

FORMATI

Indice

- 1 INTRODUZIONE**
- 2 I FORMATI**
 - 2.1 Identificazione
 - 2.2 Le tipologie di formato
 - 2.3 Formati Immagini
 - 2.3.1 Raster
 - 2.3.2 Vettoriale
 - 2.4 Altri Formati
 - 2.5 Le caratteristiche generali dei formati
- 3 CRITERI DI SCELTA DEI FORMATI**
 - 3.1 Caratteristiche
 - 3.1.1 Apertura
 - 3.1.2 Sicurezza
 - 3.1.3 Portabilità
 - 3.1.4 Funzionalità
 - 3.1.5 Supporto allo sviluppo
 - 3.1.6 Diffusione
- 4 SCELTA**
 - 4.1 Formati e prodotti per la formazione e gestione
 - 4.2 Formati per la conservazione
- 5 I FORMATI INDICATI PER LA CONSERVAZIONE**
 - 5.1 PDF - PDF/A
 - 5.2 TIFF
 - 5.3 JPG
 - 5.4 Office Open XML (OOXML)
 - 5.5 Open Document Format
 - 5.6 XML
 - 5.7 TXT
 - 5.8 Formati Messaggi di posta elettronica

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 documento, per la natura stessa dell'argomento trattato, viene periodicamente aggiornato sulla base dell'evoluzione tecnologica e dell'obsolescenza dei formati e pubblicato online sul sito 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 Identificazione

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: 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

2.2 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.

In particolare, volendo fare una prima sommaria, e non esaustiva, catalogazione dei più diffusi formati, secondo il loro specifico utilizzo possiamo elencare:

- 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, ...)

2.3 Formati Immagini

Per la rappresentazione delle immagini sono disponibili diversi formati, che possono essere distinti secondo la grafica utilizzata: raster o vettoriale.

2.3.1 Raster

Nel caso della grafica raster, l'immagine digitale è formata da un insieme di piccole aree uguali (pixel), ordinate secondo linee e colonne.

I formati più diffusi sono il .tif (usato dai fax), il .jpg, il .bmp.

2.3.2 Vettoriale

La grafica vettoriale è una tecnica utilizzata per descrivere un'immagine mediante un insieme di primitive geometriche che definiscono punti, linee, curve e poligoni ai quali possono essere attribuiti colori e anche sfumature.

I documenti realizzati attraverso la grafica vettoriale sono quelli utilizzati nella stesura degli elaborati tecnici, ad esempio progetti di edifici.

Attualmente i formati maggiormente in uso sono:

- DWG, un formato proprietario per i file di tipo CAD, di cui non sono state rilasciate le specifiche;
- DXF, un formato simile al DWG, di cui sono state rilasciate le specifiche tecniche
- 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.
- SVG, un formato aperto, basato su XML, in grado di visualizzare oggetti di grafica vettoriale, non legato ad uno specifico prodotto.

2.4 Altri Formati

Per determinate tipologie di documenti informatici sono utilizzati specifici formati. In particolare in campo sanitario i formati più usati sono:

- DICOM (immagini che arrivano da strumenti diagnostici) anche se il DICOM non è solo un formato, ma definisce anche protocolli e altro;
- HL7 ed in particolare il CDA2 (Clinical Document Architecture) che contiene la sua stessa descrizione o rappresentazione.

Le specifiche approvate per alcune tipologie di documenti quali le prescrizioni, si trovano al seguente indirizzo:

<http://www.innovazionepa.gov.it/i-dipartimenti/digitalizzazione-e-innovazione-tecnologica/attivita/tse/il-tavolo-permanente-per-la-sanita-elettronica-delle-regioni-e-delle-province-autonome-tse-.aspx>

2.5 Le 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.

Questi formati, e i programmi che li gestiscono, che sono poi quelli che consentono e facilitano l'operatività giorno per giorno sul digitale, vanno valutati in funzione di alcune caratteristiche quali:

La diffusione, ossia il numero di persone ed organizzazioni che li adotta

La portabilità, ancor meglio se essa è indotta dall'impiego fedele di standard documentati e accessibili

Le funzionalità che l'utente ha a disposizione per elaborare l'informazione e collegarla ad altre (ad esempio gestione di link)

La capacità di gestire contemporaneamente un numero congruo (in funzione delle esigenze dell'utente) di formati

La diffusione di visualizzatori che consentono una fruibilità delle informazioni in essi contenute indipendentemente dalla possibilità di rielaborarle.

