

ホワイトペーパー： PDFlib 製品群による PDF/A

ISO 19005 標準規格で記述されている PDF/A 形式群は、長期間にわたって安全にアーカイブしたり、企業や官庁の環境で確実なデータ交換に使用したりすることができるような、PDF の一貫した堅牢なサブセットを提供することを目指しています。このホワイトペーパーでは、長期間の文書アーカイブにたえる PDF/A 出力を作成するための、PDFlib の諸機能を説明します。

PDF/A-1a と PDF/A-1b PDF/A-1 は、正式には国際標準規格 ISO 19005-1 として、電子文書の確実な長期保存を目的としています。この規格は、PDF 1.4 をベースとして、色やフォント・注釈等の要素の使用に関していくらかの制限事項を設けたものです。PDF/A-1 には次の 2 種類があり、どちらも PDFlib で作成・処理することができます。

- ▶ ISO 19005-1 レベル B 準拠 (PDF/A-1b) は、文書の見え目が長期間を経ても維持されることを保証するものです。つまり PDF/A-1b は、文書がいつか将来処理された時に、同じ見え方をすることを保証します。
- ▶ ISO 19005-1 レベル A 準拠 (PDF/A-1a) は、レベル B をベースとしつつ、「タグ付き PDF」の厳格な要件を兼ねそなえたものです。これは、文書の論理構造と自然な読み順を保持するために、構造情報と、正しいテキストセマンティクスを必要とします。つまり PDF/A-1a は、文書が将来利用される時に同じ見え方をすることを保証するだけでなく、その内容 (セマンティクス) が確実に解釈されること、そして障害をもつ利用者にとってアクセシブルであることをも保証します。

以下、PDF/A と (準拠レベルを付けずに) 言ったときは、両方の準拠レベルを指すものとし、すべての PDFlib 製品の PDF/A 実装は、ISO 19005-1 (2005 年) と技術正誤表 1 (2007 年) に基づいています。

PDF/A の必須事項と制限事項 PDF/A では、PDF のいくつかの機能を必須事項とし、またいくつかの特性を禁則事項に指定しています。たとえば、テキストの正確な再現を保証するため、文書内で使っているフォントはすべて埋め込まなければなりませんし、色の正確な再現を保証するため、色はすべてデバイス非依存な形で指定しなければなりません。メタデータを XML 形式で埋め込む必要がありますし、暗号化は使用してはなりません。

しかし PDF/A では、このような単純明快な必要事項だけでなく、もっと目に見えにくい部分のさまざまな PDF の機能があわせて必須とされていたり (フォントのデータ構造内に特定の項目が必須等)、ある種の好ましからざる構造が禁止されていたり (TrueType フォントとエンコーディングの組み合わせ方等)。こうした多くの要件を実装し試験してはじめて、ソフトウェアの開発者は、規格に完全準拠した PDF/A 製品を作り上げることができます。PDF/A とは、単なる「フォントを埋め込んだ PDF」をはるかに超えた存在なのです。

PDFlib 製品スイートにおける PDF/A 対応 PDFlib は、アプリケーション開発者に対して、PDF/A に関連した次のような操作を可能にするツールキットを提供します。

- ▶ PDF/A をスクラッチから作成 (データベース内のテキストを流し込む等)
- ▶ ラスタ画像 (スキャン等) を PDF/A に変換
- ▶ 既存の PDF/A 文書进行处理 (PDF 文書の連結・分割等)
- ▶ 構造情報 (タグ付き PDF) を持たせた PDF/A-1a を作成

これらの操作はすべて、シンプルな PDFlib 関数呼び出しで実装できます。さまざまなプログラミング言語や開発環境に対するサンプルコードが、PDFlib のディストリビューションとともに提供されています。PDF/A は、グラフィックアート業界のための標準規格である PDF/X (ISO 15930) との重なりがかなり大きいことから、PDFlib における PDF/A への対応にあたっては、何年にもわたってさまざまな種類の PDF/X に対応しつづけてきた当社の蓄積が活かされています。

PDF/A 準拠の出力の作成 PDFlib による PDF/A 準拠の出力の作成は、次の手段で実現されます。

- ▶ PDFlib は、PDF バージョン番号や必須 XMP 準拠項目等、PDF/A のためのいくつかの形式設定を自動的に遂行します。
- ▶ PDFlib のクライアントプログラムでは、特定の関数呼び出しとオプション（フォント埋め込み等）を明示的に使用する必要があります。
- ▶ PDFlib のクライアントプログラムでは、特定の関数呼び出しとオプション設定（暗号化等）の使用を避ける必要があります。

PDFlib のクライアントプログラムがこれらの規則に従うならば、有効な PDF/A 出力が保証されます。PDFlib が PDF/A 作成規則に対する違反を検出したときには、例外が発生するので、アプリケーション側で処理する必要があります。エラー時には PDF 出力は一切作成されませんので、エラーが起きたときに非準拠の出力が作成されるおそれは全くありません。具体的な必須・禁則操作については PDFlib の説明書で説明しています。

