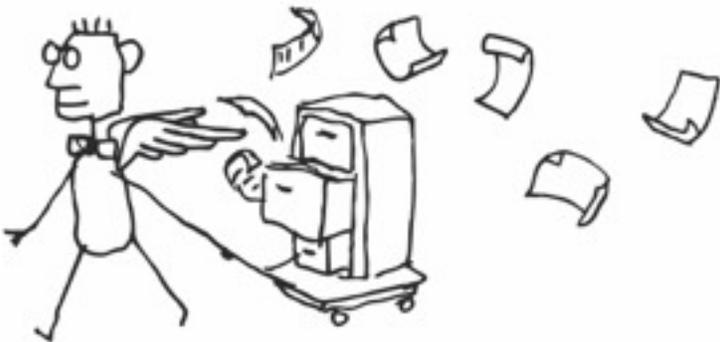


PDF/A PER L'ARCHIVIAZIONE A LUNGO TERMINE

White Paper



Lo Standard PDF/A e i suoi obiettivi



Il PDF/A è uno standard ISO per l'uso del formato PDF per l'archiviazione a lungo termine dei documenti elettronici. È stato pubblicato il primo ottobre 2005 come "ISO 19005-1: Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)." ["ISO 19005-1: Gestione del documento - formato di file per la conservazione a lungo termine del documento elettronico - parte 1: uso del PDF 1.4 (PDF/A-1)"].

Lo standard PDF/A è il risultato di più di 3 anni di lavoro congiunto di un comitato di rappresentanti di aziende e membri di diverse associazioni da tutto il mondo. L'ISO 19005-1 definisce "un formato di file basato sul PDF, conosciuto come PDF/A, che fornisce/prevede un meccanismo per rappresentare i documenti elettronici in modo da preservare il loro aspetto visivo nel tempo, indipendentemente dagli strumenti e sistemi utilizzati per creare, immagazzinare o tradurre i file." Lo standard non stabilisce né una strategia di archiviazione, né gli obiettivi di un sistema di archiviazione. Identifica un "profilo" per i documenti elettronici che assicura che questi possano essere riprodotti negli anni a venire.

È facilmente prevedibile che il PDF/A si affermi in futuro come lo standard preferito per archiviare documenti elettronici. Il formato PDF, così come il

software richiesto per creare, visualizzare ed elaborare documenti PDF, è utilizzato in tutto il mondo sia nel settore pubblico che nel privato. Il PDF è sempre più utilizzato nelle più svariate aree di applicazione in quanto formato di documento robusto e flessibile. Lo standard PDF/A ora garantisce che i documenti PDF siano accessibili con certezza anche per lunghi periodi di tempo, ed è quindi il miglior formato disponibile per l'archiviazione a lungo termine.

Esistono due diverse versioni di PDF/A-1:

- **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.

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 incorporati (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.

E' attualmente in sviluppo l'ISO 19005-1 Part-2 (PDF/A-2) presso il Comitato Tecnico. PDF/A includerà le nuove caratteristiche introdotte dalle versioni 1.5, 1.6 e 1.7 dello standard PDF e sarà retrocompatibile.

Aree di applicazione del PDF/A

Il PDF/A può essere usato per archiviare tutti i documenti elettronici destinati alla stampa, così come per l'archiviazione dei documenti cartacei scansionati. Per questo motivo il PDF/A diviene decisamente efficace in molti ambiti. Gli esempi che seguono sono presi da diverse aree di applicazione nelle quali il PDF/A è già stato implementato o nelle quali sono stati intrapresi progetti pilota.

Posta in Arrivo

Indipendentemente da come la posta arrivi in un'azienda, sia per posta tradizionale, per fax o per e-mail, essa deve essere archiviata in modo corretto. In un archivio elettronico, i documenti cartacei devono essere scansionati e le e-mail, compresi gli allegati, devono essere convertite in un formato adatto all'archiviazione. Un'infrastruttura uniforme può essere realizzata al meglio utilizzando il PDF/A.

