



Comune di Castel Baronia
Provincia di Avellino



Manuale di Gestione Documentale (art. 5 DPCM 3/12/2013) Allegato 9 - Istruzioni operative Formati elettronici dei documenti

Cod. **MANGEDOC**

Rev. **1.0**

Data: 20-07-2016

Sommario: Questo allegato descrive i criteri che indirizzano l'Ente nella scelta dei formati da adottare per la formazione, gestione e conservazione dei documenti elettronici.



REVISIONI

Rev.	Data	Redattore/i	Descrizione
1.0	20-07-2016	Maria Pia Papa	Prima stesura



INDICE

1	SCOPO	4
2	OBIETTIVO	4
3	CRITERI DI SCELTA DEI FORMATI PER LA GESTIONE	4
3.1	Il formato del documento informatico.....	4
3.2	Caratteristiche da considerare nella scelta del formato.....	5
3.2.1	Apertura	5
3.2.2	Sicurezza	5
3.2.3	Portabilità.....	5
3.2.4	Funzionalità.....	5
3.2.5	Supporto allo sviluppo	5
3.2.6	Diffusione.....	5
3.2.7	Ulteriori caratteristiche da considerare.....	6
4	CRITERI DI SCELTA DEI FORMATI PER LA CONSERVAZIONE.....	6
4.1	Non proprietà	7
4.2	Apertura.....	7
4.3	Standardizzazione	7
4.4	Trasparenza	8
5	TIPOLOGIE DI FORMATO	9
6	FORMATI ADOTTATI	11



1 SCOPO

L'evolversi delle tecnologie e la crescente disponibilità e complessità dell'informazione digitale ha indotto la necessità di gestire sempre maggiori forme di informazione digitale (testo, immagini, filmati, ecc.) e di disporre di funzionalità più specializzate per renderne più facile la creazione, la modifica e la manipolazione.

Questo fenomeno ha portato all'aumento del numero dei formati disponibili e dei corrispondenti programmi necessari a gestirli nonché delle piattaforme su cui questi operano.

Il presente documento ha lo scopo di fornire dei criteri generali per indirizzare l'Ente nella corretta scelta dei formati da adottare per la formazione, gestione e conservazione dei documenti elettronici, in coerenza con le regole tecniche emanate dalla Presidenza del Consiglio dei Ministri il 13/11/2014.

2 OBIETTIVO

Obiettivo del documento è di permettere all'Ente di adottare formati di documenti elettronici tali da garantire la leggibilità e la reperibilità del documento informatico durante il suo ciclo di vita attivo e successivamente nel tempo.

I formati adottati possono essere aggiornati periodicamente sulla base dell'evoluzione tecnologica, pertanto questo documento è sottoposto a revisione periodica.

3 CRITERI DI SCELTA DEI FORMATI PER LA GESTIONE

3.1 Il formato del documento informatico

La leggibilità di un documento informatico dipende dalla possibilità e dalla capacità di interpretare ed elaborare correttamente i dati binari che costituiscono il documento, secondo le regole stabilite dal formato con cui esso è stato rappresentato.

Il formato di un file è la convenzione usata per interpretare, leggere e modificare il file.

L'associazione del documento informatico al suo formato può avvenire, attraverso varie modalità, tra cui le più impiegate sono:

1. l'estensione del file: una serie di lettere, unita al nome del file attraverso un punto, ad esempio [nome del file].docx identifica un formato testo di proprietà della Microsoft;
2. i metadati espliciti: l'indicazione "application/msword" inserita nei tipi MIME che indica un file testo realizzato con l'applicazione Word della Microsoft;
3. il magic number: i primi byte presenti nella sequenza binaria del file, ad esempio la sequenza 0xffd8 identifica i file immagine di tipo .jpeg.

Una prima sommaria, e non esaustiva, elencazione dei più diffusi formati, secondo il loro specifico utilizzo, può essere la seguente:

- Testi/documenti (DOC, HTML, PDF,...)
- Calcolo (XLS, ...)
- Immagini (GIF, JPG, BMP, TIF, EPS, SVG, ...)
- Suoni (MP3, WAV, ...)
- Video (MPG, MPEG, AVI, WMV,...)
- Eseguibili (EXE, ...)



