

PDF/A con i prodotti PDFlib



 PDFlib® White paper

Il formato PDF/A specificato nello standard ISO 19005 definisce un subset di PDF uniforme e robusto per l'archiviazione nel lungo periodo e per lo scambio sicuro di dati nelle aziende o nelle istituzioni. Il presente white paper illustra le caratteristiche dei prodotti PDFlib ideati per creare file PDF/A adatti all'archiviazione nel lungo periodo di documenti elettronici.

PDF/A-1a e PDF/A-1b

Lo standard PDF/A-1, conosciuto in passato come ISO 19005-1, ha l'obiettivo di garantire la riproduzione affidabile nel lungo periodo dei documenti digitali. Lo standard, basato su PDF 1.4, delinea delle restrizioni per l'uso di colori, font, annotazioni e altri elementi. Esistono due diverse versioni di PDF/A-1, entrambe possono essere create ed elaborate con PDFlib:

- ▶ ISO 19005-1 Il livello di conformità B (PDF/A-1b) garantisce la riproduzione affidabile dell'aspetto visivo del documento nel lungo periodo. PDF/A-1b fa sì che il documento conservi lo stesso aspetto anche quando verrà visualizzato o stampato in futuro.
- ▶ ISO 19005-1 Il livello di conformità A (PDF/A-1a) si basa sul livello B, ma prevede delle importanti funzionalità per i »Tagged PDF«: per mantenere la struttura logica e l'ordine naturale di lettura del documento sono necessarie delle informazioni sulla struttura e una semantica affidabile del testo. In altre parole, PDF/A-1a non preserva soltanto l'aspetto del PDF nel lungo termine, ma interpreta anche il suo contenuto (semantica) rendendolo accessibile a tutti gli utenti.

Nel presente testo, quando viene fatto riferimento a PDF/A-1 senza specificare alcun livello di conformità, sono da intendersi entrambi i livelli. L'implementazione PDF/A-1 in tutti i prodotti PDFlib si basa sulla norma ISO 19005-1:2005 comprensiva della Technical Corrigendum 1 (pubblicata dalla ISO nel 2007) e tutte le TechNotes pubblicate dal PDF/A Competence Center.

Requisiti e restrizioni di PDF/A

PDF/A richiede delle determinate funzionalità PDF e ne proibisce delle altre. Ad esempio, per assicurare una riproduzione precisa del testo, tutti i font utilizzati nel testo devono essere embedded; per poter garantire la corretta riproduzione dei colori è necessario che tutti i colori vengano specificati in un formato indipendente dal sistema. I metadati devono essere integrati in formato XMP; la crittografia non deve essere utilizzata.

Oltre a queste semplici prescrizioni, PDF/A richiede comunque numerose altre caratteristiche PDF, generalmente più complesse, come la presenza di specifiche informazioni nella struttura dei font, e non ammette alcune strutture critiche, ad esempio alcune combinazioni di font TrueType e di codifiche. I programmatori sono chiamati a implementare e a controllare numerosi aspetti prima di poter realizzare dei prodotti interamente compatibili con lo standard PDF/A. PDF/A è molto di più di un »PDF con font embedded«!

Supporto PDF/A nella famiglia di prodotti PDFlib

PDFlib offre agli sviluppatori uno strumento che prevede le seguenti funzionalità per PDF/A:

- ▶ Creazione di un nuovo PDF/A, ad esempio da testo proveniente da database

- ▶ Conversione di immagini raster (ad es. immagini scannerizzate) in PDF/A
- ▶ Elaborazione di PDF/A esistenti, ad esempio unione o divisione
- ▶ Creazione di PDF/A-1a con informazioni sulla struttura (Tagged PDF)
- ▶ Inserimento dei metadati XMP nei documenti creati, comprensivo di estensioni XMP (vedi sotto).