Se ad esempio in seguito ad un aggiornamento hardware si effettua una migrazione da un programma di posta ad un altro, possono insorgere problemi di compatibilità e perdite, ovviate da un sistema di backup tramite archivio in PDF/A.

Posta in Uscita

La posta in uscita deve essere archiviata in modo simile a quella in arrivo. Un fattore importante da considerare è che la posta inviata in formato classico (copia cartacea) può anche essere archiviata con un formato elettronico affidabile. Ad oggi sono disponibili sul mercato numerosi convertitori di flussi di stampa che supportano il formato PDF/A.

Disegno Tecnico (CAD)

I disegni tecnici spesso necessitano di un software proprietario per essere visualizzati. La conversione a PDF/A permette di visualizzare i disegni con un software standard e fornisce la sicurezza che i disegni possano essere visualizzati e stampati anche dopo molti anni.

Gli stampatori e i service di pre-stampa possono convertire i loro file dal formato PDF/X, lo standard per la grafica e il layout poiché PDF/X e PDF/A sono perfettamente compatibili. Le grandi moli di documenti destinati alla stampa possono così essere catalogate e consultate all'occorrenza.

Lavoro scientifico

Molte università richiedono già che le tesi siano presentate nel formato PDF/A, quando possibile così come gli elaborati destinati alla divulgazione e distribuzione permanente.

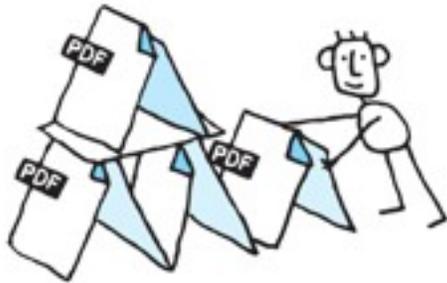
Migrazione dell'archivio

Un numero infinito di specialisti si ritiene insoddisfatto delle infrastrutture di archivio attuali per numerose ragioni. Prima fra tutte lo standard di fatto utilizzato, TIFF-G4, che è una variante in bianco e nero del TIFF, non è affidabile per l'archiviazione in quanto i documenti scansionati vengono gestiti come immagini e digitalizzati pixel per pixel: l'aspetto visivo del file viene mantenuto, in quanto font ed elementi grafici sono inclusi come immagini, ma la dimensione del file è spesso ingestibile. I testi poi se acquisiti senza OCR (riconoscimento dei caratteri) non sono ricercabili e quindi inefficaci per l'archiviazione e la consultazione semantica.

Inoltre, le disposizioni di legge attuali stanno formulando nuovi requisiti per i documenti archiviati. Molte aziende hanno già iniziato la migrazione dei loro archivi eterogenei nell'unico formato PDF/A.

Settori

L'archiviazione a lungo termine dei documenti elettronici ha importanti priorità in molti settori, pur avendo obiettivi diversi. Di seguito sono elencati alcuni esempi di alcuni settori selezionati e delle loro esigenze. La lista non è assolutamente completa, ma offre di certo una prospettiva di come il PDF/A può offrire benefici significativi per l'archiviazione documentale a lungo termine nel vostro settore.



Banche e assicurazioni

I file utilizzati dalle agenzie di credito o assicurative hanno spesso un ciclo di vita di durata superiore ai 50 anni. L'archiviazione elettronica basata sul PDF/A in questo ambito sta ottenendo una attenzione crescente non solo per i nuovi file destinati all'archiviazione ma per tutti i documenti già prodotti in passato. Il formato standardizzato PDF/A offre preziosi benefici sia per la riproducibilità, che per la consultazione.

