

## FORMATI ELETTRONICI ADOTTATI

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.

I documenti informatici destinati alla conservazione utilizzano i formati previsti nell'Allegato 2 alle Regole Tecniche DPCM 3 dicembre 2013.

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

Definisce la modalità di rappresentazione della sequenza di bit che costituiscono il documento informatico; comunemente è identificato attraverso l'estensione del file.

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

1. l'estensione: 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: es 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 0xffd8 identifica i file immagine di tipo .jpeg

Ai fini della formazione, gestione e conservazione, è necessario scegliere formati che possano garantire la leggibilità e la reperibilità del documento informatico nel suo ciclo di vita.

La scelta tra i formati dipende dalle caratteristiche proprie del formato e dei programmi che lo gestiscono.

Le caratteristiche di cui bisogna tener conto nella scelta sono:

1. Apertura
2. Sicurezza
3. Portabilità
4. Funzionalità
5. Supporto allo sviluppo
6. Diffusione

Per la scelta dei formati idonei alla formazione e gestione dei documenti informatici, sono da tenere in considerazione, oltre alle caratteristiche elencate, i seguenti elementi:

- l'efficienza in termini di occupazione di spazio fisico
- la possibilità di gestire il maggior numero possibile di metadati, compresi i riferimenti a modifiche o aggiunte intervenute sul documento.

Le pubbliche amministrazioni indicano nel manuale di gestione i formati adottati per le diverse tipologie di documenti informatici motivandone le scelte effettuate; specificano altresì i casi eccezionali in cui non è possibile adottare i formati in elenco motivandone le ragioni.

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

E' 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 per la conservazione adottati per le diverse tipologie di documenti informatici devono essere indicati nel manuale di conservazione motivandone le scelte effettuate; sono altresì specificati i casi eccezionali in cui non è possibile adottare i formati in elenco motivandone le ragioni.

Al fine di produrre e gestire documenti informatici che siano conformi alla normativa vigente e compatibili con un processo conservativo a lungo termine, Città di Cantù utilizza i formati elettronici previsti nell'allegato n. 2 al DPCM 3/12/2013 Regole Tecniche per i Sistemi di Conservazione:

- TIFF
- JPG
- Office Open XML (OOXML)
- Open Document Format
- XML
- TXT
- Formati Messaggi di posta elettronica

In particolare:

- I documenti informatici prodotti dall'Ente, per poter essere acquisiti nel Sistema di Gestione Informatica dei Documenti, devono essere prodotti o convertiti in uno dei formati previsti dall'allegato n. 2 del DPCM 3/12/2013; tale conversione avviene preferibilmente nei formati PDF o PDF/A; JPG. Il formato OOXML (es. DOCX o XLSX) è comunque idoneo alla conservazione e quindi, documenti in tali formati possono essere caricati sul Sistema di Gestione Informatica dei Documenti, salvo quanto previsto nei successivi punti;
- i documenti informatici prodotti dall'Ente, al fine di essere sottoscritti con firma digitale, vengono migrati in PDF o PDF/A prima di essere sottoscritti e registrati nel Sistema;
- La produzione di copie informatiche non per immagine di documenti sia informatici che cartacei avviene nei formati PDF o PDF/A;
- per l'acquisizione in formato digitale di documenti, nativamente prodotti in formato cartaceo, mediante l'attività di scansione, l'Ente utilizzerà il formato PDF;
- per l'acquisizione di documenti sottoscritti con firma digitale o altra sottoscrizione elettronica inviati all'AOO da parte di utenti esterni, il Comune richiede l'utilizzo dei formati PDF, PDF/A e JPG (preferibilmente PDF e PDF/A) ad eccezione delle fatture elettroniche, le quali sono in formato XML firmato digitalmente;
- per l'acquisizione sul sistema di documenti che non necessitano di firma digitale o altra sottoscrizione elettronica il Comune accetta, a seconda delle finalità per cui i documenti sono utilizzati, tutti i formati previsti dall'allegato n. 2 del DPCM 3/12/2013, adatti ai fini della conservazione digitale; in tal caso può essere utilizzato anche il formato OOXML (es. DOCX, XLSX, ecc...)
- ai fini della conservazione dei messaggi di posta elettronica lo standard a cui fare riferimento è RFC 2822/MIME, mentre per quanto riguarda il formato degli allegati al suddetto messaggio si utilizzeranno, a seconda della tipologia del documento trattato e delle esigenze, i formati elettronici precedentemente indicati;
- al fine della pubblicazione di documenti sul proprio sito istituzionale l'Ente utilizza preferibilmente i formati PDF o PDF/A;
- per la produzione del registro giornaliero di protocollo da inviare al conservatore accreditato si utilizza il formato PDF.

