

MANUALE DI GESTIONE DEL PROTOCOLLO INFORMATICO, DEI FLUSSI DOCUMENTALI E DEGLI ARCHIVI

ALLEGATO 5 - Formati idonei per la formazione e conservazione dei documenti informatici

1. Introduzione

Il presente documento fornisce indicazioni iniziali sui formati dei documenti informatici che per le loro caratteristiche sono, al momento attuale, da ritenersi coerenti con le regole tecniche del documento informatico, del sistema di conservazione e del protocollo informatico.

I formati descritti sono stati scelti tra quelli che possono maggiormente garantire i principi dell'interoperabilità tra i sistemi di conservazione e in base alla normativa vigente riguardante specifiche tipologie documentali.

Il presente allegato, per la natura stessa dell'argomento trattato, viene periodicamente aggiornato sulla base dell'evoluzione tecnologica e dell'obsolescenza dei formati e allineato con la pubblicazione dell'Agenzia per l'Italia digitale.

2. I formati

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.

2.1 Le tipologie di formato

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

Per esigenze lavorative gestionali possono essere trattati documenti in formati diversi da quelli indicati al paragrafo 2.3, come ad esempio in formato WORD (.doc), EXCEL (.xls), POWER POINT (.ppt), purché accompagnati dalla versione dello stesso documento in uno dei formati accettati, preferibilmente PDF/A.

2.2 Caratteristiche generali dei formati

L'informazione digitale è facilmente memorizzata, altrettanto facilmente accedere e riutilizzarla, modificarla e manipolarla, in altre parole, elaborarla ed ottenere nuova informazione.

In particolare devono soddisfare quanto previsto da AGID

- apertura
- sicurezza

- portabilità
- funzionalità
- supporto allo sviluppo
- diffusione

2.2.1 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.

Questa condizione si verifica sia quando il formato è documentato e pubblicato da un produttore o da un consorzio al fine di promuoverne l'adozione, sia quando il documento è conforme a formati definiti da organismi di standardizzazione riconosciuti.

Nelle indicazioni di questo documento si è inteso privilegiare i formati già approvati dagli Organismi di standardizzazione internazionali quali ISO e ETSI.

2.2.2 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.

2.2.3 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.

2.2.4 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.

2.2.5 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.).

2.2.6 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.

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.

2.3 Formati idonei per la conservazione

La scelta dei formati idonei alla conservazione oltre al soddisfacimento delle caratteristiche suddette deve essere strumentale a che il documento assuma le caratteristiche di immutabilità e di staticità previste dalle regole tecniche.

In particolare è necessario tener conto nella scelta dei seguenti elementi:

- non devono poter contenere macroistruzioni o codici eseguibili, ovvero devono essere disponibili gli strumenti capaci di rilevarne la presenza con sufficiente sicurezza;
- devono essere standard e documentati, ovvero le relative specifiche devono essere pubblicamente accessibili, complete ed esaustive;
- devono essere robusti, accurati, ampiamente adottati ed usabili
- devono essere indipendenti dalle piattaforme tecnologiche, in modo da poter visualizzare un documento senza particolari vincoli di natura informatica o il pagamento di royalty;
- devono essere conformi alle disposizioni emanate dalle autorità competenti in materia di archiviazione e conservazione digitale.

Per quanto fin qui considerato, è opportuno privilegiare i formati che siano standard internazionali (de jure e de facto) o, quando necessario, formati proprietari le cui specifiche tecniche siano pubbliche, dandone opportuna evidenza nel manuale di conservazione dei documenti informatici.

Ulteriore elemento di valutazione nella scelta del formato è il tempo di conservazione previsto dalla normativa per le singole tipologie di documenti informatici.

I formati di seguito indicati sono un primo elenco di formati da usare per la conservazione:

• PDF/A (Portable Document Format/Archive) formato sviluppato con l'obiettivo specifico di rendere possibile la conservazione documentale a lungo termine su supporti digitali.
• ODF (Open Document Format) 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. Secondo questo formato, un documento è descritto da più strutture XML, relative a contenuto, stili, metadati ed informazioni per l'applicazione.
• XML (Extensible Markup Language) formato di testo flessibile. E' un linguaggio di markup, ovvero un linguaggio marcatore basato su un meccanismo sintattico che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo.
• OOXML (Office Open XML) 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.
• TXT è un file che contiene solo caratteri di scrittura semplici, che compongono un testo leggibile direttamente dagli utenti senza bisogno di installare programmi appositi.
• RTF (Rich Text Format) è un file ASCII con stringhe di comandi speciali in grado di controllare le informazioni riguardanti la formattazione del testo: il tipo di carattere e il colore, i margini, i bordi del documento, ecc.
• TIFF (Tagged Image File Format) formato immagine di tipo raster.
• DXF (Drawing Interchange Format, o Drawing Exchange Format) , un formato simile al DWG(Autocad) , di cui sono state rilasciate le specifiche tecniche.

- | |
|---|
| <ul style="list-style-type: none">• Shapefile un formato vettoriale proprietario per sistemi informativi geografici (GIS) con la caratteristica di essere interoperabile con con i prodotti che usano i precedenti formati. Il formato è stato sviluppato e regolato da ESRI, allo scopo di accrescere l'interoperabilità fra i sistemi ESRI e altri GIS. Di fatto è diventato uno standard per il dato vettoriale spaziale, e viene usato da una grande varietà di sistemi GIS. |
| <ul style="list-style-type: none">• SVG (Scalable Vector Graphics), un formato aperto, basato su XML, in grado di visualizzare oggetti di grafica vettoriale, non legato ad uno specifico prodotto. |

Come già indicato nelle premesse questo elenco sarà periodicamente aggiornato, sulla base delle nuove tecnologie e dei nuovi standard definiti da AGID.

Qualora detta documentazione debba possedere specifiche valenze giuridiche, tra cui ad esempio l'opponibilità a terzi, deve essere prodotta nei formati indicati nei punti precedenti, o altri che offrano analoghe o maggiori garanzie a motivo dell'evoluzione tecnologica, e firmata digitalmente.

Per i formati indicati deve essere garantita dagli applicativi informatici, la corretta visualizzazione dei contenuti.

Date le caratteristiche richieste per i documenti informatici in tema di inalterabilità e immutabilità, non sono accettati né trasmessi file compressi, come ad esempio i file con estensione “.ZIP” oppure “.RAR”.

I documenti informatici prodotti dall'Amministrazione su formati diversi (ad esempio, in estensione “.doc” di Microsoft Word, “.xls” di Microsoft Excel) prima della loro sottoscrizione con firma elettronica o comunque nel momento che si considerano perfezionati, sono convertiti nel formato PDF/A - o nei formati sopraindicati se maggiormente confacenti al tipo di documento considerato - al fine di garantirne la leggibilità, la non alterabilità durante le fasi di accesso e conservazione e l'immutabilità nel tempo del contenuto e della struttura.