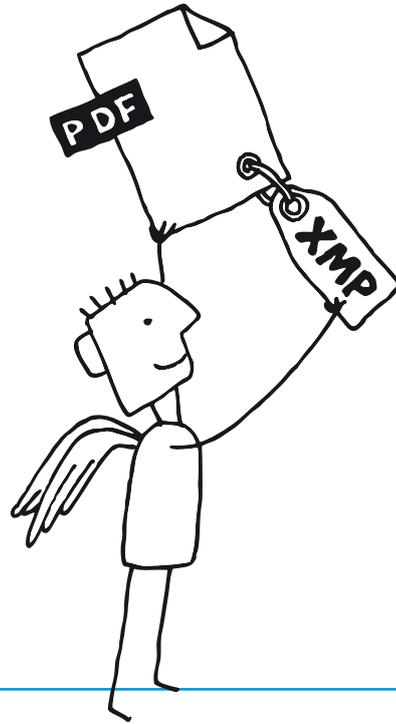


Supporto di metadati XMP nei prodotti PDFlib



 **PDFlib** White paper

L'importanza dei metadati

La parola metadato significa letteralmente »dato su un (altro) dato«. I metadati sono stati descritti come il biglietto da visita di un documento digitale. I metadati comprendono spesso una serie di proprietà, ognuna delle quali assume un determinato significato all'interno di un dato documento. Alcuni esempi di proprietà comuni:

- ▶ L'autore del documento PDF.
- ▶ La data di creazione di un documento PDF o di quando un'immagine è stata scattata con la fotocamera.
- ▶ Il nome dell'utente che ha fatto la foto.
- ▶ Il numero di serie di un documento personalizzato.
- ▶ L'articolo gestito a magazzino (SKU - stockkeeping unit) descritto nel documento.
- ▶ L'anno di produzione del prodotto descritto nel documento.
- ▶ Il numero di riferimento di un documento legale.

Con un numero sempre crescente di pubblicazioni, documentazioni, traduzioni e altri servizi realizzati in modo interamente digitale, i metadati rivestono un ruolo cruciale per la gestione dei documenti digitali durante tutto il loro ciclo di vita.

Extensible Metadata Platform (XMP) di Adobe

Quando Adobe riconobbe la necessità di realizzare un formato comune per i metadati utilizzabile con tutte le applicazioni e i formati, venne implementata la Extensible Metadata Platform (XMP). Si tratta di un formato XML modellato sulla base di RDF (Resource Description Framework), la pietra fondante dell'iniziativa Semantic Web, della W3C. Oltre a rendere le specifiche XMP liberamente disponibili, Adobe mette a disposizione dei programmatori un tool XMP open-source.

I metadati XMP si muovono con il file stesso e possono essere integrati in numerosi file comuni tra i quali PDF, TIFF e JPEG. Le proprietà dei metadati vengono raggruppate nei cosiddetti schemi. Ogni schema viene identificato da un namespace URI univoco e contiene un numero arbitrario di proprietà.

Le specificazioni includono più di dieci schemi predefiniti con centinaia di proprietà per le caratteristiche comuni di documenti e immagini. Lo schema XMP predefinito più diffuso è il Dublin Core, abbreviato *dc*. Esso contiene delle proprietà generali come *Titolo*, *Creato da*, *Oggetto* e *Descrizione*. Oltre agli schemi predefiniti è possibile creare degli schemi personalizzati in modo tale da poter soddisfare le esigenze di un'azienda o di un settore in termini di metadati specifici.

Lo schema XMP per documenti PDF è stato introdotto con Acrobat 5 e PDF 1.4 nel 2001. Lo schema che ha preceduto i file XMP per i PDF era formato da semplici coppie chiave/valore, le cosiddette informazioni sul documento, l'unica formula utilizzata per trasportare i metadati prima dell'introduzione di XMP. Mentre le informazioni sul documento continuano ad essere supportate in Acrobat e nei file PDF, i metadati XMP sono una soluzione molto più potente che permette ai metadati di non andare persi in caso di conversione di formato, ad es. durante il passaggio da TIFF scannerizzati a PDF.

XMP è implementato in tutti i prodotti publishing di Adobe e viene supportato da numerose software house e user group indipendenti. Adobe Bridge, parte di Creative Suite, opera con i metadati XMP in diversi formati di file. I metadati XMP possono essere visualizzati e modificati nel pannello *File Info/ Document Properties* in Acrobat (*File, Properties..., Additional metadata...*), Photoshop, InDesign e altre applicazioni Adobe. Mentre il pannello File Info raggruppa le proprietà dei metadati in base agli schemi XMP predefiniti, i pannelli personalizzati possono essere utilizzati per adattare la visualizzazione dei metadati e dei campi editabili in base alle esigenze delle varie applicazioni.

