

# ホワイトペーパー： PDFlib 製品群における XMP メタデータ対応

メタデータの重要性 メタデータという用語は、文字どおりには「データに関するデータ」を意味します。電子文書に対して、メタデータはよくその名刺にたとえられます。多くのメタデータは複数のプロパティから成り、各プロパティは文書の文脈で特定の意味を持ちます。よく利用されるメタデータのプロパティを例示します。

- ▶ PDF 文書の作成者。
- ▶ PDF 文書が作成されたり、JPEG 画像がカメラで撮られたりした日付。
- ▶ 画像を撮影した撮影者の名前。
- ▶ パーソナライズされた文書のシリアル番号。
- ▶ 文書に記載されている品目の在庫管理単位 (SKU)。
- ▶ 文書に記載されているエンジニアリング製品の製造年。
- ▶ 裁判における文書参照番号。

出版・文書作成・翻訳等、完全に電子化された形で作業が進められていくワークフローがますます増えていく中において、メタデータは、電子文書の発生から消滅までを管理するうえで重要な役割を担っています。

Adobe の Extensible Metadata Platform (XMP) Adobe は、アプリケーションやファイル形式を超えて使える共通メタデータ形式の必要性を認識し、Extensible Metadata Platform (XMP) を設計しました。これは XML ベースの形式であり、セマンティック Web イニシアチブの根幹をなす、W3C の RDF (Resource Description Framework) に則っています。Adobe は XMP の仕様を無償提供し、オープンソースの XMP ツールキットを、ソフトウェア開発者のために提供しています。

XMP は、ファイルと一体になって持ち運ばれるものであり、PDF・TIFF・JPEG 等、広く利用されている多くのファイル形式への埋め込みが可能です。メタデータのプロパティは、スキーマごとにグループ化されています。各スキーマは一意的な名前空間 URI によって識別され、任意の数のプロパティを持ちます。

XMP の仕様には、さまざまな代表的な種類の文書や画像の特性に合わせて、1 ダース以上の定義済みスキーマが用意され、その中に数百種のプロパティが定められています。最も広く利用されているのは、Dublin Core、略して *dc* という定義済み XMP スキーマです。このスキーマには、*Title*・*Creator*・*Subject*・*Description* といった一般的なプロパティが含まれています。定義済みのスキーマとは別に、カスタムのスキーマを定義して、社内や業界独自のメタデータ要請に対応することも可能です。

PDF 文書での XMP は、2001 年の Acrobat 5・PDF 1.4 から導入されています。PDF における XMP の前任者は、単純なキーと値の対、すなわちいわゆる文書情報項目を集めたただけのもので、XMP の導入前はこれが唯一のメタデータ保持手段として提供されていました。今でも Acrobat や PDF は文書情報項目に対応していますが、XMP メタデータはもっとはるかに強力な概念であり、たとえばスキャンした TIFF を PDF に変換する等、形式を変換してもメタデータを引き継ぐことが可能になっています。

XMP は Adobe の提供する全製品に実装され、それ以外の何ダースものソフトウェアベンダーやユーザーグループもこれに対応しています。Creative Suite に含まれる Adobe Bridge は、さまざまなファイル形式の中で XMP メタデータを扱います。XMP メタデータを「ファイル情報」/「文書プロパティ」パネルで表示・編集できる Adobe アプリケーションとしては、Acrobat (「ファイル」 「プロパティ ...」 「その他のメタデータ ...」)・Photoshop・InDesign 等が挙げられます。なお、「ファイル情報」パネルでは、メタデータのプロパティは定義済み XMP スキーマにしたがってグループ化されていますが、カスタムのパネルを定義して、いろいろな業務分野での要請に合わせて、メタデータの表示や編集可能なフィールドをあつらえることも可能です。

XMP による垂直協業 XMP は、さまざまな業界団体において、個別のメタデータ要求に対応するために利用される度合がますます高まっています。いくつかの例を挙げます。

