

## Premessa

In generale tutti i formati che rispondono alle seguenti caratteristiche possono essere considerati accettabili sia per quanto riguarda l'archivio corrente che per quanto riguarda la conservazione:

- aperti, documentati, non proprietari, trasparenti: si tratta di formati che sono standard "de jure" o standard "de facto", le cui specifiche siano comunque rese pubbliche. Il criterio di trasparenza implica anche una maggiore facilità nella fruizione del documento tramite un maggior numero di sw anche con semplici funzioni di base (ad esempio un semplice editor di testo come Blocco note). Gli standard "de jure" sono preferibili agli standard "de facto"; nel primo caso un organismo internazionale ha certificato le specifiche del formato, ad esempio: il PDF e' ISO 32000, l' ODF e' ISO 26300;
- robusti: in caso di corruzione e' possibile comunque eseguire un recupero parziale e in ogni caso una minima perdita di bit non ne compromette la comprensione;
- stabili nel tempo: frequenza di modifiche nel tempo delle specifiche del formato, tali da garantire comunque compatibilità all'indietro (utilizzo del documento con versioni obsolete del sw utilizzato per generarlo) e in avanti (utilizzo del documento con versioni più recenti del sw utilizzato per generarlo) senza sostanziale perdita di informazioni;
- diffusi: quanto più il sw che gestisce alcuni formati e' diffuso tanto più sarà stabile nel tempo il formato;
- auto-consistenti: contengono tutte le informazioni necessarie per essere visualizzati in modo corretto e completo;
- non modificabili: il documento non deve poter essere modificabile, almeno non facilmente;
- accessibili: possibilità di accesso alle informazioni del documento da parte di persone diversamente abili (il PDF prodotto tramite scansione non e' un documento accessibile, il PDF prodotto tramite conversione via sw e' un documento accessibile);
- assenza di protezioni: non siano presenti password o vincoli particolari collegati al documento

Tuttavia esiste un'altra serie di considerazioni che riduce ulteriormente il numero dei formati che possono essere considerati pienamente accettabili.