- Archiviazione e Compressione (ZIP, RAR, ...)
- Formati email (SMTP/MIME, ...)

Ai fini della formazione, gestione e conservazione, questo Ente ritiene di dover scegliere formati tali da garantire la leggibilità e la reperibilità del documento informatico durante il suo ciclo di vita attivo e successivamente nel tempo.

3.2 Caratteristiche da considerare nella scelta del formato

Come richiesto dalla letteratura e dalla normativa cogente, le principali caratteristiche che devono orientare la scelta da parte dell'Ente sono:

1. apertura
2. sicurezza
3. portabilità
4. funzionalità
5. supporto allo sviluppo
6. diffusione

3.2.1 Apertura

Un formato si dice "aperto" quando è conforme a specifiche pubbliche, cioè disponibili a chiunque abbia interesse ad utilizzare quel formato. La disponibilità delle specifiche del formato rende sempre possibile la decodifica dei documenti rappresentati in conformità con dette specifiche, anche in assenza di prodotti che effettuino tale operazione automaticamente.

3.2.2 Sicurezza

La sicurezza di un formato dipende da due elementi: il grado di modificabilità del contenuto del file e la capacità di essere immune dall'inserimento di codice maligno.

3.2.3 Portabilità

Per portabilità si intende la facilità con cui i formati possano essere usati su piattaforme diverse, sia dal punto di vista dell'hardware che del software, inteso come sistema operativo. Di fatto è indotta dall'impiego fedele di standard documentati e accessibili.

3.2.4 Funzionalità

Per funzionalità si intende la possibilità da parte di un formato di essere gestito da prodotti informatici, che prevedono una varietà di funzioni messe a disposizione dell'utente per la formazione e gestione del documento informatico.

3.2.5 Supporto allo sviluppo

E' la modalità con cui si mettono a disposizione le risorse necessarie alla manutenzione e sviluppo del formato e i prodotti informatici che lo gestiscono (organismi preposti alla definizione di specifiche tecniche e standard, società, comunità di sviluppatori, ecc.).

3.2.6 Diffusione

La diffusione è l'estensione dell'impiego di uno specifico formato per la formazione e la gestione dei documenti informatici.



Questo elemento influisce sulla probabilità che esso venga supportato nel tempo, attraverso la disponibilità di più prodotti informatici idonei alla sua gestione e visualizzazione.

3.2.7 Ulteriori caratteristiche da considerare

Inoltre, nella scelta dei prodotti altre caratteristiche importanti sono la capacità di occupare il minor spazio possibile in fase di memorizzazione (a questo proposito vanno valutati, in funzione delle esigenze dell'utente, gli eventuali livelli di compressione utilizzabili) e la possibilità di gestire il maggior numero possibile di metadati, compresi i riferimenti a chi ha eseguito modifiche o aggiunte.

4 CRITERI DI SCELTA DEI FORMATI PER LA CONSERVAZIONE

Una particolare attenzione va rivolta alla scelta dei formati in riferimento alla 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 oggi incontra sicuramente serie difficoltà ad accedervi.

E' molto importante stabilire quali sono i criteri oggettivi per la scelta dei formati dei documenti elettronici che assicurino la loro leggibilità a distanza di venti, cinquanta o più anni, quindi avere ben chiari i requisiti che devono essere presi in considerazione nella scelta di un formato compatibile con un processo di conservazione digitale.

L'individuazione dei requisiti desiderabili per i formati elettronici è stata oggetto di un'intensa attività di studio e ricerca da parte di numerosi enti ed organizzazioni nazionali ed internazionali. A tale proposito si citano un interessante testo del Prof. Stefano Pigliapoco (Università di Macerata), *"La memoria digitale delle amministrazioni pubbliche. Requisiti, metodi e sistemi per la produzione, archiviazione e conservazione dei documenti informatici"*, pubblicato nel 2005 dall' editore Maggioli ed una dispensa più recente (febbraio 2010) di Stefano Allegrezza, dal titolo *"Requisiti e standard dei formati elettronici per la produzione di documenti informatici"*, da cui sono tratti alcuni concetti espressi qui di seguito.