デバイス非依存な色指定 一貫した色再現を維持するために、PDF/A ではデバイス非依存な色の使用が必須で、通常これは ICC プロファイルか CIE Lab 色指定で実現されます。出力インテントを設定して文書の色特性を記述することもできます。こうした概念はグラフィックアート業界では常識ですが、業務系の PDF 開発者は必ずしも、カラーマネジメントの諸概念に通暁しているとはかぎりません。そこで PDFlib は、入力データの出所にかかわらず、デバイス非依存な出力の作成を容易にします。

- ▶ PDF/A 出力を、出力インテント内の ICC プロファイルありでもなしでも作成できます。
- ▶ よくある場合の黒のテキストに対しては、PDFlib は、出力インテントに対する ICC プロファイルが指定されているかどうかに応じて、適切な色空間（Lab か DeviceGray か）を自動的に選択します。
- ▶ 外部 ICC プロファイルや、画像に埋め込まれたプロファイルは、精密な色制御を可能にします。
- ▶ 代表的な運用類型のための ICC プロファイルが、PDFlib のディストリビューションとともに提供されているので、有効な PDF/A 出力を迅速に実現することができます。

ラスタ画像（TIFF・JPEG 等）は、文書作成の中で不可欠な役割を担います。スキャンした紙文書や、デジタルカメラの写真は、文書作成ワークフローにおける代表的なラスタ画像です。最近のワークフローでは、ラスタ画像データはすでにデバイス非依存になっているかもしれないので（その手段としては ICC カラープロファイルが埋め込まれているのが普通）、その場合は PDF/A 互換ですが、レガシ画像データはたいていはデバイス依存で、たとえば ICC プロファイルのない白黒や RGB のスキャンであったりします。PDFlib はこのどちらの場合にも対応します。

- ▶ ラスタ画像に埋め込まれた ICC プロファイルは尊重されます。
- ▶ 画像に外部 ICC プロファイルを適用することもできます。
- ▶ 出所の不明なレガシデータのための代替策として、PDFlib は、多くのハードウェア・ソフトウェア環境にマッチする sRGB プロファイルを内蔵しています。
- ▶ 出力インテントに対して ICC プロファイルを指定すれば、画像に ICC プロファイルを適用しなくても、デバイス依存な画像データが使えるようになります。

PDFlib の説明書では、代表的な運用類型における PDF/A カラー戦略を説明していません。

XMP らしく作成 PDF/A では、文書に関する情報を、XMP メタデータを用いて PDF 自体の中に記録しておくことが義務づけられています。XMP とは、標準やカスタムのデータを記録するための強力かつ柔軟なフレームワークを提供するものです (XMP に関する別立ての当社ホワイトペーパーもごらんください)。すでに XMP メタデータをワークフローで使っている人なら、XMP ストリームをまるごと自分で作ってしまって、それを PDFlib で PDF/A 出力に入れ込ませるという手もあります。しかし XMP になじみのない開発者なら、こんな奥の深い分野に首を突っ込む必要はありません。PDFlib は、PDF/A として必要な XMP 出力を作成しますし、規格の義務づけるとおりに、クラシックな文書情報フィールドを、それぞれ対応する XMP 構文へ自動的にマップします。こうした結果、開発者が XMP の威力を徹底活用したいと思えば存分にそれが可能な一方で、メタデータ要件が単純な場合であれば PDFlib の自動 XMP 生成で事足りるようになっているのです。

XMP 拡張スキーマ XMP は本質的に拡張可能です。すなわち、会社や業界独自のメタデータ要請は、XMP 拡張スキーマを構築して満たすことができます。PDF/A はこの概念に対応していますが、ただし将来の取得の便宜を図るため、そのスキーマに関する記述を機械読み取り可能な形で、特定の規則に従って文書内に埋め込むことを義務づけています。PDFlib 製品群は、PDF/A のための XMP 拡張スキーマに対応しています (実際、PDFlib 7.0.3 は拡張スキーマ対応をそなえた初の製品でした)。PDFlib は、ユーザーが与える XMP メタデータを、拡張スキーマも含めて検証して、生成出力が PDF/A 完全準拠になることを保証しています。

www.pdflib.com では、PDF/A における XMP に関するもっと詳しい説明や、XMP 拡張スキーマのオンラインバリデータが得られます。

既存の PDF/A 文書进行处理 既存の PDF/A 準拠の文書からページを取り込む際には、従うべき規則が増えます。たとえば Adobe Acrobat では、2 つの PDF/A 準拠の文書を連結する際、いとも簡単に PDF/A 準拠でなくなった出力文書を作り出せてしまいます (警告も一切出ません)。既存の PDF/A 文書を扱う際には、PDFlib+PDI はすべての入力・出力文書の PDF/A 諸特性を注意深く調べて、出力も確実に PDF/A 準拠にしています。さらに細かく制御するため、取り込み文書の出力インテントを出力 PDF へコピーして、既存文書の PDF/A の色特性を引き継ぐことも可能です。