Industria

Il settore dell'ingegneria è particolarmente adatto come candidato per l'archiviazione a lungo termine. I disegni industriali per la costruzione di ponti, aeroplani o macchine, in molti casi, vanno archiviati almeno per 99 anni. Potrebbe inoltre essere necessario archiviare numerose versioni dei documenti.

Case editrici

Il PDF/X si è affermato in passato come formato standard per lo scambio di documenti destinati ai processi di stampa. Un file PDF/X può essere salvato al contempo come file PDF/A. La soluzione risulta utile ed efficace nei casi in cui le pubblicazioni debbano essere immagazzinate in un archivio a lungo termine dopo la loro produzione.

Settore pubblico e Governo

L'uso del PDF/A in questo settore dipende molto dal tipo di agenzia pubblica o governativa. Il PDF/A può trovare applicazione in quelle agenzie in cui è necessario conservare registrazioni digitali o digitalizzate, come ad esempio progetti, registrazioni dell'edilizia e notifiche.

In particolare il Codice dell'Amministrazione Digitale, che traccia il quadro legislativo generale entro cui può e deve attuarsi la digitalizzazione dell'azione amministrativa italiana prevede che:

- art. 42: "le pubbliche amministrazioni valutano in termini di rapporto tra costi e benefici il recupero su supporto informatico dei documenti e degli atti cartacei dei quali sia obbligatoria o opportuna la conservazione e provvedono predisposizione dei conseguenti piani di sostituzione degli archivi cartacei con archivi informatici, nel rispetto delle regole tecniche adottate ai sensi dell'articolo 71"
- art. 44: "Il sistema di conservazione dei documenti informatici garantisce:
 - a. l'identificazione certa del soggetto che ha formato il documento e dell'amministrazione o dell'area organizzativa omogenea di riferimento di cui all'articolo 50, comma 4, del decreto del Presidente della Repubblica 28 dicembre 2000, n. 445;
 - b. l'integrità del documento;
 - c. la leggibilità e l'agevole reperibilità dei documenti e delle informazioni identificative, inclusi i dati di registrazione e di classificazione originari;
 - d. il rispetto delle misure di sicurezza previste dagli articoli da 31 a 36 del decreto legislativo 30 giugno 2003, n. 196, e dal disciplinare tecnico pubblicato in allegato B a tale decreto."

Vantaggi del formato

Lo standard PDF/A definisce i requisiti per la creazione di documenti adatti all'archiviazione, sulla base del formato PDF, ampiamente diffuso. Il PDF/A combina le capacità universali del PDF con l'affidabilità e la consistenza essenziali per l'archiviazione a lungo termine.

PDF/a: il PDF certo e conservabile

Un file PDF/A contiene tutto ciò che serve per riprodurlo e non contiene nulla che potrebbe comprometterne o alterarne la riproduzione.

Più capace e compatto del TIFF G4 e del JPEG

I formati raster immagazzinano tutti i contenuti della pagina come pixel. Nel PDF/A il testo si basa sui caratteri ed è quindi ricercabile. I PDF/A con pagine scansionate possono avere un testo OCR integrato e posizionato direttamente dietro al testo visibile, ricostruendo il testo in un livello invisibile. Questo permette al documento di essere consultabile, conservando completamente l'aspetto di fac-simile.

Più universale dei file originati da applicazioni proprietarie

I file originati da programmi proprietari generalmente hanno bisogno delle rispettive applicazioni per essere aperti e visualizzati. Per il PDF/A è sufficiente una sola applicazione indipendente dalla piattaforma di creazione.

La firma digitale PDF/A

Il file in formato PDF supporta firme interne (embedded). Lo standard PDF/A consente l'utilizzo

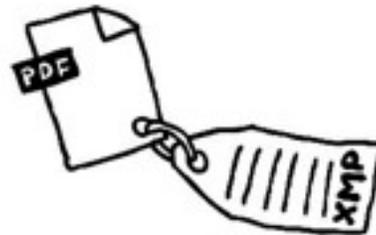
della firma digitale con un numero minimo di requisiti. I documenti in formato PDF/A possono essere quindi firmati in modo affidabile conservando al contempo la compatibilità con lo standard e tutti i vantaggi offerti dalla firma digitale. Questo supporto è inoltre estendibile alle nuove tecnologie nel campo delle firme che saranno introdotte in futuro.