I requisiti dei formati si distinguono in requisiti generali, applicabili a tutte le tipologie di formati, e requisiti specifici, relativi ad una particolare categoria di formati, quali, ad esempio, i documenti di testo, le immagini, i contenuti audio, etc..

I requisiti generali, a loro volta, possono essere:

- **requisiti generali di primo livello**, che vanno presi in considerazione in prima battuta per operare una selezione iniziale
 - **non proprietà**
 - **apertura**
 - **standardizzazione**
 - **trasparenza**
- **requisiti generali di secondo livello**, non meno importanti dei primi, che vanno presi in esame successivamente dopo aver selezionato i formati sulla base dei requisiti di primo livello, tra i quali citiamo robustezza, stabilità, auto-contenimento, auto-documentazione, indipendenza dal dispositivo, assenza di meccanismi tecnici di protezione, assenza di limitazioni sull'utilizzo, accessibilità, non modificabilità, sicurezza, efficienza.



Si riportano di seguito alcuni chiarimenti relativi ai requisiti generali di primo livello, rimandando alle sopra citate pubblicazioni per una trattazione più completa di tutti i requisiti.

4.1 Non proprietà

Un formato si dice proprietario quando è stato creato da una organizzazione privata (ad esempio, una software house), che ne detiene i diritti di proprietà intellettuale; di conseguenza le sue specifiche vengono gestite esclusivamente da tale organizzazione.

Un formato si dice, invece, non proprietario (o libero) quando la gestione delle sue specifiche non è prerogativa di un'organizzazione privata ma è affidata ad una comunità di sviluppatori che cooperano per la gestione condivisa delle stesse, o ad un organismo di standardizzazione.

Ad esempio, sono proprietari (di proprietà Microsoft) i ben noti formati DOC, XLS e PPT (prodotti, rispettivamente, con Microsoft Word, Microsoft Excel e Microsoft PowerPoint), mentre è non proprietario il formato ODF (prodotto con la suite di office automation OpenOffice.org).

Ai fini della conservazione digitale è preferibile utilizzare formati non proprietari in quanto non sono legati all'esistenza di una specifica azienda che ne detiene la proprietà e che potrebbe, in qualsiasi momento, modificarne le specifiche, renderle inaccessibili, o imporre restrizioni sul loro utilizzo.

4.2 Apertura

La definizione di apertura è quella già riportata nel par. 3.2.1.

Il fatto che un formato sia aperto è indipendente dal fatto che sia proprietario o meno.

Ad esempio, il formato DOC della Microsoft (così come l'XLS e il PPT), oltre ad essere proprietario, è stato, per diversi anni, anche chiuso perché le sue specifiche non erano mai state rese note; invece il DOCX è aperto, dal momento che Microsoft ne ha pubblicato fin dall'inizio le specifiche complete.

Un altro esempio è il formato PDF che, pur essendo stato per molti anni proprietario, è sempre rimasto aperto in quanto le sue specifiche erano liberamente accessibili.

Il requisito dell'apertura è di importanza fondamentale in quanto solo se le specifiche sono note è possibile la realizzazione di software in grado di interpretare correttamente la sequenza di bit che costituisce l'oggetto digitale.

Tuttavia, il requisito dell'apertura non è sufficiente se non è affiancato dal requisito della completa documentazione, ovvero di una descrizione completa ed esaustiva del formato. Un esempio di formato pienamente documentato è il PDF/A, le cui specifiche sono state riconosciute come standard ISO 19005-1:2005.

4.3 Standardizzazione

Un formato è standard quando le sue specifiche sono state definite o approvate da un organismo di standardizzazione ufficiale (quali l'ISO, l'ANSI, l'ECMA, il W3C, etc.) (standard de jure) oppure quando le sue specifiche sono diventate uno standard grazie alla sua ampia diffusione (in questo caso si parla di standard de facto).

I formati che hanno ottenuto un riconoscimento come standard da parte di un organismo di standardizzazione sono meno soggetti ad obsolescenza.