- ▶ AdsML コンソーシアムでは、広告情報・コンテンツ交換のための仕様とプロセスを作成しています。
- ▶ 国際新聞通信委員会 (IPTC) は、ニュース組織が集まって設立した業界団体です。そこでは、ニュースデータ交換のための業界標準を策定しています。XMP のための「IPTC Core」スキーマを頒布しており、画像等のニュース素材に対するメタデータの伝達に広く利用されています。
- ▶ 医学画像交換のための DICOM 標準規格では、PDF の使用に対応しており、患者データ・研究内容・機材詳細等のメタデータを保持するためのカスタム XMP スキーマを記述しています。
- ▶ Publishing Requirements for Industry Standard Metadata (PRISM) では、雑誌・ニュース・カタログ・書籍・定期刊行物の内容を処理するためのメタデータ語彙を定義しています。

ISO 標準規格が義務づける XMP ISO 規格として、既存のものや発行見込のものがいくつかあり、それぞれグラフィックアート業界・アーカイブ・技術文書等、特定の応用分野に絞った PDF サブセットを記述しています。2001 年と 2002 年に策定されたプリプレス標準規格である PDF/X-1 と X-3 を除けば、PDF に対するすべての ISO 規格に、XMP メタデータの使用が含まれています (ISO 32000 以外の多くの場合はむしろそれが義務づけられています)。

- ▶ PDF/A-1 を標準化した ISO 19005-1 (2005 年発行) : 「長期保存のための電子文書ファイル形式 — PDF 1.4 の利用」。PDF/A-1 では、準拠ファイルを識別するために XMP が必須であり、XMP 拡張スキーマによってカスタムメタデータにも対応しています。カスタムデータが将来なるべく利用できるよう、拡張スキーマはすべて、それに関する記述を PDF/A 内に、決められた形で埋め込まなければなりません。PDF/A-1 では文書情報項目の使用も許されていますが、XMP だけでワークフローが組み立てられるよう、よく使われる PDF 文書情報項目はそれぞれ特定の定義済み XMP プロパティと同期させておかねばなりません。文書情報項目と XMP プロパティとの間のこの「橋渡し」は、この規格で定義されています。PDF/A-1 における XMP 対応は、XMP 2004 の仕様に基づいています。
- ▶ PDF/E を標準化した ISO 24517-1 (2008 年発行見込) : 「PDF を用いた技術文書形式 — PDF 1.6 の利用」。PDF/E における XMP 対応は PDF/A-1 とほぼ同等ですが、ただし、より新しい XMP 2005 の仕様に基づいています。
- ▶ PDF/X-4 を標準化した ISO 15930-7 (2008 年発行) : 「PDF 1.6 を用いた印刷データの完全な交換 (PDF/X-4) および外部プロファイル参照を持った印刷データの部分的交換 (PDF/X-4p)」。PDF/A-1 と同様、PDF/X-4 でも規格準拠を表すために XMP が必須です。文書情報項目は PDF/X-4 では使用可能ですが、それぞれ対応する XMP 項目と同期させておかねばなりません。XMP 拡張スキーマを使ってカスタムメタデータを表すことは許されます。ただし PDF/A-1 と違って、決められた形で説明を埋め込まなくても使えます。PDF/X-4 における XMP 対応は、XMP 2005 の仕様に基づいています。
- ▶ PDF/X-2 を標準化した ISO 15930-5 (2003 年発行) および PDF/X-5 を標準化した ISO 15930-8 (2008 年発行見込) : 「PDF 1.6 を用いた印刷データの部分的交換 (PDF/X-5)」。PDF/X-2・X-5 の文書は他の PDF/X 文書を参照し、そうした参照の参照先は、さまざまな XMP 項目を用いて識別されます。このため XMP は、PDF/X-2・X-5 では不可欠の構成要素です。
- ▶ ISO 32000 (2008 年発行見込) : 「文書管理 — 可搬文書形式 — PDF 1.7」。ISO 32000 は PDF 1.7 を標準化したものです。その技術的な中身は PDF 1.7 (Acrobat 8 のファイル形式) と等価であり、XMP メタデータに完全対応しています。

