

Dokument-Formate in DMS-Systemen

1. Einführung

Elektronische Dokumente haben heute in vielen Bereichen das klassische Papierdokument abgelöst. Auch im professionellen Dokumenten-Management beziehungsweise bei der Archivierung von Dokumenten ist dies zunehmend der Fall. Nach der Entscheidung für ein - vollständiges oder partielles - digitales Dokumenten-Management ist eine grundsätzliche Entscheidung über die Datenformate zu treffen, die künftig im Rahmen des DMS verwendet werden sollen. Häufig wird für Archivierungszwecke in diesem Zusammenhang neben dem lange Zeit üblichen Bildformat TIFF das PDF- beziehungsweise PDF/A-Format empfohlen. Mit dem von Microsoft entwickelten XPS-Format, das wie Adobes Portable Document Format beliebige druckfähige Dateien in ein einheitliches universelles Format überführen kann, kommt ein weiteres mögliches Zielformat für archivierte Dokumente hinzu.

Eine Unterform digitaler Dokumente sind elektronische Formulare. Um dort eingetragene Daten in einem DMS komfortabel verwalten und beim Einsatz von dynamischen Formularen professionell weiterverarbeiten zu können, sollten diese in XFA (XML Forms Architecture), einem XML-basierten Format, abgelegt werden. XFA-Formulardaten können in einer PDF-Datei oder als eigenständige XDP-Datei gespeichert werden.

2. PDF und PDF/A

Das PDF/A-Format basiert auf dem von Adobe entwickelten Dateiformat PDF, das seit 1995 eine immer größere Verbreitung erfahren hat, durch die Internationale Organisation für Normung (ISO) zum offiziellen Standard (aktueller ISO-Standard für das PDF-Format: ISO-32000) erhoben wurden, und so schließlich auf globaler Ebene zum Standardformat in der dokumentbasierten elektronischen Kommunikation und Zusammenarbeit herangewachsen ist. Das PDF-Format wird von Adobe laufend weiterentwickelt, die jeweils aktuellen Regeln und technischen Hintergründe zu seiner Implementierung waren vor der Erhebung zum ISO-Standard in einer sogenannten PDF-Format-Spezifikation festgelegt. Seitdem PDF ein ISO-Standard ist, werden in unregelmäßigen Abständen Ergänzungen als Supplements (aktuell als Extension Levels 3 und 5) durch Adobe veröffentlicht. Mit der PDF Association (pdfa.org), einem Zusammenschluss weltweit führender Unternehmen und Fachleute im Bereich PDF-Technologie, soll der Informations- und Erfahrungsaustausch auf den Gebieten PDF und Langzeitarchivierung gemäß ISO 19005: PDF/A gefördert werden.

Die Vorteile des PDF-Formats sind unter anderem:

- Unabhängigkeit vom Ursprungsprogramm, mit dem eine Datei erstellt wurde und von der Rechnerplattform (Betriebssysteme Windows, MAC und Linux)
- Unabhängigkeit von der verwendeten Hardware: Lesbarkeit auf PC, PDA, Smartphone oder E-Book
- Dateien sind nach der Erstellung sofort durchsuchbar. Bilddateien (zum Beispiel im TIFF-Format) müssen zuerst mit Texterkennungs-/OCR-Software weiterverarbeitet werden
- Integration rechtsgültiger elektronischer Signaturen (Unterschriften) möglich
- Zugriffs- und Bearbeitungsrechte können vergeben oder eingeschränkt werden (Schutz des Dokuments vor unberechtigtem Zugriff und vor Veränderung)
- Einheitliche Visualisierung der Inhalte
- PDF-Dateien sind komprimiert und haben deshalb üblicherweise einen Bruchteil der Größe des Ursprungsformats oder von Bilddateien