Ai fini della conservazione, è, quindi, importante scegliere formati che siano standard; gli standard de jure sono, inoltre, da preferire agli standard de facto, dal momento che solo il processo ufficiale di standardizzazione garantisce che non vi siano interessi di parte nella definizione delle specifiche di un formato e nella sua implementazione.

Strettamente connesso al requisito della standardizzazione è quello dell'**ampia adozione**, che fa riferimento al grado di utilizzo di un formato. Questo requisito è molto importante in quanto l'ampia adozione costituisce uno dei principali "deterrenti" contro i rischi legati all'obsolescenza tecnologica. È evidente che, se un formato è ampiamente adottato, esso sarà meno soggetto ad essere abbandonato dalle aziende produttrici di software, le quali saranno in grado di sviluppare strumenti più semplici per la creazione, la fruizione, la migrazione e l'emulazione di file codificati secondo quel formato, senza necessità di specifici investimenti da parte delle istituzioni archivistiche¹.

4.4 Trasparenza

L'ultimo requisito generale di primo livello è quello della trasparenza, che tiene conto del grado di semplicità con cui è possibile ottenere la fruizione di un file⁶⁰. Un formato è trasparente se è possibile la fruizione dei contenuti digitali codificati secondo quel formato utilizzando semplici strumenti di base (ad esempio, mediante un editor di testo⁶¹ nel caso dei documenti di tipo testuale). In realtà il requisito della trasparenza non è assoluto, nel senso che non esistono formati elettronici completamente trasparenti, essendo sempre necessaria l'intermediazione di un sistema informatico per la fruizione di un contenuto digitale; tuttavia esistono formati più trasparenti ed altri meno trasparenti.

Per fare un esempio, mettendo a confronto tre formati, il TXT, l'RTF e il DOC, si può senz'altro affermare che:

- il TXT è in assoluto quello più trasparente in quanto permette di inserire il testo sotto forma di caratteri ASCII, ma non prevede alcuna formattazione;
- l'RTF è meno trasparente del TXT, ma più trasparente del DOC, essendo possibile, in linea di principio e con un po' di buona volontà, interpretare senza l'ausilio di un computer anche i tag utilizzati dal formato;
- il formato DOC risulta il meno trasparente tra tutti.

Un altro aspetto da tenere presente è che la compressione può ridurre la trasparenza. Nel caso sia necessario utilizzarla, è opportuno scegliere algoritmi di compressione che siano aperti, non proprietari, ampiamente documentati, non soggetti ad alcuna licenza e possibilmente standard.

Grazie alla loro semplicità, la maggior parte dei programmi riesce ad interpretare facilmente i formati trasparenti, ed è prevedibile che ciò sarà possibile anche in futuro nel caso in cui il formato dovesse diventare obsoleto o dovessero andare smarrite le sue specifiche.

L'importanza del requisito della trasparenza risulta evidente anche dai casi, ormai numerosi, di perdita di dati digitali che si sono verificati nel corso dell'ultimo cinquantennio.

¹ Cfr. il sito della Library of Congress (US) <<http://www.digitalpreservation.gov/formats/index.shtml>>: «If a format is widely adopted, it is less likely to become obsolete rapidly, and tools for migration and emulation are more likely to emerge from industry without specific investment by archival institutions».

Si veda anche Frequently Asked Questions (FAQs) ISO 19005-1:2005 PDF/A-1, luglio 2006, disponibile sul sito dell'AIIM all'indirizzo <http://www.aiim.org/documents/standards/19005-1_FAQ.pdf>: «Adoption - widespread use may be the best deterrent against preservation risk».



5 TIPOLOGIE DI FORMATO

L'allegato 2 al DPCM 13/11/2014 fornisce una descrizione dei formati che, per le loro caratteristiche, possono essere considerati ai fini della conservazione. L'Ente effettua la propria scelta dei formati da adottare orientandosi sulla base dei requisiti fissati nei capitoli precedenti e delle tipologie di formato indicate nella tabella seguente:

Formato	Caratteristiche
PDF	<p>Il PDF (Portable Document Format) è un formato creato da Adobe nel 1993 che attualmente si basa sullo standard ISO 32000. E' stato concepito per rappresentare documenti complessi in modo indipendente dalle caratteristiche dell'ambiente di elaborazione del documento. Nell'attuale versione gestisce varie tipologie di informazioni quali: testo formattato, immagini, grafica vettoriale 2D e 3D, filmati.</p> <p>Un documento PDF può essere firmato digitalmente in modalità nativa attraverso il formato ETSI PAdES.</p> <p>Il formato è stato ampliato in una serie di sotto-formati tra cui il PDF/A.</p> <p>Sviluppato da Adobe Systems (http://www.adobe.com/) - Estensione .pdf</p> <p>Tipo MIME application/pdf - Formato aperto - Specifiche tecniche Pubbliche</p>
PDF/A	<p>Il PDF/A è stato sviluppato con l'obiettivo specifico di rendere possibile la conservazione documentale a lungo termine su supporti digitali.</p> <p>Tra le caratteristiche di questa tipologia di file abbiamo:</p> <ul style="list-style-type: none">- assenza di collegamenti esterni,- assenza di codici eseguibili, quali javascript, ecc.,- assenza di contenuti crittografati. <p>Queste caratteristiche rendono il file indipendente da codici e collegamenti esterni che ne possono alterare l'integrità e l'uniformità nel lungo periodo.</p> <p>Le più diffuse suite d'ufficio permettono di salvare direttamente i file nel formato PDF/A.</p> <p>Sono disponibili prodotti per la verifica della conformità di un documento PDF al formato PDF/A.</p>
TIFF	<p>Sviluppato da Aldus Corporation in seguito acquistata da Adobe.</p> <p>Estensione .tif - Tipo MIME: image/tiff</p> <p>Formato non aperto - Specifiche tecniche Pubbliche</p> <p>Di questo formato immagine raster vi sono parecchie versioni, alcune delle quali proprietarie, che ai fini della conservazione nel lungo periodo sarebbe bene evitare.</p> <p>Questo è un formato utilizzato per la conversione in digitale di documenti cartacei. Il suo impiego va valutato attentamente in funzione del tipo di documento da conservare in considerazione dei livelli di compressione e relativa perdita dei dati.</p>
JPG	<p>Sviluppato da Joint Photographic Experts Group - Estensioni .jpg, .jpeg</p> <p>Tipo MIME: image/jpeg - Formato aperto - Specifiche tecniche Pubbliche</p>



	<p>Il formato JPEG può comportare una perdita di qualità dell'immagine originale.</p> <p>Anche in questo caso, come nel caso dei TIFF, avendo una grossa diffusione, può essere preso in considerazione, ma il suo impiego, correlato ad un opportuno livello di compressione va valutato attentamente in funzione del tipo di documento da conservare.</p> <p>JPG è il formato più utilizzato per la memorizzazione di fotografie ed è quello più comune su World Wide Web.</p> <p>Lo stesso gruppo che ha ideato il JPG ha prodotto il JPEG 2000 con estensione .jp2 (ISO/IEC 15444-1) che può utilizzare la compressione senza perdita di informazione. Il formato JPEG 2000 consente, inoltre, di associare metadati ad un'immagine. Nonostante queste caratteristiche la sua diffusione è tutt'oggi relativa.</p>
Office Open XML (OOXML)	<p>Sviluppato da Microsoft (http://www.microsoft.com - http://www.microsoft.it)</p> <p>Estensioni principali .docx, .xlsx, .pptx - Formato aperto - Derivato da XML</p> <p>Specifiche tecniche pubblicate da Microsoft dal 2007</p> <p>Possibile presenza codice maligno</p> <p>Comunemente abbreviato in OOXML, è un formato di file, sviluppato da Microsoft, basato sul linguaggio XML per la creazione di documenti di testo, fogli di calcolo, presentazioni, grafici e database.</p> <p>Open XML è adottato dalla versione 2007 della suite Office di Microsoft.</p> <p>Il formato Office Open XML dispone di alcune caratteristiche che lo rendono adatto alla conservazione nel lungo periodo, tra queste l'embedding dei font, la presenza di indicazioni di presentazione del documento, la possibilità di applicare al documento la firma digitale XML.</p> <p>I metadati associabili ad un documento che adotta tale formato sono previsti dallo standard ISO 29500:2008.</p>
Open Document Format (ODF)	<p>Sviluppato da OASIS (http://www.oasis-open.org/) Oracle America (già Sun Microsystems) (http://www.oracle.com/it/index.html)</p> <p>Estensioni .ods, .odp, .odg, .odb - Formato aperto - Derivato da XML</p> <p>Specifiche tecniche pubblicate da OASIS dal 2005.</p> <p>ODF (Open Document Format, spesso referenziato con il termine OpenDocument) è uno standard aperto, basato sul linguaggio XML, sviluppato dal consorzio OASIS per la memorizzazione di documenti corrispondenti a testo, fogli elettronici, grafici e presentazioni.</p> <p>Secondo questo formato, un documento è descritto da più strutture XML, relative a contenuto, stili, metadati ed informazioni per l'applicazione.</p> <p>Lo standard ISO/IEC IS 26300:2006 è ampiamente usato come standard documentale nativo, oltre che da OpenOffice.org, da una ampia serie di altri prodotti disponibili sulle principali piattaforme: Windows, Linux, Mac.</p> <p>È stato adottato come standard di riferimento da moltissime organizzazioni governative e da diversi governi ed ha una "penetrazione" di mercato che cresce giorno per giorno.</p>
XML	<p>Sviluppato da W3C - Estensione .xml - Formato aperto</p>



