PDF/A Metadata – XMP, RDF & Dublin Core

This article deals with the following topic areas:
What is Metadata?
Why Metadata?
Metadata in PDF
All about XMP
How to get metadata into PDF?

What is Metadata?

"Data about Data"

Metadata is data about data. It describes the content, character, provisos, and other characteristics of data. The term 'metadata' was coined in the 1960s by Jack Myers.

Metadata is everywhere

Metadata helps us to understand the world around us. Without it, we are lost. Metadata also exists outside the computing world, on plans and maps, for example.

A simple example of map metadata.

Why Metadata?

There are many reasons for using metadata. Companies can benefit from using it when they set up and maintain their databases. If time is spent on the intelligent management of document data at the document creation stage, this initial extra effort is sure to pay off in the long term.

For example, the accounting center can access this metadata later on and the transmission of data may be easier. Search processes can also be significantly improved if uniform concepts are used.

Uses for Metadata
Metadata helps to provide an optimum overview of data and ideal data handling in many areas. This includes the following examples:

Identification
- Title
- Area Covered
- Themes Responsible party
- Providence – Where did it come from?

Distribution
- Distributor
- Formats
- Media
- Online
- Price

Entity and Attribute Information
- Features
- Attributes
- Attribute Values

And many, many more...

Benefits for Internal Data

Metadata minimizes the work effort required for many tasks in companies and authorities. Employees can find the data records they require more quickly and do not have to look for additional information on certain data elsewhere. With regard to personnel fluctuations, metadata prevents the loss of information that might otherwise be mislaid. Metadata also provides many automation possibilities that could not be realized without it. Further time savings result from the fact that metadata avoids data record duplicates within an organization.
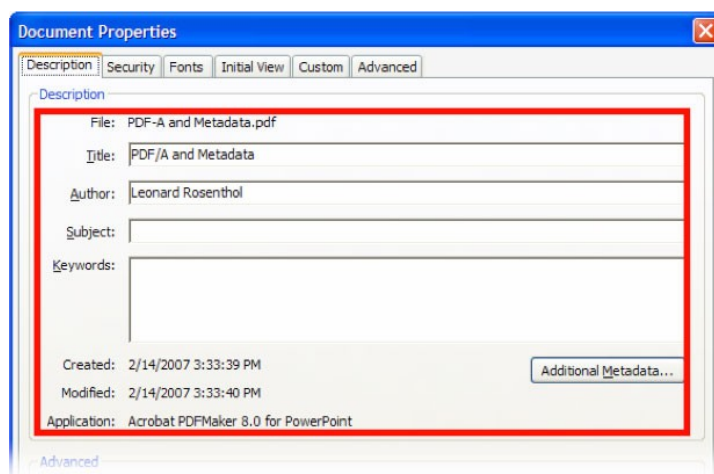
Benefits for External Data

Correct metadata simplifies the entry of transactions at the accounting office. Another plus point is that a greater number of people can utilize the data. This saves time because data record duplicates at different branches can be avoided.

Metadata in PDF

How is metadata organized in PDF? In Adobe Acrobat, you can view and change metadata in the 'Document Properties' dialog box.

Info Dictionary

The info dictionary has been a part of PDF since PDF version 1.0. This area belongs to the document itself and contains a collection of name/value pairs. Predefined pairs include *Title, Author, Subject, Keywords*, and others. You can also add your own values

Document Properties in Adobe Acrobat
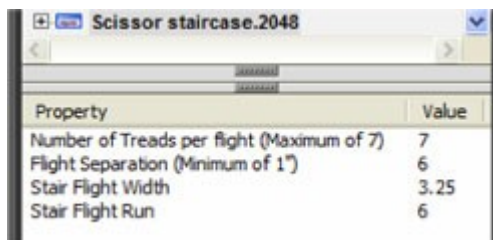
PieceInfo Dictionary

The PieceInfo dictionary was introduced with PDF 1.3 (Acrobat 4). This dictionary is either document- or page-related. The Application Private Data section is used by Adobe Photoshop and Adobe Illustrator.

XMP (Extensible Metadata Platform)

XMP is a more recent development. It was introduced with PDF 1.4 (Acrobat 5). XMP is based on RDF (Resource Definition Framework). RDF is a W3C standard for XMP-based metadata (for more information, see www.w3.org/RDF). XMP can be linked with XObjects on document pages. XObjects are also known as images and repeating objects. In addition, XMP can be linked with fonts and ICC profiles.

Object Data (also called User Properties)

Object data was introduced with Acrobat 7 and is based on PDF 1.6. This area is connected with individual content elements. The name/value pairs can be realized using character strings, digits, or logical links.
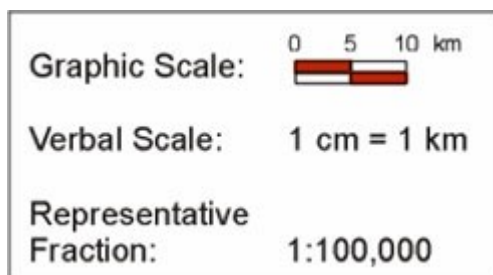

Example of object data

Measurement Properties

Measurement properties have also been permitted since PDF 1.6. These properties are page-related and supply information on sizes and units of measure. This allows PDF units to be linked with units from the real world (for example, 1cm = 1km).