Layout ed elementi grafici

Le immagini interne ad un file PDF/A devono essere sempre riproducibili quindi vengono integrate nel file come elemento fisso. Per la gestione dei colori, PDF/A ingloba l'impostazione di output prevista (per monitor, stampa, etc.) o l'imposizione di un profilo colore predefinito in spazi colori ben conosciuti (RGB, sRGB, CMYK, Lab, etc.) così da poter essere gestito.

Metadati per la ricostruzione dei documenti

I metadati, informazioni supplementari "nascoste" nei file, possono essere personalizzati dal creatore del file e quindi trasmettere istruzioni supplementari sul documento ai posteri senza che questo venga reso palese al lettore del documento.



PDF/A per grandi volumi di dati

Quando l'esigenza di archiviazione deriva dalla gestione di dati di dimensioni notevoli, spesso derivanti da database o dati strutturati creati "al volo" partendo da elementi variabili (come ad esempio nome, indirizzo, codice cliente per generare una fattura) la soluzione efficace è sfruttare le potenzialità di una libreria di linguaggio conforme allo standard PDF/A: è in questo ambito che **PDFlib** si propone come la soluzione ideale.

PDFlib: il più diffuso strumento per generare PDF

PDFlib è il più diffuso strumento per generare e manipolare "al volo" i file in formato PDF. I principali obiettivi di PDFlib sono la creazione dinamica di PDF su server nonché l'implementazione della funzionalità "Salva come PDF" nelle applicazioni già esistenti. Si rivela utile per numerosi altri usi come ad esempio:

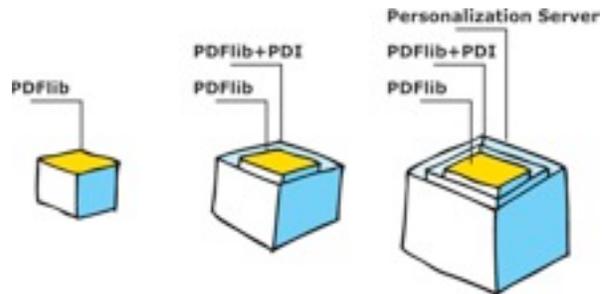
- creazione dinamica di fatture
- mass mailing inserendo dati personali in documenti PDF già esistenti
- fatture e report importati da documenti Office© di Microsoft
- integrazione con altri strumenti del pacchetto Office©
- indicizzazione dei file PDF per i motori di ricerca
- crittografia di documenti PDF riservati .

La famiglia di prodotti PDFlib è disponibile in tre versioni: PDFlib, PDFlib+PDI (PDF import) e PDFlib Personalization Server (PPS).

PDFlib+PDI comprende tutte le funzionalità di PDFlib e la funzione PDF Import Library (PDI). Grazie a PDI potrete aprire i documenti PDF esistenti e aggiungere delle pagine all'output di PDFlib.

Oltre alle funzionalità di PDFlib+PDI, **PDFlib Personalization Server (PPS)** offre con i PDFlib block delle funzionalità per elaborare i dati variabili. PPS

rende le applicazioni indipendenti dalle modifiche di layout.



Con il plugin **PDFlib Block** per Acrobat, il designer crea dei segnaposti per introdurre testi e immagini variabili. Queste aree, chiamate blocchi, contengono numerose proprietà, come dimensioni dei font, colore e proporzioni. Intervenedo sul server, PDFlib Personalization Server è in grado di riempire questi blocchi con dati variabili, provenienti ad esempio da un database, e creare l'output finale.