XMP per le aziende

Sempre più aziende utilizzano XMP per soddisfare le proprie esigenze in termini di metadati. Alcuni esempi:

- ▶ Il consorzio AdsML crea specifiche e processi per lo scambio di informazioni e contenuti pubblicitari.
- ▶ La International Press Telecommunications Council (IPTC) è un Gruppo industriale fondato da agenzie di stampa. Sviluppa standard industriali per lo scambio di notizie. Ha pubblicato l'«IPTC Core» per XMP, uno schema largamente utilizzato per trasferire i metadati di immagini e di altri elementi delle notizie.
- ▶ Lo standard DICOM per lo scambio di immagini mediche supporta l'uso di PDF e specifica uno schema XMP personalizzato per memorizzare i dati dei pazienti, la descrizione dello studio, i dettagli sugli equipaggiamenti e altri metadati.
- ▶ Il Publishing Requirements for Industry Standard Metadata (PRISM) definisce un vocabolario di metadati per elaborare i contenuti di riviste, news, cataloghi, libri e giornali.

XMP obbligatorio per gli standard ISO

Vi sono numerosi standard ISO che specificano i subset PDF per certe tipologie di applicazioni, come il settore grafico, di archiviazione e quello ingegneristico. A parte i vecchi standard di pre stampa PDF/X-1 e X-3, introdotti nel 2001 e nel 2002, tutti gli standard ISO per i PDF comprendono l'uso di metadati XMP (obbligatori con eccezione della norma ISO 32000):

- ▶ PDF/A-1 in ISO 19005-1 (pubblicato nel 2005): «Electronic document file format for long-term preservation – Use of PDF 1.4». PDF/A-1 richiede i metadati XMP per identificare i file conformi e supporta i metadati personalizzati attraverso le estensioni XMP. In futuro, per massimizzare l'utilizzo dei metadati personalizzati, tutte le estensioni dovranno essere embedded in PDF/A-1. PDF/A-1 permette l'utilizzo delle informazioni sul documento, ma richiede, per garantire un workflow basato esclusivamente su XMP, la sincronizzazione tra le comuni informazioni del documento PDF e diverse proprietà XMP predefinite. Lo standard definisce questo «passaggio» tra le informazioni del documento e le proprietà XMP. Il supporto XMP in PDF/A-1 si basa sulle specifiche XMP 2004.
- ▶ PDF/E in ISO 24517-1 (pubblicato nel 2008): «Engineering document format using PDF – Use of PDF 1.6». Il supporto XMP in PDF/E è praticamente identico a quello di PDF/A-1, ma si basa sulla nuova specificazione XMP 2005.
- ▶ PDF/X-4 in ISO 15930-7 (pubblicato nel 2008): «Complete exchange of printing data (PDF/X-4) and partial exchange of printing data with external profile reference (PDF/X-4p) using PDF 1.6». Simile allo standard PDF/A-1, XMP è necessario per la conformità allo standard PDF/X-4. Le informazioni sul documento possono essere usate in PDF/X-4, ma devono essere sincronizzate con le corrispondenti voci XMP. Le estensioni XMP per metadati personalizzati sono consentite. Tuttavia, a differenza di PDF/A-1, questi possono essere utilizzati senza dover integrare una descrizione formale. Il supporto XMP in PDF/X-4 si basa sulle specifiche XMP 2005.
- ▶ PDF/X-2 in ISO 15930-5 (pubblicato nel 2003) e PDF/X-5 in ISO 15930-8 (pubblicato nel 2008): «Partial exchange of printing data using PDF 1.6 (PDF/X-5)». I documenti PDF/X-2 e X-5 rimandano ad altri documenti PDF/X come fonti di contenuto esterno. Il target di questo rimando viene identificato da diverse voci XMP. Questa caratteristica rende il file XMP una componente fondamentale di PDF/X-2 e PDF/X-5.
- ▶ ISO 32000 (pubblicato nel 2008): «Document management – Portable document format – PDF 1.7». ISO 32000 è la versione standardizzata di PDF 1.7. Il contenuto tecnico è identico a PDF 1.7 (il formato di Acrobat 8) con un supporto completo per i metadati XMP.

Il Dublin Core, uno degli schemi per metadati XMP più diffusi, è stato standardizzato nella ISO 15836 (pubblicato nel 2003): «Information and documentation — The Dublin Core metadata element set».