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.

È facilmente comprensibile come, nella fase di gestione del digitale, l'utente debba avere a disposizione la massima flessibilità possibile in termini di formati e funzionalità disponibili.

Gli unici limiti sono quelli che un'organizzazione impone a se stessa quando per esigenze di interscambio ed interoperabilità, può determinare i formati, e i relativi programmi di gestione, che maggiormente soddisfano le contingenti esigenze operative.

3 Criteri di scelta dei formati

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.

3.1 Caratteristiche

Le caratteristiche di cui bisogna tener conto nella scelta sono:

1. apertura
2. sicurezza
3. portabilità

4. funzionalità
5. supporto allo sviluppo
6. diffusione

3.1.1 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. In quest'ultimo caso tuttavia si confida che quest'ultimi garantiscono l'adeguatezza e la completezza delle specifiche stesse.

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

3.1.2 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

3.1.3 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.

3.1.4 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.

3.1.5 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.).

3.1.6 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.

4 Scelta

4.1 Formati e prodotti per la formazione e gestione

Per la scelta dei formati idonei alla formazione e gestione dei documenti informatici, sono da tenere in considerazione le caratteristiche indicate nei paragrafi precedenti.

Ulteriori elementi da valutare sono l'efficienza in termini di occupazione di spazio fisico e 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.**

4.2 Formati 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.

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 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.

5 I formati indicati per la conservazione

I formati di seguito indicati sono un primo elenco di formati che possono essere usati per la conservazione.

Come già indicato nelle premesse questo elenco sarà periodicamente aggiornato.

5.1 PDF - PDF/A

Il PDF (Portable Document Format) è un formato creato da Adobe nel 1993 che attualmente si basa sullo standard ISO 32000. È 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.

Un documento PDF può essere firmato digitalmente in modalità nativa attraverso il formato ETSI PAdES.

Il formato è stato ampliato in una serie di sotto-formati tra cui il PDF/A.

Sviluppato da	Adobe Systems http://www.adobe.com/
Estensione	.pdf
Tipo MIME	application/pdf
Formato aperto	Sì
Specifiche tecniche	Pubbliche
Standard	ISO 32000-1 (PDF) ISO 19005-1:2005 (vers. PDF 1.4) ISO 19005-2:2011 (vers. PDF 1.7)
Ultima versione	1.7
Collegamento utile	http://www.pdfa.org/doku.php

Il PDF/A è stato sviluppato con l'obiettivo specifico di rendere possibile la conservazione documentale a lungo termine su supporti digitali

Tra le caratteristiche di questa tipologia di file abbiamo:

- assenza di collegamenti esterni,
- assenza di codici eseguibili quali javascript ecc.,
- assenza di contenuti crittografati.

Queste caratteristiche rendono il file indipendente da codici e collegamenti esterni che ne possono alterare l'integrità e l'uniformità nel lungo periodo.

Le più diffuse suite d'ufficio permettono di salvare direttamente i file nel formato PDF/A.

Sono disponibili prodotti per la verifica della conformità di un documento PDF al formato PDF/A.

5.2 TIFF

Sviluppato da	Aldus Corporation in seguito acquistata da Adobe
Estensioni	.tif
Tipo MIME	image/tiff
Formato aperto	No
Specifiche tecniche	Pubbliche
Ultime versioni	TIFF 6.0 del 1992 TIFF Supplement 2 del 2002
Collegamenti utili	http://partners.adobe.com/public/developer/tiff/index.html

Di questo formato immagine raster, in versione non compressa o compressa senza perdita di informazione. Di questo formato vi sono parecchie versioni, alcune delle quali proprietarie (che ai fini della conservazione nel lungo periodo sarebbe bene evitare). In genere le specifiche sono pubbliche e non soggette ad alcuna forma di limitazione.

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.

Esistono, infine, alcuni formati ISO basati sulla specifica TIFF 6.0 di Adobe (che è quella "ufficiale" del TIFF). Si tratta del formato ISO 12639, altrimenti noto come TIFF/IT, rivolto particolarmente al mondo del publishing e della stampa e dell'ISO 12234, altrimenti detto TIFF/EP, più orientato alla fotografia digitale.