	<p>Specifiche tecniche pubblicate da W3C (http://www.w3.org/XML/)</p> <p>Extensible Markup Language (XML) è un formato di testo flessibile derivato da SGML (ISO8879).</p> <p>Su XML si basano numerosi linguaggi standard utilizzati nei più diversi ambiti applicativi. Ad esempio:</p> <ul style="list-style-type: none">• SVG usato nella descrizione di immagini vettoriali• XBRL usato nella comunicazione di dati finanziari• ebXML usato nel commercio elettronico• SOAP utilizzato nello scambio dei messaggi tra Web Service
TXT	<p>Oltre a XML, per quanto concerne i formati non binari “in chiaro”, è universalmente utilizzato il formato TXT.</p> <p>Ai fini della conservazione nell’uso di tale formato, è importante specificare la codifica del carattere (Character Encoding) adottata.</p>
Formati messaggi di posta elettronica	<p>Ai fini della conservazione, per preservare l’autenticità dei messaggi di posta elettronica, lo standard a cui fare riferimento è RFC 2822/MIME.</p> <p>Per quanto concerne il formato degli allegati al messaggio, valgono le indicazioni di cui ai precedenti paragrafi.</p>

6 FORMATI ADOTTATI

Nel panorama piuttosto variegato di formati esistenti allo stato attuale, non essendo possibile fare previsioni su quale formato riuscirà ad imporsi come standard de jure e/o de facto, nell’intento di fondare su basi solide il processo di conservazione digitale, questo Ente fa ricadere la propria scelta su formati che soddisfino al massimo grado anzitutto i requisiti generali di primo livello (non proprietà, apertura, standardizzazione e trasparenza) e poi il maggior numero possibile di requisiti generali di secondo livello.

In quest’ottica, quindi, stabilisce quanto segue:

Trattamento	Formati adottati
Formazione e trattamento documenti informatici	PDF, PDF/A, TIFF, JPG, OOXML, ODF, XML, TXT
Archiviazione nel repository documentale (Halley Document Server)	Prima dell’archiviazione, i file vengono convertiti in formato PDF, PDF/A
Firma digitale di documenti informatici	Prima della sottoscrizione, i file vengono convertiti in formato PDF/A o PDF.
Produzione di copie informatiche di documenti informatici	PDF, PDF/A
Acquisizione in formato digitale di documenti analogici, mediante scansione	PDF
Acquisizione dall’esterno di documenti non sottoscritti con firma digitale	Formati indicate per la conservazione, ai sensi dell’allegato 2 del DPCM 13/11/2014: PDF, PDF/A, TIFF, JPG, OOXML, ODF, XML, TXT
Messaggi di posta elettronica	RFC 2822/MIME



Allegati ai messaggi di posta elettronica	Sono ammessi i formati elettronici già stabiliti ai punti precedenti
---	--