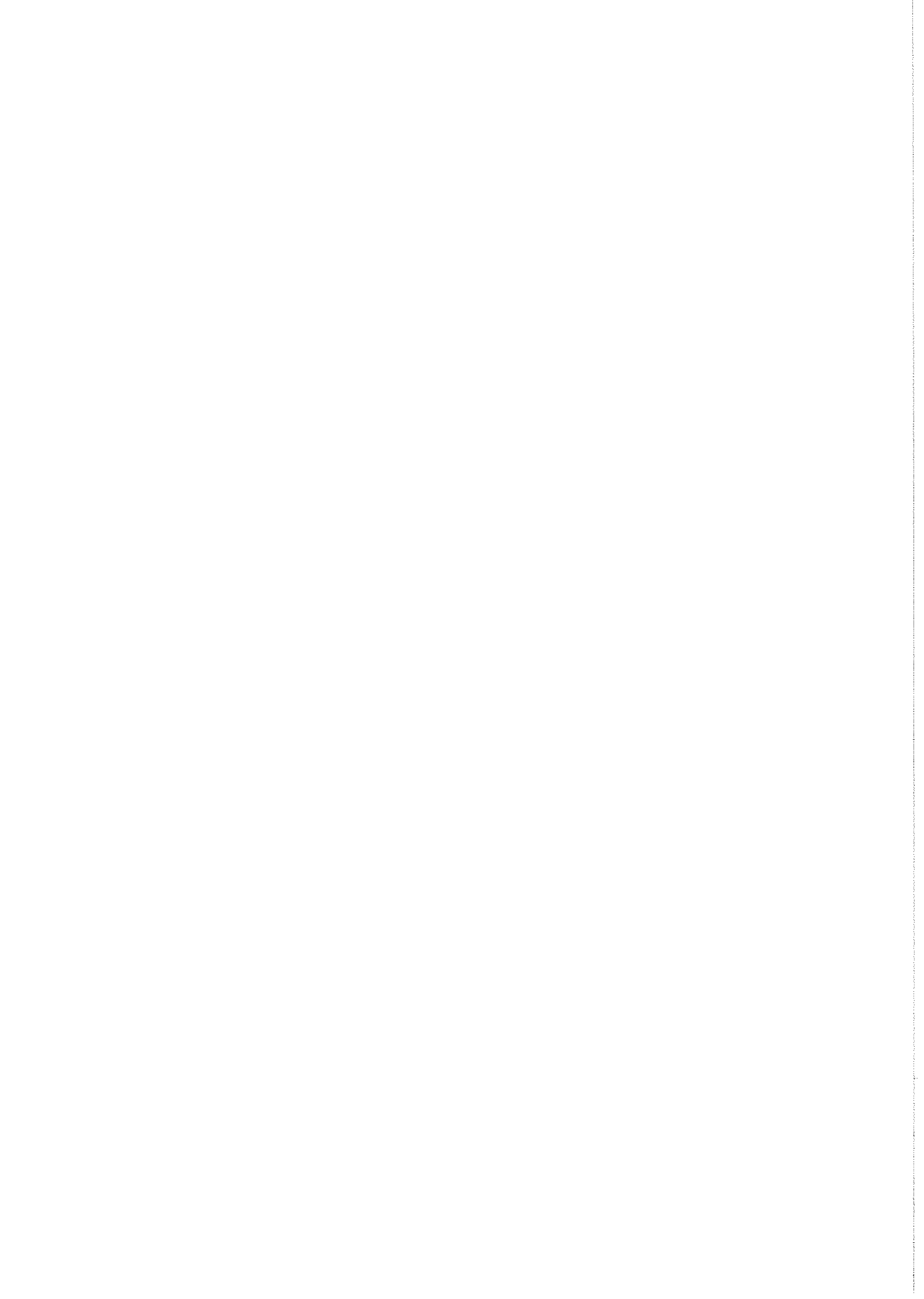
Nei documenti della famiglia MS-Office, in particolare Word (formato DOC) e della famiglia Open office (formato ODT) e' possibile inserire parti variabili (es: campi DATA che si autoaggiornano ogni volta che il documento viene aperto, macro o parti di codice eseguibile che fanno riferimento a banche dati che possono essere modificate o cessate nel tempo e comunque non più disponibili al momento dell'invio in conservazione), file audio e video. La presenza di informazioni che possono modificare il loro valore nel tempo può comprometterne la conservazione dal punto di vista informativo, sia a breve che a lungo termine: non garantiscono infatti l'immodificabilità del documento nel tempo.

Ricordiamo inoltre che ai sensi della attuale normativa il documento informatico non deve contenere macroistruzioni, riferimenti esterni, codici eseguibili od altri elementi tali da attivare funzionalità che possono modificarne il contenuto e che il documento informatico sottoscritto con firma digitale o altro tipo di firma qualificata non ha l'efficacia prevista dall'articolo 2702 del codice civile, cioè della scrittura privata.

Questi formati inoltre possono presentare significative differenze formali nella presentazione del contenuto al momento della visualizzazione tramite sw o versioni diverse da quelle con cui sono stati prodotti e risultano molto facilmente modificabili al momento della visualizzazione e della memorizzazione. In alcuni casi la corretta presentazione è garantita solo dalla versione del sw con cui è stato prodotto, che, nel caso dei formati doc, non è liberamente disponibile.

Potrebbero quindi creare problemi di visualizzazione e di apertura al momento del loro ricevimento da parte di altri.

D'altro canto questi formati rispondono sicuramente al criterio della diffusione.



Il nuovo formato di Word, DOCX, è un formato aperto perché la Microsoft ne ha reso pubbliche le specifiche (diventato nel 2008 standard "de jure"; UNI CEI ISO/IEC IS 29500:2008), tuttavia è un formato ancora poco diffuso che necessita di un software specifico e a pagamento per poterlo visualizzare correttamente (MS-Office 2007). Stessa considerazione vale per tutti i formati della famiglia Office 2007. Questa famiglia di formati non risponde al momento al criterio dell'ampia diffusione: oltre ad avere gli stessi problemi citati per i formati DOC e ODT.