Measurement properties have been permitted since PDF 1.6

All about XMP

XMP Overview

XMP (Extensible Metadata Platform) enables metadata to be integrated into all relevant Adobe applications (and in other places) using a uniform schema.

EXtensible

▹ Easily add new metadata properties
▹ Extend the CS application UI
▹ Support from 30+ major asset management vendors

Metadata

▹ Data about...data
▹ XML based on the W3C standard for encoding metadata (RDF)
▹ create "smart assets" by embedding XML metadata in binary files
▹ File format neutral, e.g., JPEG, TIFF, EPS, PDF, Adobe native
▹ Metadata schema neutral (customizable)

Platform

▹ Across all CS products
▹ Based on industry W3C standards
▹ Freely available via open source → partners
▹ intelligent / automation workflows; improved productivity

XMP Benefits

Business		Technical	
Promotes Intelligent Media	Fosters re-use, re-purposing, re-expression across domains Promotes brand equity, intellectual capital & other intangible asset	**Self Describing**	Not limited to a specific schemaEvery file can have meta data
Open Platform	Enables metadata capture, preservation & propagation,across devices, applications, file formats, institutions	**Accessible**	Based on industry standards (W3C)Openly available Extends metadata beyond the context of a database
Intelligent media based on XMP			

Key Elements of XMP
- **Framework:** XML structure for storing information
- **Data package:** How and where to store and call up information
- **Specification:** Description of the standard and its relationship to other standards
- **XMP toolkit (SDK):** Available free of charge, open source
- **Modifiable fields:** User interface for user interaction with metadata
- **Platform:** Standardized access to metadata throughout Adobe CS

A Closer look at XMP
XMP is based on the W3C standard. The Adobe metadata framework constitutes the first ever wide-scale, comprehensive, and practical use of RDF (Resource Description Format). The Adobe XMP Platform has the following elements:

- **XMP framework:** RDF framework or rendering metadata from multiple schemas
- **XMP schema:** Schema for the description of properties, contained in namespaces
- **XMP data package technology:** Method for embedding XML fragments in binary code
- **XMP SDK:** Support for third-party solutions (interface and enhancements)

XMP Schemas

RDF specifies IDs in XML sequences, structured by *source, property,* and *value* (or alternatively *subject, predicate,* and *object*). RDF schemas define vocabulary. Adobe designed the *standard XMP schemas.* The XMP framework permits the integration of any schema that is structured in accordance with the specification. Area-specific schemas (such as *IPTC* and *NewsML*) can also be described in XMP data packages.

Property	Value Type	Description
BitRate	Integer	Bits per second
Dimensions	Dimensions	Size of playback view rectangle
Interleaved	Boolean	If true, NTSC fields, otherwise frames
NaturalRate	Real number	Fields/Frames per second

An example of a XMP Schema for Video.

XMP Basic Schema

Property	Value Type	Category	Description
xmp:CreateDate	Date	Internal	The date and time the resource was originally created.
xmp:CreatorTool	AgentName	Internal	The name of the first known tool used to create the resource. If history is present in the metadata, this value should be equivalent to that of xmpMM:History's softwareAgent property.
xmp:MetadataDate	Date	Internal	The date and time that any metadata for this resource was last changed. It should be the same as or more recent than xmp:ModifyDate.
xmp:ModifyDate	Date	Internal	The date and time the resource was last modified. NOTE:The value of this property is not necessarily the same as the file's system modification date because it is set before the file is saved.

The XMP Basic schema provides properties that provide basic descriptive information

- ▹ The schema namespace URI is http://ns.adobe.com/xmp/1.0/
- ▹ The preferred schema namespace prefix is *xmp*.

Adobe PDF Schema

Property	Value Type	Category	Description
pdf:Keywords	Text	External	Keywords.
pdf:PDFVersion	Text	Internal	The PDF file version (for example: 1.0, 1.3, and so on).
pdf:Producer	AgentName	Internal	The name of the tool that created the PDF document.

The Adobe PDF schema provides a set of properties used with Adobe PDF documents.

▹ The schema namespace URI is http://ns.adobe.com/pdf/1.3/
▹ The preferred schema namespace prefix is pdf

Dublin Core Metadata Initiative – dublincore.org/

The Dublin Core metadata element set (also known simply as *Dublin Core*) is a vocabulary with fifteen properties that describe the properties of the source. Dublin Core is part of a larger set that consists of metadata vocabulary and technical specifications, overseen by the *Dublin Core Metadata Initiative (DCMI)*.