Supporto XMP con la suite di prodotti PDFlib

Nel 2004 è stato introdotto nella gamma di prodotti PDFlib un semplice supporto XMP. Con il supporto PDF/A-1 in PDFlib 7 (pubblicato nel 2006), le funzionalità XMP sono state estese per soddisfare le esigenze di PDF/A-1. In particolare è stata implementata la sincronizzazione automatica delle informazioni del documento con le proprietà XMP (così come specificato in PDF/A-1) e la creazione automatica di numerose proprietà XMP interne necessarie per PDF/A-1. In questo modo gli utenti di PDFlib sono in grado di generare i file XMP per PDF/A-1 senza dover confrontarsi con gli aspetti interni del formato XMP. Gli utenti avanzati possono inserire direttamente in PDFlib tutti i metadati XMP degli schemi predefiniti e integrarli nei documenti PDF creati. Poiché PDFlib è disponibile per tutti i principali sistemi operativi e non richiede prodotti di parti terze, garantisce il supporto XMP su tutte le piattaforme.

Inoltre, PDFlib supporta anche le estensioni XMP come da PDF/A-1. Gli utenti possono integrare le descrizioni delle estensioni per i metadati personalizzati così come richiesto da PDF/A-1. Poiché PDFlib verifica la consistenza interna e la conformità agli standard delle estensioni XMP inserite dagli utenti e delle corrispondenti descrizioni, i file generati saranno sempre conformi allo standard PDF/A-1.

Questa funzionalità ha reso PDFlib la prima soluzione mondiale che supporta le estensioni XMP per PDF/A-1. Grazie alla partecipazione di PDFlib GmbH al PDF/A Competence Center, tutte le attività in ambito di PDF/A vengono coordinate con gli altri soggetti che operano con i software PDF/A, garantendo il massimo livello di conformità agli standard di allineamento alla pratiche del settore. Maggiori dettagli su XMP in PDF/A sono disponibili all'indirizzo www.pdflib.com.

Oltre ai file XMP del documento, i prodotti PDFlib supportano anche gli XMP a livello di oggetto. Ad esempio un XMP di pagina può contenere delle informazioni per l'elaborazione; gli XMP a livello di immagine possono contenere le informazioni sui diritti d'autore richieste per le licenze, ecc.

Inserire XMP nei PDF con PDFlib PLOP e PLOP DS

Oltre alle numerose funzionalità come crittografia, decrittografia, ottimizzazione e firma digitale, PDFlib PLOP e PLOP DS possono inserire i metadati XMP nei documenti PDF già esistenti. Questa funzione risulta particolarmente comoda quando i documenti PDF non contengono tutte le proprietà dei metadati richieste. È particolarmente utile nei workflow con PDF/A poiché in PLOP e PLOP DS il supporto XMP è compatibile con PDF/A. Ad esempio, i file XMP personalizzati con le estensioni possono venire inseriti nei documenti PDF/A anche nei workflow che non supportano le estensioni.

Estrazione di XMP dai PDF con PDFlib pCOS

L'interfaccia pCOS è il metodo utilizzato da PDFlib GmbH per recuperare tutti i tipi di informazioni dai documenti PDF. È disponibile sia come prodotto stand-alone, sia come soluzione integrata in numerosi altri prodotti. pCOS mette a disposizione un semplice metodo di programmazione per estrarre i metadati XMP dai documenti PDF. Grazie alla normalizzazione dei metadati XMP in Unicode, l'utente non dovrà preoccuparsi dei problemi di codifica.

Il recupero di dati XMP funziona indipendentemente dalla compressione, dalla crittografia e dalla struttura degli oggetti PDF. Mentre il meccanismo di impacchettazione di Adobe consente la semplice inclusione e il recupero di pacchetti dati XMP in vari formati, il formato PDF presenta diverse caratteristiche che ne rendono il funzionamento complicato. Ad esempio, i documenti PDF possono contenere numerose sezioni di aggiornamento che causano la presenza di diverse istanze di flusso PDF all'interno dello stesso documento, e questo anche se solo una di queste istanze è rilevante. Una semplice ricerca nel testo del blocco XMP potrebbe produrre un'istanza non corretta; solo i programmi che si attengono attentamente alla struttura del PDF recupereranno sempre il blocco di metadati XMP corretto. Questa è la ragione per cui il tool XMP gratuito di Adobe non è in grado di supportare pienamente il recupero di informazioni XMP dai PDF, supportandolo invece per altri formati come TIFF e JPEG.