Tutte queste operazioni possono essere eseguite con dei semplici comandi di PDFlib. PDFlib prevede dei codici di esempio per numerosi linguaggi di programmazione e ambienti di sviluppo. Ulteriori tecniche di programmazione per PDF/A sono disponibili nel PDFlib Cookbook. Grazie alla relativa somiglianza tra gli standard PDF/A e PDF/X (ISO 15930) per il settore grafico, il supporto PDF/A offerto in PDFlib ha tratto vantaggio dalla nostra pluriennale esperienza con numerose versioni di PDF/X. Per semplificare l'integrazione dei font come richiesto da PDF/A, il Japanese Resource Kit per i prodotti PDFlib comprende i più comuni font giapponesi con una licenza integrabile compresa nella licenza del software e i profili ICC specifici per il paese, le CMaps e la documentazione per gli utenti giapponesi.

Creazione di documenti conformi a PDF/A

I file conformi a PDF/A vengono creati con PDFlib nel seguente modo:

- ▶ PDFlib esegue le impostazioni formali per PDF/A, come la versione del PDF e le informazioni di identificazione richieste da XMP.
- ▶ Il programma client PDFlib deve utilizzare alcuni comandi e opzioni (ad es. per il font embedding).
- ▶ Il programma client PDFlib non deve utilizzare altri comandi o impostazioni (ad es. la crittografia).

Se il programma client PDFlib si attiene a queste regole, i documenti prodotti saranno in PDF/A valido. Qualora PDFlib rilevi una violazione delle regole di creazione dei PDF/A, verrà indicata un'eccezione che dovrà essere gestita dall'applicazione. In caso di errore non verrà generato nessun file PDF; in questo modo si è sempre sicuri di non creare, in presenza di errori, file non conformi. Maggiori informazioni sulle operazioni richieste o vietate sono contenute nella documentazione PDFlib.

Specifiche colore indipendenti dal sistema

Per garantire l'uniformità della riproduzione dei colori, PDF/A richiede l'uso di colori indipendenti dal sistema, solitamente definiti dai profili ICC o dalle specificazioni CIE Lab. L'output descrive le caratteristiche di colore del documento. Se da un lato questi concetti sono largamente utilizzati nel settore grafico, gli sviluppatori di soluzioni PDF non sono familiari con i temi inerenti la gestione del colore. PDFlib semplifica la creazione di output indipendenti dal sistema, qualsiasi siano i dati in ingresso.

- ▶ L'output PDF/A può essere creato con o senza profili ICC nell'output intent.
- ▶ Nel caso tradizionale di un testo di colore nero, PDFlib selezionerà automaticamente uno spazio colore appropriato (Lab o DeviceGray) in base al fatto che sia stato specificato o meno un profilo ICC per l'output intent.
- ▶ I profili ICC esterni o i profili embedded nelle immagini permettono un controllo dettagliato dei colori.
- ▶ Grazie ai profili ICC per le applicazioni più comuni, scaricabili gratuitamente, è possibile creare i file PDF/A in breve tempo.

Le immagini raster, ad es. TIFF e JPEG, svolgono un ruolo fondamentale nella creazione dei documenti. I documenti cartacei scannerizzati e le fotografie digitali sono esempi comuni di immagini raster utilizzate nei normali documenti. Se nei workflow moderni le immagini raster risultano sempre più spesso indipendenti dal sistema (solitamente grazie ad un profilo colore ICC embedded) e in virtù di ciò compatibili con PDF/A, la maggior parte delle immagini continuano ad essere dipendenti dal sistema, ad esempio le scannerizzazioni in bianco e nero o le RGB senza un profilo ICC associato. PDFlib supporta entrambe le situazioni:

- ▶ Utilizzo dei profili ICC embedded nelle immagini raster.
- ▶ Applicazione dei profili ICC esterni.
- ▶ Come soluzione per i file di origine sconosciuta, PDFlib prevede un profilo ICC RGB integrato per numerosi ambienti hardware e software.
- ▶ Specificando un profilo ICC per l'output intent, le immagini indipendenti dal sistema possono essere utilizzate senza dover usare un profilo ICC con le immagini.

La documentazione di PDFlib affronta le strategie PDF/A legate al colore per le più comuni applicazioni.