タグ付き PDF を持たせた PDF/A-1a を作成 PDF/A-1a は、PDF/A-1b にタグ付き PDF を加えたものと考えられます。これは文書に対する構造情報を必須とし、また、フォントに関して特定の条件を課して、テキストを適切に解釈できるようにしています。こうした結果、PDF/A-1a 文書は身体障害を持った利用者にとって完全にアクセシブルになっています。PDF/A-1a 文書は、見た目だけでなく中身の意味をも保持しています。

PDFlib の PDF/A-1a に対する対応は、タグ付き PDF を作り出すための機能をベースにしています。そこには、各内容アイテムを、文書の構造ツリー内のそれぞれ特定の場所に位置付けることができる機能がありますし、文書構造と関連しない内容アイテム (ヘッダ・フッタやノンブル等) については、アーティファクトとして、すなわちその文書の再利用時 (文書がソフトウェアで読み上げられる時や、何か他の形式へ変換される時等) には無視されるものとしてタグ付けされる機能もあります。代替テキストを画像に付けることができ、そのテキストは Acrobat によって、目の不自由な利用者のために読み上げられます。

ただし、タグ付き PDF を作るためには、その文書の論理的な構造を詳しくわかっている必要があります。PDFlib は PDF に関連する詳しいことは面倒を見ますが、文書の内容からその構造を推測することはできません。

タグ付き PDF への対応は、PDFlib 6 から導入されています。PDFlib 7 では、文書の構造ツリーに注釈を入れ込んで、リンク等のインタラクティブ要素のアクセシビリティを向上させることのできる機能が追加されています。

すでにタグ付き PDF に対応していたことをベースとして、PDFlib 7 は、PDF/A-1a に準拠した出力を作り出すことができます。だから PDFlib は、この高度な PDF/A レベルに対応した初のツールになったのです。

PDF/A を検証 標準規格に基づいたワークフローを実装する際には、その出力結果を規格にてらして検証するためのツールを投入することが得策です。PDF/A 標準規格についてはいろいろなソフトウェアツールが入手可能で、所与の PDF 文書が ISO 標準に準拠しているかどうかを検査することができます。

PDFlib が作り出す PDF/A は、Adobe Acrobat 8 のプリフライトツールに実装されている検証規則に完全に準拠しています。なお、Acrobat 7 は PDF/A に関する ISO 標準規格の最終リリースよりも以前に発売されたため、その PDF/A 対応はいくつかの点で PDF/A 標準規格を逸脱しており、若干問題があります。ですので Acrobat 7 は、PDF/A の検証には用いるべきではありません。

PDFlib GmbH は、PDF/A 検証ツールの諸ベンダーとアクティブに協働して、PDF/A 標準規格に関する理解と解釈を確実に共有しています。

現時点で PDFlib GmbH は、PDF/A 文書を検証するための製品や、任意の PDF 文書を PDF/A へ変換するための製品は提供していません。

PDFlib PLOP による PDF/A 準拠文書の処理 PDFlib PLOP・PLOP DS 製品は、いくつかの PDF 処理機能を提供していますが、それらはすべて PDF/A 対応になっています。具体的には、PDF/A 入力ファイル进行处理すると、その処理が標準規格に準拠した形で完遂できたならば PDF 準拠の出力が生成され、そうでなければエラーメッセージが出ます。いくつか例を挙げます。

- ▶ PDF/A 文書を暗号化しようとする、暗号化は PDF/A では許されていないので、エラーメッセージが出ます。PDF/A 文書を PLOP で暗号化するためには、明示的に PDF/A 状態を代償にする必要があります。
- ▶ PLOP DS を使って PDF/A 文書に電子署名を適用すると、PDF/A 標準規格に準拠した形で署名フィールドが作成されます。
- ▶ PLOP を使って、既存の PDF/A 文書に XMP メタデータを挿し込むこともできます。PLOP 3.1 は、PDF/A のための XMP 拡張スキーマに対応しているため、既存の文書の後処理工程に使うこともできますし、あるいは文書をはじめに作成するソフトが XMP 拡張スキーマに対応していない場合であっても、これを使って XMP 拡張スキーマを徹底活用することができます。

PDF/A 技術センター PDFlib GmbH は PDF/A 技術センターの創立メンバーの一員です。これは、業界における PDF/A の認知向上と、ベンダー間の互換性の確保を目指す組織です。当社では、PDF/A の諸議題に関する技術ノートを発行している技術作業部会にアクティブに関わりつつ、PDF/A のためのテストスイートの作業も行なっています。詳しくは www.pdfa.org をご覧ください。

PDFlib GmbH
Franziska-Bilek-Weg 9
80339 München, Germany
電話 +49・89・452 33 84-0
info@pdflib.com
www.pdflib.com