Ricerca dei metadati XMP con PDFlib TET PDF IFilter

TET PDF IFilter implementa l'interfaccia IFilter di Microsoft e può essere utilizzato con diversi prodotti di ricerca desktop e enterprise di Microsoft o di altre aziende, come Windows Desktop Search (WDS), Office SharePoint Server (MOSS), Indexing Server e SQL Server. Il supporto XMP in TET PDF IFilter permette di sfruttare i metadati XMP in ambienti dove vengono impiegate le soluzioni di ricerca di Microsoft.

L'implementazione avanzata di TET PDF IFilter supporta le proprietà di sistema di Windows per i metadati. Oltre al contenuto delle pagine, indicizza i metadati XMP e le informazioni sul documento standard e personalizzate. L'indicizzazione dei metadati può essere configurata in molti livelli:

- ▶ Le informazioni sul documento e le proprietà XMP comuni vengono mappate con le proprietà standard di Windows, ad es. *Titolo, Oggetto, Autore*.
- ▶ TET PDF IFilter aggiunge utili pseudo-proprietà, come dimensione della pagina, livello di conformità PDF/A, elenco dei font.

- ▶ Tutte le più importanti proprietà XMP possono essere ricercate, come *dc:rights*, *xmpRights:UsageTerms*, *xmp:CreatorTool*.
- ▶ Le proprietà XMP personalizzate (definite dall'utente) possono essere ricercate, ad es. le voci di classificazione specifiche dell'azienda.
- ▶ Oltre ai metadati dei documenti, i XMP attaccati alle immagini, ad es. il nome del fotografo o le informazioni sul copyright.

Come opzione, TET PDF IFilter può salvare i metadati in formato testo indicizzato, in modo da poter essere utilizzati da motori che non supportano i metadati, come SQL Server.

I workflow beneficiano della ricerca nei documenti basata su XMP

La gestione dei metadati XMP può essere integrata in numerose applicazioni che richiedono la possibilità di effettuare ricerche nei documenti digitali. Seguono due esempi tipici.

Publishing: i professionisti utilizzano Adobe e altri software di publishing per creare in modo interattivo documenti e metadati. Essi assegnano ai documenti parole chiave, nome dell'autore, informazioni sul copyright e altre proprietà XMP comuni. Per cercare o raggruppare i documenti in base alle proprietà dei metadati assegnate, possono servirsi di Adobe Bridge che prevede l'utilizzo di schemi XMP comuni come Dublin Core e IPTC.

Documentazione tecnica: un vasto numero di documenti viene creato manualmente o automaticamente per essere conservato in archivi aziendali o dipartimentali. A questi documenti è possibile accedere con un comune tool Windows per il recupero delle informazioni, come Microsoft Office SharePoint Server (MOSS) sui sistemi server, Windows Desktop Search (WDS) sulle workstation, o altri prodotti di retrieval. Dopo aver integrato TET PDF IFilter a questi prodotti, gli utenti sono in grado di cercare i documenti basati sui metadati XMP, i contenuti della pagina attuale o anche le proprietà dell'immagine. Mentre gli schemi XMP predefiniti soddisfano i requisiti di base, gli schemi XMP personalizzati possono essere utilizzati nelle ricerche per soddisfare le esigenze specifiche dell'azienda.



PDFlib GmbH

Franziska-Bilek-Weg 9
80339 Monaco, Germania
Tel. +49 • 89 • 452 33 84-0
support@pdflib.com
www.pdflib.com/developer/pdfa

PDFlib GmbH è un'azienda specializzata nel settore delle tecnologie PDF. I prodotti PDFlib vengono utilizzati in tutto il mondo dal 1997. L'azienda segue attentamente gli sviluppi e i trend del mercato, come gli standard ISO per il formato PDF. PDFlib GmbH distribuisce i suoi prodotti in tutti i continenti e detiene una posizione da leader di mercato in Nordamerica, Europa e Giappone.



PDF/A Competence Center

www.pdfa.org

Lo standard ISO PDF/A sta riscuotendo notevole interesse sul mercato. Associazione internazionale fondata nel 2006, il PDF/A Competence Center ha lo scopo di rispondere alla crescente domanda di informazioni sul tema PDF/A e di promuovere lo scambio di idee ed esperienze. L'enorme esperienza di numerosi esperti di PDF/A è disponibile a tutte le parti interessate. L'associazione organizza seminari e conferenze contribuendo alla standardizzazione del settore.

PDFlib GmbH è uno dei membri fondatori del PDF/A Competence Center.