XMP made easy

PDF/A affida l'uso dei metadati XMP per la memorizzazione delle informazioni sul documento allo stesso PDF. XMP offre un framework potente e affidabile per memorizzare le proprietà dei metadati standard e personalizzate (vedi anche il white paper dedicato ai metadati XMP). Se i metadati vengo-

no già utilizzati nel workflow è possibile creare dei flussi XMP che PDFlib integrerà nell'output PDF/A. Gli sviluppatori che non sono familiari con i metadati XMP non sono comunque costretti ad approfondire questo vasto tema. PDFlib creerà l'output XMP richiesto per i documenti PDF/A, mappando automaticamente i campi informativi dei documenti classici nei corrispondenti costrutti XMP così come richiesto dagli standard. In questo modo, i programmatori potranno sfruttare tutta la potenza dei metadati XMP, mentre la funzione di creazione automatica dei metadati XMP offerta da PDFlib permetterà di soddisfare le esigenze in situazioni in cui sono necessari dei semplici metadati.

Estensioni XMP

XMP è una soluzione caratterizzata da un'elevata estensibilità, in questo modo è possibile rispondere alle esigenze specifiche di un'azienda o di un determinato settore creando delle estensioni dei metadati XMP. PDF/A supporta questo approccio, ma per facilitare il recupero delle informazioni anche in un secondo tempo è necessario che nel documento venga integrata, secondo le modalità definite, una descrizione leggibile dello schema. I prodotti PDFlib sono stati i primi a supportare le estensioni XMP per i PDF/A. PDFlib verifica i metadati XMP e le estensioni fornite dall'utente così da assicurare che l'output generato sia completamente conforme a PDF/A. Maggiori informazioni sui metadati nei documenti PDF/A sono disponibili all'indirizzo www.pdflib.com.

Elaborazione di documenti PDF/A già esistenti

Quando vengono importate delle pagine da documenti conformi a PDF/A, si applicano delle ulteriori regole. Ad esempio, con Adobe Acrobat è facile combinare due documenti PDF/A ed ottenere un documento in uscita che non risulta più conforme alle specifiche PDF/A (senza che questo venga segnalato). Elaborando i documenti PDF/A esistenti, PDFlib+PDI esamina attentamente le proprietà PDF/A di tutti i documenti in ingresso e in uscita per garantire che siano conformi alle specifiche PDF/A. Per un ulteriore controllo, l'output intent di un documento importato può essere copiato nel PDF in uscita, clonando in questo modo le proprietà dei colori PDF/A di un documento già esistente. Allo stesso modo, i metadati XMP dei documenti importati possono essere copiati o uniti.

Creazione di PDF/A-1a con Tagged PDF

PDF/A-1a può essere considerato come PDF/A-1b più Tagged PDF: necessita di informazioni sulla struttura del documento e imposta delle condizioni sui font per far sì che il testo possa essere interpretato correttamente. In questo modo, i documenti PDF/A-1a sono completamente accessibili anche agli utenti con disabilità. Oltre a preservare l'aspetto visivo del documento, viene mantenuto anche il significato del contenuto.

Il supporto PDF/A-1a di PDFlib si basa sulle funzionalità per produrre Tagged PDF: ogni elemento di contenuto può essere posizionato in una determinata posizione nella struttura ad albero del documento; gli elementi di contenuto non rilevanti per la struttura del documento (ad es. header e footer, impaginazione) possono essere taggati come elementi personalizzati, in questo modo essi verranno ignorati quando il documento verrà riutilizzato (ad es. quando il documento viene letto da un software o convertito in un altro formato). Nelle immagini è possibile inserire del testo alternativo; il testo può essere letto anche da utenti con problemi visivi.

Per creare i Tagged PDF è necessario disporre di numerose informazioni sulla struttura logica del documento. PDFlib si occupa dei dettagli relativi al PDF, ma non è in grado di dedurre la struttura del documento sulla base del suo contenuto.

Sulla base del supporto Tagged PDF esistente, PDFlib può produrre un output conforme alle specifiche PDF/A-1a.