最も広く利用されている定義済み XMP メタデータスキーマの一つである Dublin Core は、ISO 15836 (2003 年発行) : 「情報とドキュメンテーション — Dublin Core メタデータ要素セット」として標準化されています。



PDFlib 製品スイートにおける XMP 対応 シンプルな XMP 対応は、2004 年に PDFlib 製品ファミリに導入されています。PDFlib 7 (2006 年発売) が PDF/A-1 に対応したことに伴い、XMP のさまざまな機能が、PDF/A-1 のいろいろな要請に合わせて拡張されました。とくに、文書情報項目を XMP プロパティに自動同期させる (PDF/A-1 の橋渡し仕様の通りに) 機能や、PDF/A-1 に必須のいくつかの内部 XMP プロパティを自動生成する機能が実装されました。その結果、PDFlib のユーザーは、PDF/A-1 のための XMP を、XMP 形式の中身にかかわらずわされずに生成できるようになっています。高度なユーザーなら、定義済み XMP メタデータスキーマを直接まるごと PDFlib に与え、生成 PDF 文書に取り込ませることもできます。PDFlib はあらゆる主要なオペレーティングシステムで利用できるうえ、サードパーティ製品を一切必要としないので、これを使えばどのプラットフォームでも XMP 対応が可能です。

これに加えて PDFlib 7.0.3 では、PDF/A-1 に従って XMP 拡張スキーマにも対応しました。ユーザーは PDF/A-1 の中に、カスタムメタデータのための「拡張スキーマコンテンツスキーマ」を埋め込むことができます。PDFlib は、ユーザーから与えられた XMP 拡張スキーマが内部的に整合しているか、規格に準拠しているかを完全検証しますので、その出力は PDF/A-1 標準規格に準拠していることが保証されています。

この機能によって PDFlib 7.0.3 は、世界で初めて PDF/A-1 のための XMP 拡張スキーマに対応した製品となりました。PDFlib GmbH は PDF/A 技術センターに参画していますので、PDF/A に関するすべての活動は、PDF/A ソフトウェアを出している他の諸ベンダーとの緊密な連携によって、望みうる最高度の標準規格準拠と業界慣習反映を実現しています。

XMP 検証機能は、PDF/A 出力を作成しないときでも有効なので、PDFlib 7.0.3 における XMP 対応の改良の恩恵は、XMP を利用するすべての人に及びます。

[www.pdflib.com](http://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。

PDFlib PLOP 3.1 で XMP を PDF に入れ込む 暗号化・復号・最適化・電子署名といった機能に加えて、PDFlib PLOP では、既存の PDF 文書に XMP メタデータを入れ込むことも可能です。この機能は、既存の PDF 文書の中に、必要なすべてのメタデータプロパティが入っていないような場合に便利です。PLOP における XMP 対応は PDF/A 対応になっているので、とくに PDF/A ワークフローではこれは有用です。たとえば、拡張スキーマを持たせたカスタム XMP を、拡張スキーマに対応していないワークフロー由来の PDF/A 文書に入れ込むことができます。

PDFlib pCOS で XMP を PDF から抽出 pCOS インタフェースとは、あらゆる種類の情報を PDF 文書から取り出すことのできる、PDFlib GmbH の方式です。これはスタンドアロン製品としても利用できますし、他のすべての製品にも内蔵されています。pCOS は、XMP メタデータを PDF 文書から抽出するためのシンプルなプログラミング手段を提供します。XMP メタデータは Unicode に規格化されるので、利用者はエンコーディングの問題に悩まされずに済みます。