Nel prospetto di seguito riportato sono indicate alcune delle proprietà più significative di cui godono i formati individuati dal DPCM 3/12/2013, ai fini della gestione e conservazione dei documenti informatici; tale strumento è stato pensato per agevolare il Comune nella scelta dei formati da adottare.

Le sigle riportate nel prospetto sono esplicitate nella successiva legenda:

FORMATO ADOTTATO	PROPRIETÀ RILEVANTI AI FINI DELLA GESTIONE E CONSERVAZIONE DEI DOCUMENTI
PDF	<p><b>S</b> (<i>de jure</i>, ISO 32000) - <b>P</b> - <b>A</b> - <b>CD</b> - <b>AC</b> (se viene fatto uso corretto dei tag) - <b>NM</b> - <b>E</b> - <b>SR</b> - <b>AMP</b> (il PDF presenta alcuni meccanismi di protezione ma è possibile creare file PDF che non ne presentino) - <b>AD</b> - <b>ACT</b> (attenzione al fatto che non necessariamente un file PDF è autocontenuto, può presentare rinvii ad oggetti esterni o dipendere dai font utilizzati) - <b>AA</b> - <b>AL</b> (i brevetti sono concessi a titolo gratuito per tutta la loro durata)</p>
PDF/A	<p><b>S</b> (<i>de jure</i>, ISO 19005) - <b>A</b> - <b>NP</b> - <b>CD</b> - <b>AA</b> - <b>ACT</b> - <b>AD</b> - <b>P</b> - <b>AMP</b> - <b>AL</b> - <b>AC</b> (se viene fatto uso corretto dei tag) - <b>NM</b> - <b>SR</b> - <b>E</b></p>
TIFF	<p>Formato immagine in versione non compressa o compressa senza perdita di informazioni (compressione di tipo <i>lossless</i>), <b>S</b> (<i>de jure</i>, ISO 12639 e ISO 12234, basati su TIFF 6.0) - <b>A</b> - <b>NM</b> (quando utilizzato per l'archiviazione di file con contenuti prevalentemente testuali) - <b>AL</b> - <b>AA</b> - <b>AD</b> - <b>R</b> - <b>SB</b> - <b>SR</b> - <b>P</b>; alcune versioni del formato TIFF sono proprietarie e dunque sarebbe meglio evitarle ai fini della conservazione</p>
JPG	<p>Formato immagine in versione compressa con perdita di informazioni (compressione di tipo <i>lossy</i>, l'utilizzo va valutato attentamente in relazione all'opportuno grado di compressione e al tipo di documento da conservare), <b>S</b> (<i>de jure</i>, ISO/IEC 10918) - <b>A</b> - <b>NM</b> (quando utilizzato per l'archiviazione di file con contenuti prevalentemente testuali) - <b>AA</b> - <b>P</b> - <b>SR</b>; la versione JPEG 2000 (ISO/IEC 15444) consente l'utilizzo di compressione senza perdita di informazioni, tuttavia gode di una diffusione piuttosto relativa.</p>
OOXML	<p><b>S</b> (<i>de jure</i>, ISO/IEC DIS 29500) - <b>A</b> - <b>NP</b> - <b>CD</b> - <b>NB</b> (basato su XML) - <b>P</b> - <b>SB</b> - <b>R</b> - <b>AC</b> - <b>E</b> - <b>AA</b> - <b>T</b> - <b>R</b> - <b>AD</b> - <b>ACT</b> - <b>AMP</b> (I file sono predefinitamente non cifrati, anche se il formato prevede questa possibilità) - <b>AL</b> - <b>NM</b> (il formato è modificabile ma è possibile creare una versione del file non modificabile) - <b>SR</b> (ma prevede la possibilità di attivare le macro) - <b>E</b></p>

<b>ODF</b>	<b>S</b> ( <i>de jure</i> , ISO/IEC 26300) - <b>A</b> - <b>CD</b> - <b>NB</b> (basato su XML) - <b>NP</b> - <b>AA</b> - <b>T</b> - <b>R</b> - <b>ACT</b> - <b>AD</b> - <b>P</b> - <b>AMP</b> (I file sono predefinitamente non cifrati, anche se il formato prevede questa possibilità) - <b>AL</b> - <b>AC</b> - <b>SB</b> - <b>E</b>
<b>XML</b>	<b>S</b> ( <i>de jure</i> , W3C) - <b>A</b> - <b>NB</b> - <b>NP</b> - <b>AC</b> - <b>CD</b> - <b>SB</b> - <b>R</b> - <b>P</b> - <b>AA</b> - <b>AL</b> - <b>E</b> - <b>T</b>
<b>TXT</b>	<b>A</b> - <b>NB</b> - <b>AC</b> - <b>SR</b> - <b>P</b> - <b>R</b> - <b>AA</b> - <b>AL</b> - <b>NP</b> - <b>CD</b> - <b>SB</b> - <b>T</b> - <b>E</b>