Elaborazione di documenti conformi a PDF/A con PDFlib PLOP

PDFlib PLOP offre diverse funzionalità per i PDF, tutte in grado di supportare PDF/A. In questo modo, seguendo una procedura di elaborazione conforme allo standard, dai documenti PDF/A in ingresso è possibile generare dei file conformi alle specifiche PDF/A. In caso contrario viene visualizzato un messaggio di errore. Alcuni esempi:

- ▶ È possibile servirsi di PLOP per inserire dei metadati XMP nei documenti PDF/A già esistenti. Poiché PLOP supporta le estensioni XMP per le specifiche PDF/A, può essere utilizzato come uno strumento di post elaborazione per i documenti già esistenti, oppure per sfruttare le estensioni XMP anche quando queste non vengono supportate dal software di creazione dei documenti.
- ▶ Poiché la crittografia non è supportata, se si tenta di crittografare i documenti PDF/A verrà visualizzato un messaggio di errore. Per crittografare un documento PDF/A in PLOP è necessario rinunciare alla specifica PDF/A.

Firme digitali per PDF/A con PDFlib PLOP DS

PLOP DS include tutte le funzionalità di PDFlib PLOP. Inoltre permette di applicare una firma digitale nei documenti PDF. Poiché PLOP DS supporta le regole PDF/A per i campi della firma, i documenti PDF/A firmati saranno conformi allo standard PDF/A.

Verifica di PDF/A

Quando si sviluppa un workflow basato sugli standard, è buona prassi utilizzare dei tool appositi per verificare i risultati e confrontarli con gli standard. Per quanto riguarda lo standard PDF/A sono disponibili dei tool per verificare se un documento PDF è conforme o meno allo standard ISO.

Membro fondatore del PDF/A Competence Center, PDFlib GmbH è stata attivamente impegnata nello sviluppo e nell'implementazione della Isartor Test Suite per PDF/A-1 (vedi www.pdfa.org). Il tool Isartor è lo strumento di riferimento per la verifica PDF/A. PDFlib GmbH offre prodotti per creare ed elaborare documenti PDF/A, senza prevedere però tool per la verifica dei documenti PDF/A o per convertire i documenti PDF in PDF/A.

Nel corso delle procedure QA eseguite dalla nostra azienda per verificare la conformità PDF/A, abbiamo confrontato tutti gli strumenti di verifica delle specifiche PDF/A disponibili. I risultati di questo confronto sono consultabili in »Bavaria Report on PDF/A Validation Accuracy« (disponibile sul nostro sito web). Il Bavaria Report ha avuto un forte impatto nel processo di consolidamento dei tool di verifica PDF/A.

I PDF/A creati con i prodotti PDFlib sono stati verificati con Preflight in Adobe Acrobat 9 e con tutti gli altri strumenti di verifica PDF/A per garantire che nessun aspetto dello standard venga violato.

Verso lo standard PDF/A-2

Il settore dei PDF ha fatto enormi passi in avanti da quando, nel 2005, è stato pubblicato lo standard PDF/A-1. Fra i numerosi traguardi raggiunti, PDF 1.7 è stato standardizzato come ISO 32000-1 nel 2008. Per permettere l'integrazione delle nuove funzionalità nei workflow con PDF/A, stanno per essere apportate delle modifiche agli standard di archiviazione: PDF/A-2 (ISO 19005-2) contemplerà diverse nuove funzionalità attualmente non previste da PDF/A-1, tra le altre novità in questo ambito la compressione JPEG 2000, i contenuti opzionali (livelli), la trasparenza, il PDF Portfolio, ecc. Lo standard PDF/A-2 non renderà obsoleto il PDF/A-1 e non costringerà gli utenti a migrare alla nuova versione. Sarebbe un evidente controsenso per uno standard destinato a garantire la riproduzione affidabile nel lungo periodo!

ISO 19005-2 verrà pubblicato nel 2010. PDFlib GmbH offrirà il supporto per lo standard PDF/A-2 non appena questo sarà disponibile nella sua versione finale.



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.