XMP の取得は、圧縮や暗号化・PDF オブジェクト構造に関係なく動作します。Adobe の定めている XMP のパッケージのしくみでは、XMP のデータパッケージはさまざまなファイル形式に簡単に埋め込めて簡単に取り出せるようになっているのですが、PDF の文書ではちょっと面倒な現象がいくつかあって、話がややこしくなっています。たとえば、PDF 文書は更新セクションをいくつか含むことがあるので、その場合は 1 つの XMP ストリームがファイル内の複数箇所にだぶって現れてしまうのですが、そのうち正しいのは 1 個だけです。単純なテキスト検索で XMP ブロックを検索しても、間違った箇所を取って来かねませんので、PDF のオブジェクト構造を注意深くたどっていくソフトウェアだけが、どんな場合でも正しく XMP メタデータを取得することができます。だから Adobe の無償の XMP Toolkit は、TIFF や JPEG といった他のファイル形式の中の XMP には対応しているのに、PDF からの XMP 抽出には完全対応していないのです。

PDFlib TET PDF IFilter で XMP メタデータを検索 TET PDF IFilter は、PDFlib GmbH から発売された最新の製品です。これは Microsoft の IFilter インタフェースを実装し

ており、Microsoft やサードパーティのさまざまな、Windows Desktop Search (WDS)・Office SharePoint Server (MOSS)・Indexing Server・SQL Server といったデスクトップ・組織内検索製品とともに使うことができます。TET PDF IFilter の XMP 対応を利用すると、Microsoft の検索ソリューションが投入されている環境で、非常に簡単に XMP メタデータを活用することができます。

TET PDF IFilter の高度なメタデータ実装は、メタデータのための Windows プロパティシステムに対応しています。ページ内容だけでなく、XMP メタデータや、標準・カスタム文書情報項目もインデックスします。メタデータのインデクシングは、以下のいずれかのレベルに設定できます。

- ▶ 文書情報項目と、よく用いられる XMP プロパティを、標準の Windows プロパティ (タイトル・サブジェクト・作者等) にマップする。
- ▶ 有用な PDF 特有の擬似プロパティ (ページサイズ・PDF/A 準拠レベル・フォントリスト等) を TET PDF IFilter が追加する。
- ▶ すべての定義済み XMP プロパティ (*dc:rights*・*xmpRights:UsageTerms*・*xmp:Creator Tool* 等) を検索可能にする。
- ▶ カスタム (ユーザー定義) の XMP プロパティ (社内独自分類項目等) を検索可能にする。

TET PDF IFilter は設定によって、メタデータを生テキストインデックスに含めさせることもできます。そうすると、メタデータ対応機能のない全文テキスト検索エンジン (SQL Server 等) でもメタデータの検索が可能になります。

XMP ベースの文書検索で恩恵を受けるワークフローシナリオ XMP メタデータの処理は、電子文書の検索を必要とするさまざまなシナリオに組み込むことができます。以下に典型的な例を 2 つ挙げます。

出版：クリエイティブのプロフェッショナルたちは、Adobe 等のパブリッシングソフトウェアを使って、文書やメタデータをインタラクティブに作っています。彼らは文書に、キーワード・作成者名・著作権情報等、広く利用される XMP プロパティを割り当てます。彼らは Adobe Bridge を使って、割り当てられているメタデータプロパティに従って文書を検索したりグループ化したりすることができ、Dublin Core や IPTC といった広く利用されている XMP スキーマに関心を絞っています。

技術文書：大量の文書が手作業や自動で作成され、部課や社内のコレクションに集められています。こうした文書コレクションは、サーバシステム上では Microsoft Office SharePoint Server (MOSS)、ワークステーション上では Windows Desktop Search (WDS) 等の抽出製品等、広く利用されている Windows 抽出ツールでアクセスされます。こうした製品に TET PDF IFilter を組み合わせると、利用者は文書群に対して、XMP メタデータのプロパティでも、ページの実内容でも検索をかけられるようになります。定義済み XMP スキーマでも基本的な需要は満たしていますが、社内独自の要請を満たすため、カスタマイズした XMP スキーマをクエリで使うことも可能です。

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