Legenda delle caratteristiche dei formati elettronici:

**Accessibile (AC):** un formato si dice accessibile in riferimento alla capacità di un file di risultare fruibile da qualsiasi categoria di utenti in maniera semplice, in altre parole può dirsi accessibile se fruibile non solo da utenti normodotati ma anche da quelli con varie forme di disabilità (ad es. ipovedenti);

**Ampiamente adottato (AA):** un formato si dice ampiamente adottato quando gode di un'adozione estremamente ampia;

**Aperto (A):** un formato si dice aperto quando è descritto da specifiche tecniche pubbliche, liberamente accessibili ed esaustive;

**Assenza di limitazioni sull'utilizzo (AL):** un formato con assenza di limitazioni all'utilizzo non è gravato dall'esistenza di brevetti o di richieste di pagamenti (*royalty*), che possano limitarne l'uso, con pregiudizio della conservazione;

**Assenza di meccanismi tecnici di protezione (AMP):** un formato che non presenti meccanismi tecnici di protezione, consente di replicare il contenuto dei file prodotti per mezzo di esso su nuovi supporti, effettuare migrazioni e normalizzazioni, ai fini della fruibilità nel lungo periodo;

**Autocontenuto (ACT):** un formato si dice autocontenuto quando presenta la capacità di includere tutte le risorse necessarie alla sua rappresentazione;

**Autodocumentato (AD):** un formato si dice autodocumentato quando è in grado di supportare l'inclusione di metadati;

**Completamente documentato (CD):** un formato si dice completamente documentato quando chiunque voglia implementare applicazioni in grado di leggere, produrre o modificare file in quel formato, potrà trovare nelle specifiche tutte le informazioni necessarie;

**Efficiente (E):** un formato si dice efficiente in relazione alle dimensioni ridotte dei file con esso prodotti, a parità di contenuti, rispetto a quelli prodotti con altri formati;

**Non binario (NB):** un formato si dice non binario quando un file formato per mezzo di esso è rappresentato solamente da caratteri in una delle codifiche esistenti (es. ASCII o UNICODE) risultando, pertanto, leggibile con qualsiasi editor di testo;

**Non modificabile (NM):** un formato si dice non modificabile quando il programma che lo gestisce non ne consente la modifica in maniera semplice;

**Non proprietario (NP):** un formato si dice non proprietario quando non è legato ad una particolare azienda che potrebbe modificarne o renderne inaccessibili le specifiche tecniche;

**Portabile (P):** un formato si dice portabile quando il suo utilizzo non è vincolato a specifiche piattaforme sia dal punto di vista dell'hardware che dal punto di vista del software;

**Robusto (R):** un formato si dice robusto quando, in caso di corruzione del file, consente il recupero, totale o parziale, dei suoi contenuti;

**Sicuro (SR):** un formato si dice sicuro quando non può contenere virus o altre forme di codice maligno;

**Stabile (SB):** un formato si dice stabile quando non subisce variazioni nel tempo oppure le nuove versioni dello stesso si succedono con cadenze temporali ragionevoli;

**Standard (S):** un formato si dice standard quando le sue specifiche tecniche sono definite o approvate da un ente di standardizzazione (standard *de jure*) oppure quando il formato gode di un'ampia diffusione presso una comunità (standard *de facto*);

**Trasparente (T):** un formato si dice trasparente quando è possibile, con semplicità, effettuare l'analisi diretta di un file, utilizzando strumenti di base (ad es. semplici editor di testo).

Per quanto riguarda i formati per i file immagine è determinante valutare se si tratta di **formati compressi** o **non compressi**: i formati compressi consentono la riduzione della dimensione del file, permettendo di risparmiare spazio ai fini della memorizzazione dello stesso. La compressione può essere di tipo *lossless* o *lossy*: il primo tipo consente di comprimere il file senza la perdita di informazioni, evitando la riduzione di qualità, il secondo tipo di compressione, invece, comporta la perdita di alcune informazioni, causando una diminuzione della qualità del file al momento della sua rappresentazione, ottenendo però dimensioni più ridotte rispetto ad una compressione di tipo *lossless*.