Oltre ai programmi specifici che consentono la visualizzazione del solo tipo di documento supportato, esistono nel web dei sw di visualizzazione generici (viewer), in grado di rendere leggibili diversi formati di file di testo.

Questi sw possono essere scaricati sul PC oppure utilizzabili direttamente dal browser (vedi ad esempio: <http://guidami.blogspot.com/2010/12/aprire-online-documenti-microsoft.html> e <http://www.officeviewers.com/>)

Un formato che risponde alle caratteristiche elencate all'inizio del paragrafo è il formato XML, sviluppato dal consorzio W3C. È il formato secondo il quale vengono generate dai sistemi di protocollo tutte le ricevute previste per l'interoperabilità: segnatura di protocollo, conferma di ricezione, eccezione, etc...

Il formato XML, se non accompagnato da un foglio di stile tramite il quale è possibile visualizzarne il contenuto in modo formattato (se al suo interno sono previste informazioni di formattazione), risulta un elenco di variabili e valori assegnati a queste variabili, risulta quindi molto utilizzato nello scambio di dati tra applicazioni e sistemi. Il file segnatura.xml che si scambiano i sistemi di protocollo interoperabili ne è un esempio: il file segnatura.xml contiene i dati della protocollazione in uscita dal mittente. Il file viene intercettato dal sistema di protocollo del destinatario e i dati presenti vengono utilizzati per precompilare la registrazione in entrata. Un altro esempio di documento informato XML è il tracciato dati che viene scambiato tra le applicazioni in una logica di cooperazione applicativa. Il registro di protocollo, è un ulteriore esempio di documento che può essere realizzato in formato XML.

In effetti più che un formato propriamente detto XML (eXtensible Markup Language) è un linguaggio di marcatura: un documento XML è scritto in formato testo con tag (marcatori) anch'essi in formato testo, è quindi leggibile con semplici test editor come il Blocco Note o tramite un qualsiasi browser.

Si riportano brevi note, sicuramente non esaustive, relative a formati per immagini o di tipo audio e video.

**Formati relativi a file contenenti immagini:** TIFF, JPEG e GIF. I formati JPEG e GIF forniscono file di immagini compressi, producendo file di dimensioni ridotte rispetto al TIFF. A causa di questa compressione tuttavia risulta meno robusto del TIFF, cioè: in caso corruzione del file l'immagine visualizzata risulta più degradata nel caso del formato JPEG che nel caso del formato TIFF. Il formato GIF fornisce immagini di bassa qualità perché supporta un numero inferiore di colori rispetto al JPEG.

**Formati relativi a file di tipo audio e video.** L'MP3 è il formato audio, basato sullo standard di compressione MPEG 1; è quello che risponde al criterio di massima diffusione ma non a quello di robustezza. Il WAV, formato audio non compresso, risponde al criterio di diffusione e robustezza ma produce file di notevoli dimensioni.

Per i formati video le estensioni più diffuse sono: AVI, MP4, WMV.

Per le pubbliche Amministrazioni i formati audio e video possono essere considerati come una produzione numericamente marginale o di nicchia, rispetto alla maggiore quantità di documenti di testo, ma possono rivestire una notevole rilevanza nel caso in cui si stabilisca che la verbalizzazione ufficiale di sedute di organi collegiali possano essere rappresentate dalle registrazioni audio o video delle stesse.

Infine si segnala che i formati ZIP, RAR e 7-zip non sono formati di documenti, ma formati di contenitori di documenti, pertanto questi possono contenere documenti (anche strutturati in cartelle) in formati non accettati e questo non risulta immediatamente visibile al ricevente. In particolare se il file compresso e' a sua volta firmato digitalmente, la verifica della firma non consente l'immediata visualizzazione del contenuto informativo.

### Conclusioni

Per gli usuali documenti testuali e in altri casi, quali disegni tecnici o riproduzione in immagine di documenti, attualmente il migliore e più semplice formato accettabile, sia per la gestione corrente che per la successiva conservazione, e' il PDF e PDF/A, che offrono anche le migliori garanzie di corretta visualizzazione al momento della loro apertura tramite sw liberamente disponibili.