5.3 JPG

Sviluppato da	Joint Photographic Experts Group
Estensioni	.jpg, .jpeg
Tipo MIME	image/jpeg
Formato aperto	Sì
Specifiche tecniche	Pubbliche
Standard	ISO/IEC 10918:1
Ultima versione	2009
Collegamenti utili	http://www.jpeg.org/ www.iso.org

Il formato JPEG può comportare una perdita di qualità dell'immagine originale. 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.

JPG è il formato più utilizzato per la memorizzazione di fotografie ed è quello più comune su World Wide Web.

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.

5.4 Office Open XML (OOXML)

Sviluppato da	Microsoft http://www.microsoft.com http://www.microsoft.it
Estensioni principali	.docx, .xlsx, .pptx
Tipo MIME	
Formato aperto	Sì
Derivato da	XML
Specifiche tecniche	pubblicate da Microsoft dal 2007
Standard	ISO/IEC DIS 29500:2008
Ultima versione	1.1
Possibile presenza codice maligno	Sì
Collegamenti utili	http://msdn.microsoft.com/en-us/library/aa338205.aspx http://standards.iso.org/ittf/PubliclyAvailableStandards www.iso.org

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.

Open XML è adottato dalla versione 2007 della suite Office di Microsoft.

Lo standard prevede, oltre alle indicazioni fondamentali (strict), alcune norme transitorie (transitional) introdotte per ammettere, anche se solo temporaneamente, alcune funzionalità presenti nelle vecchie versioni del formato e la cui rimozione avrebbe potuto danneggiare gli utenti, facendogli perdere funzionalità.

Per quanto riguarda il supporto di Microsoft Office allo standard ISO/IEC 29500:2008:

- MS Office 2007 legge e scrive file conformi a ECMA-376 Edition 1.
- MS Office 2010 legge e scrive file conformi a ISO/IEC 29500:2008 transitional e legge file conformi a ISO/IEC 29500:2008 strict

Documenti conformi ad ISO/IEC 29500:2008 strict sono supportati da diversi prodotti informatici disponibili sul mercato.

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.

I metadati associabili ad un documento che adotta tale formato sono previsti dallo standard ISO 29500:2008.

5.5 Open Document Format

Sviluppato da	OASIS http://www.oasis-open.org/ Oracle America (già Sun Microsystems) http://www.oracle.com/it/index.html
Estensioni	.ods, .odp, .odg, .odb
Tipo MIME	application/vnd.oasis.opendocument.text
Formato aperto	Sì
Derivato da	XML
Specifiche tecniche	pubblicate da OASIS dal 2005
Standard	ISO/IEC 26300:2006 UNI CEI ISO/IEC 26300
Ultima versione	1.0
Collegamenti utili	http://books.evc-cit.info/ http://www.oasis-open.org www.iso.org

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.

Secondo questo formato, un documento è descritto da più strutture XML, relative a contenuto, stili, metadati ed informazioni per l'applicazione.

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.

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.

5.6 XML

Sviluppato da	W3C
Estensioni	.xml
Tipo MIME	application/xml text/xml
Formato aperto	Sì
Specifiche tecniche	pubblicate da W3C http://www.w3.org/XML/
Collegamenti utili	http://www.w3.org/

Extensible Markup Language (XML) è un formato di testo flessibile derivato da SGML (ISO 8879).

Su XML si basano numerosi linguaggi standard utilizzati nei più diversi ambiti applicativi. Ad esempio:

- 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

5.7 TXT

Oltre a XML, per quanto concerne i formati non binari “in chiaro”, è universalmente utilizzato il formato TXT.

Ai fini della conservazione nell'uso di tale formato, è importante specificare la codifica del carattere (Character Encoding) adottata.

5.8 Formati Messaggi di posta elettronica

Ai fini della conservazione, per preservare l'autenticità dei messaggi di posta elettronica, lo standard a cui fare riferimento è RFC 2822/MIME.

Per quanto concerne il formato degli allegati al messaggio, valgono le indicazioni di cui ai precedenti paragrafi.