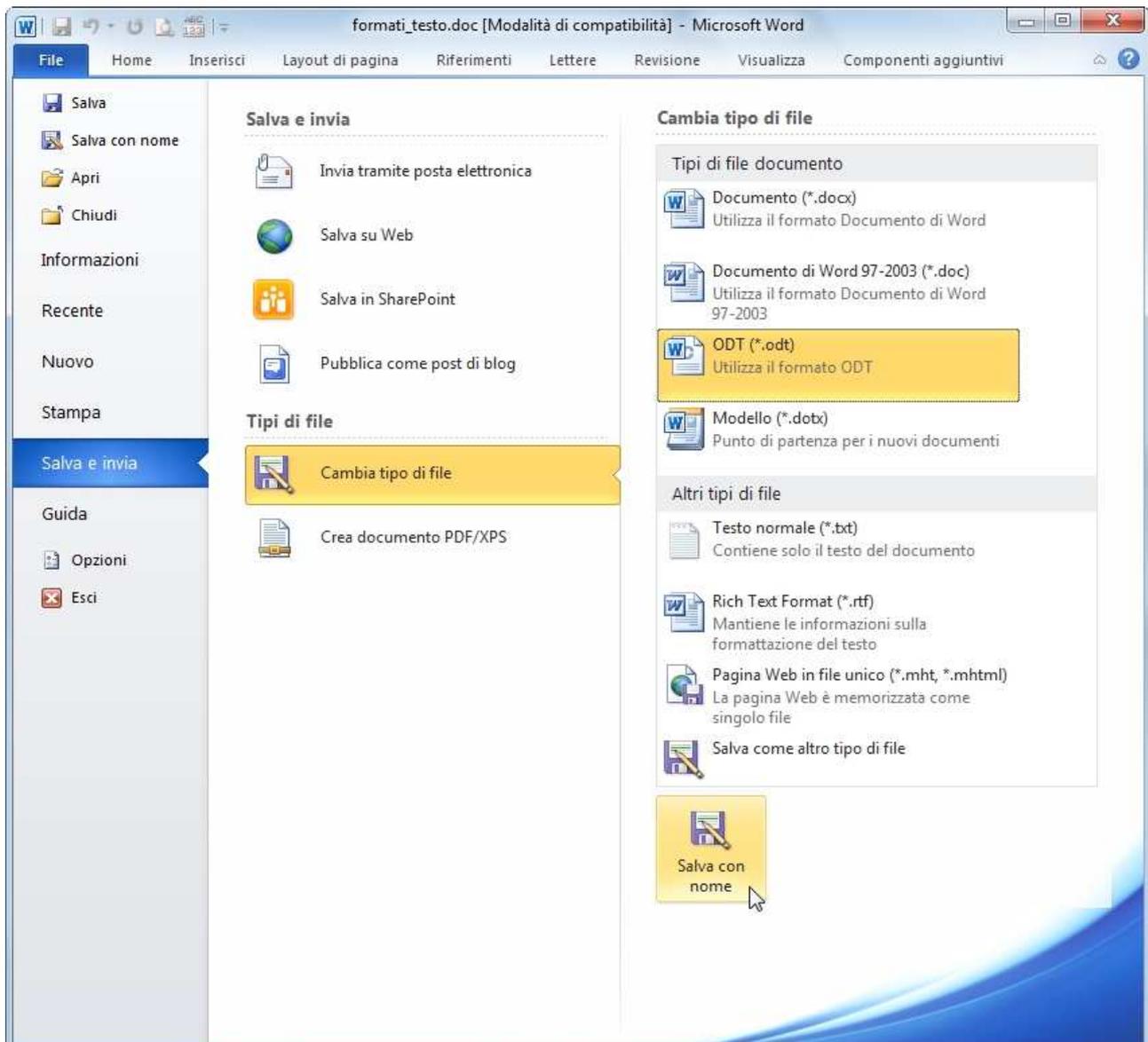


Figura 5. Il salvataggio dei documenti in file con estensione ODT, in Word 2010.