PDFlib offre anche altri strumenti di utilità supplementari quali:

- **PDFlib PLOP**, un tool versatile per linearizzare, ottimizzare, riparare, analizzare, crittografare e decrittografare i documenti PDF; la versione PDFlib PLOP DS permette di inserire una firma digitale nei file PDF;
- I **PDFlib Text Extraction Toolkit (TET)** è uno strumento brevettato destinato agli sviluppatori per l'estrazione sicura di testo e immagini dai documenti PDF;
- **PDFlib pCOS** offre un metodo semplice ed elegante per recuperare le informazioni, anche non visibili, contenute all'interno di un documento PDF. Con pCOS è possibile ad esempio rilevare i metadati, gli elementi interattivi (link, ecc.) o le dimensioni della pagina.

Supporto PDF/A nella famiglia di prodotti PDFlib

PDFlib offre agli sviluppatori uno strumento che supporta 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 scansite) 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 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 scaricabile con il prodotto.

Elaborazione di documenti PDF/A già esistenti

Quando vengono importate delle pagine da documenti conformi a PDF/A, si applicano 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 alcune 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.

Informazioni supplementari su PDFLib

PDFlib funziona su quasi tutte le piattaforme informatiche. Sono disponibili varianti a 32 e 64 bit per tutte le versioni standard di Windows, Mac OS X, Linux e Unix, nonché per IBM eServer iSeries e zSeries mainframe.

Per garantire le migliori performance possibili e ridurre l'overhead, il PDFlib core è scritto in codice C ottimizzato. Grazie ad una semplice API (Application Programming Interface) è possibile accedere alle funzionalità di PDFlib da numerosi ambienti di sviluppo (ad esempio .NET per C#, VB.NET, ASP.NET, COM, C e C++ etc).

Benefici di PDFlib: affidabilità, semplicità e assistenza professionale

In tutto il mondo sono migliaia i programmatori che utilizzano il software PDFlib. I prodotti PDFlib soddisfano tutti i requisiti in termini di qualità e performance per l'utilizzo sui server. Tutti i prodotti PDFlib sono adatti per l'utilizzo continuo su server e per le operazioni di batch senza supervisione.

I prodotti PDFlib sono estremamente veloci e sono in grado di elaborare fino a centinaia di pagine al secondo. L'interfaccia di programmazione è intuitiva e di facile apprendimento.

In caso di problemi potete contare sul servizio di assistenza professionale. Per garantire che il workflow a livello aziendale non subisca mai interruzioni, PDFlib offre la possibilità di sottoscrivere un contratto di servizio. In questo modo potrete accedere alle versioni più recenti e potrete contare su tempi di risposta celeri qualora si dovesse verificare un problema.

I nostri prodotti supportano tutte le lingue internazionali e Unicode. Vengono utilizzati da utenti in ogni angolo del mondo.

Licenza

Sono a disposizione diverse tipologie di licenza: server, integrativa, aziendale e codice sorgente. È altresì possibile sottoscrivere un contratto di servizio per

usufruire di un'assistenza professionale con tempi brevi di risposta e aggiornamenti gratuiti.

PDFlib è un marchio registrato PDFlib GmbH, un'azienda specializzata nelle tecnologie legate a PDF e distribuita in esclusiva per l'Italia da **omnidea srl**.

Informazioni supplementari in merito alla famiglia PDFlib possono essere reperite sul sito www.pdflib.it.



omnidea srl

via delle industrie,15
30175 Venezia Marghera

Tel: +39 041 509 3441

Fax: +39 041 509 3641

info@omnidea.it

www.omnidea.it



PDFlib GmbH

Franziska-Bilek-Weg 9
80339 München Germany

Tel +49 89 452 33840

Fax +49 89 45233 8499

info@pdflib.com

www.pdflib.com



è un marchio registrato PDFlib GmbH
distribuito in esclusiva per l'Italia da

www.pdflib.it