Il secondo in particolare, specificamente definito per la conservazione (A indica appunto Archive o "Archiving") presenta ottime caratteristiche in quanto auto-consistente e definito da uno

Se non appare tale dicitura, certamente il formato del documento non è pienamente conforme allo standard PDF/A.

Se appare la dicitura indicata, il formato del documento potrebbe essere conforme allo standard PDF/A, ma la verifica effettiva si ottiene solo tramite prodotti specializzati. Si veda al riguardo il sito [www.pdfa.org](http://www.pdfa.org) alla sezione "Validate PDF/A".

Tra i prodotti elencati si segnalano:

- Adobe Acrobat 9 Professional, che esegue la verifica di conformità del documento alle specifiche dello standard di formato ISO 19005 definite. La funzione di verifica viene attivata selezionando il link presente nella sezione "Informazioni PDF" del documento stesso. Il risultato della verifica viene fornito come "stato". Informazioni di dettaglio sul documento sono ottenibili quale risultato dell'analisi della funzione "Preflight".
- PdfaPilot di [www.callassoftware.com](http://www.callassoftware.com), la cui versione server, per piattaforma Unix, è attualmente utilizzata dal servizio camerale di validazione formato PDF/A, messo a disposizione degli utenti Telemaco per la verifica puntuale (e non massiva) dei documenti in formato PDF/A1-b

Lo standard PDF/A è suddiviso in due parti. Il PDF/A-1, l'unico ad oggi approvato, è suddiviso a sua volta in due livelli:

- PDF/A-1a= massimo richiesto dallo standard
- PDF/A-1b= minimo richiesto dallo standard (incluso nel PDF/A-1a).

E' infine da sottolineare che la conversione in formato PDF di un documento, redatto e gestito fino al momento della sua stesura finale tramite un software specifico, quali quelli della suite di Office, o di OpenOffice, è ormai accessibile a tutti, in modalità semplice e gratuita, tramite software free (es: PDF Creator) o tramite i tool di conversione presenti nei menù sia di MS-Office che di Open Office

Al loro interno le Amministrazioni possono quindi facilmente organizzarsi affinché la produzione dei documenti e la gestione documentale (interna e in uscita) sia basata essenzialmente sul formato PDF o meglio PDF/A.

Formati quali DOC, ODT e RTF possono essere ritenuti accettabili soprattutto perché rispondono al criterio della diffusione (soprattutto nel caso del DOC) e della apertura, ma dovrebbero essere limitati all'uso per scambio di modelli modificabili o di bozze di documenti da aggiornare e non per le versioni finali e stabili dei documenti stessi, magari firmati digitalmente. Inoltre ODT ha il vantaggio della gratuità del software che lo produce.

Formati quali quelli dei fogli di calcolo non dovrebbero essere accettati e soprattutto mai firmati elettronicamente, in quanto quasi certamente contengono macroistruzioni, codici eseguibili e possibili riferimenti esterni che possono alterare la rappresentazione dei dati contenuti; per questi formati, nel caso sia necessario trasmettere ad altri o ricevere da altri dati elaborati con fogli di calcolo dovrebbe essere chiesta/suggerita una preventiva conversione in PDF, per garantire la stabilità dei dati stessi e la possibilità di acquisirli come documento stabile e conservabile nel sistema di gestione documentale

Nei caso di documenti strutturati o necessari per sviluppare funzioni di interoperabilità tra sistemi o per il trasferimento di dati sicuramente il formato attualmente preferibile è XML.

Il formato XML è un formato pienamente accettabile.

Il formato immagine che offre migliori garanzie ai fini della conservazione è il TIFF.

Formati audio e video potranno essere accettati se rispondenti alle caratteristiche elencate all'inizio del paragrafo, i formati più diffusi di questa tipologia sono: MP3, AVI, WMV, MPEG-4.

I formati audio e video tuttavia sono in continua evoluzione; soprattutto i formati video