The complete set of vocabularies, *DCMI Metadata Terms [DCMI TERMS]*, also contains a set of source categories (resource classes) – the *DCMI Type Vocabulary [DCMI-TYPE]*.

The conditions in the DCMI vocabularies are intended to be used in conjunction with other, compatible vocabularies and in combination with application profiles, on the basis of the *DCMI Abstract Model [DCAM]*.

The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio. "Core" because its elements are broad and generic, usable for describing a wide range of resources.

Dublin Core (dc) Schema

Property	Value Type	Category	Description
dc:contributor	bag ProperName	External	Contributors to the resource (other than the authors).
dc:coverage	Text	External	The extent or scope of the resource.
dc:creator	seq ProperName	External	The authors of the resource (listed in order of precedence, if significant).
dc:date	seq Date	External	Date(s) that something interesting happened to the resource.
dc:description	Lang Alt	External	A textual description of the content of the resource. Multiple values may be present for different languages.
dc:format	MIMEType	Internal	The file format used when saving the resource. Tools and applications should set this property to the save format of the data. It may include appropriate qualifiers.
dc:identifier	Text	External	Unique identifier of the resource.
dc:language	bag Locale	Internal	An unordered array specifying the languages used in the resource.
dc:publisher	bag ProperName	External	Publishers.
dc:relation	bag Text	External	Relationships to other documents.
dc:rights	Lang Alt	External	Informal rights statement, selected by language.
dc:source	Text	External	Unique identifier of the work from which this resource was derived.
dc:subject	bag Text	External	An unordered array of descriptive phrases or keywords that specify the topic of the content of the resource.
dc:title	Lang Alt	External	The title of the document, or the name given to the resource. Typically, it will be a name by which the resource is formally known.
dc:type	bag open Choice	External	A document type; for example, novel, poem, or working.

DocInfo -> XMP Crosswalk

The table shows how entries and properties from the DocInfo and XMP areas relate to each other and can be translated.

Document information dictionary		XMP	
Entry	PDF type	Property	XMP type
Title	text string	dc:title	Lang Alt
Author	text string	dc:creator	seq ProperName
Subject	text string	dc:description["x-default"]	bag Text
Keywords	text string	pdf:Keywords	Text
Creator	text string	xmp:CreatorTool	AgentName
Producer	text string	pdf:Producer	AgentNam
CreationDate	date	xmp:CreateDate	Date
ModDate	date	xmp:ModifyDate	Date

Other Schemas supported in Adobe Creative Suite
- Basic XMP *
- Dublin Core *
- Rights Management *
- Media Management *
- Adobe PDF *

* Supported by Acrobat & PDF

- Photoshop (IPTC subset)
- EXIF
- Job Ticket Management
- XMP Page Text
- Camera Raw – metadata edits
- Photoshop History
- Audio & Video
- Stock Photos

Custom Schemas

The schemas defined in this document are core schemas that are believed to be applicable to a wide variety of needs. If possible, it is always desirable to use properties from existing schemas. However, XMP was designed to be easily extensible by the addition of custom schemas. If your metadata needs are not already covered by the core schemas, you can define and use your own schemas.

If you are considering creating a new namespace, observe the following:

Avoid including properties that have the same semantics as properties in existing namespaces.

If your properties might be useful to others, try to collaborate in creating a common namespace, to avoid having a multitude of incompatible ones. To define a new schema, you should write a human-readable schema specification document. The specification document should be made available to any developers who need to write code that understands your metadata.

NOTE: Future versions of XMP might include support for machine-readable schema specifications, but such support will always be in addition to the requirement for human-readable schema specification documents.

How to get metadata into PDF?

There is a range of solutions that can be used to automatically or manually add metadata to PDF. This overview shows some of the possibilities:

Info Dictionary		XMP	
Manuell	**Automatisch**	**Manuell**	**Automatisch**
Adobe Acrobat Standard/Professional	Acrobat SDK	Adobe Acrobat Standard/Professional	Acrobat SDK
Other 3rd party PDF viewers	Adobe PDFLibrary	Apago's PDF Enhancer	Adobe PDFLibrary
	of 3rd party	PdfLicenseManager	Adobe XMPToolkit
	Libraries and Tools		Apago's PDF Enhancer
			iText

Related Resources

Adobe Acrobat
- **Acrobat:** http://www.adobe.com/products/acrobat/
- **Acrobat SDK:** http://www.adobe.com/devnet/acrobat/
- **Adobe PDFLibrary:** http://www.adobe.com/devnet/pdf/library/

XMP
- **XMP Website:** http://www.adobe.com/xmp
- **XMP Toolkit 4.0 Labs:** http://labs.adobe.com/xmp

Other
- **PdfLicenseManager:** http://creativecommons.org/weblog/entry/6054
- **PDF Enhancer:** http://www.apagoinc.com/prod_home.php?prod_id=2
- **iText:** http://www.lowagie.com/iText

Leonard Rosenthol, Adobe/aoe