PDF/A ist ein Unterformat von PDF, das zum Standard für einen speziellen Verwendungszweck von PDF-Dateien erhoben wurde, nämlich für die regelkonforme, elektronische Langzeitarchivierung von Dokumenten. Mit seiner Hilfe sollen Archivierungsmethoden für elektronische Dokumente optimiert werden. Es ist ein Baustein für eine optimale Archivierungsstrategie, in die daneben jeweils die spezifischen Gegebenheiten des Umfelds (Unternehmensstandards, Anforderungen, eingesetzte Applikationen und Datenquellen) einbezogen werden müssen. Diesen Standard hat die ISO am 28. September 2005 veröffentlicht und dafür die offizielle Bezeichnung „ISO 19005-1:2005 - Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 (PDF/A-1)“ vergeben. Zurzeit wird in den ISO-Gremien an einer aktualisierten Version des Standards gearbeitet: ISO 19005-2 (PDF/A-2). PDF/A-2 soll dem erweiterten Funktionsumfang der PDF-Referenz 1.7 Rechnung tragen. Der neue Standard wird nicht mehr auf der PDF-Referenz von Adobe basieren, sondern auf ISO Norm 32000-1 (PDF 1.7).

Dateien, die dem PDF/A-Standard entsprechend erstellt werden, sollen elektronische Inhalte so darstellen, dass ihr Erscheinungsbild im Betrachtungsprogramm (PDF-Viewer) über eine lange Zeit immer gleich bleibt. Dies soll unabhängig von der Hardware und Software möglich sein, mit der die PDF-Datei erstellt, gespeichert und angezeigt wird. Aus diesem Grund dürfen in einer PDF/A-konformen Datei auch keine Verweise auf externe Quellen wie Bilder, Texte oder nicht eingebettete Schriften enthalten sein und eine Verschlüsselung ist nicht zulässig. Im Ergebnis legt der PDF/A-

Standard deshalb hinsichtlich der Bildschirm- und Druckausgabe fest, welche der gemäß der PDF-Referenz grundsätzlich möglichen Eigenschaften einer PDF-Datei mit Blick auf das Ziel der Langzeit-Archivierung notwendig, optional möglich, nicht erlaubt oder eingeschränkt erlaubt sein sollen.

Anforderungen an PDF/A-konforme Dokumente im Detail:

Vollständigkeit

Bezüge (zum Beispiel Links auf externe Inhalte) und Ressourcen (zum Beispiel die benutzten Schriftarten und Bilder), die nicht in der PDF-Datei selbst enthalten sind und deren Zugänglichkeit deshalb außerhalb der Kontrolle der Datei liegt, führen dazu, dass die Datei als nicht PDF/A-konform eingestuft wird. In einer PDF/A-1-Datei muss eine Kennzeichnung als PDF/A-1 in Form von Metadaten nach dem Extensible Metadata Platform-(XMP)-Standard eingebettet sein.

Eindeutigkeit

Farben müssen in Farb-Profilen (Quellprofile, ICC-Profile) so definiert sein, dass eine eindeutige Farbdarstellung möglich ist. Ein Beispiel ist sRGB, der Standardfarbraum für die Monitoranzeige unter Windows. Zudem müssen die Codierungsinformationen für Fonts eindeutig sein. Und schließlich sind alternative Bilder, also eine Variante mit niedriger Auflösung für die Bildschirmanzeige und eine Variante mit hoher Auflösung für den Ausdruck auf Papier, nicht zulässig.

Zugänglichkeit

Die Verschlüsselung von PDF-Dateien und das Sperren von Funktionen zum Drucken und zum Kopieren von Texten und Bildern sind in PDF/A-konformen Dateien nicht zulässig.

Verzicht auf kontextabhängige oder dynamische Funktionen

JavaScript und Aktionen dürfen nicht in PDF/A-Dateien eingebunden werden, da ihre Ausführung den Inhalt und/oder die Darstellung der PDF-Datei verändern oder beeinflussen kann. Auch Audio- oder Videodateien dürfen nicht eingebettet werden.

Ebenen und Transparenzen

Ebenen (PDF-Layer) und transparente Objekte sind nicht erlaubt, wenn eine Datei PDF/A-konform sein soll.

3. Ein Alternativformat zu PDF: XPS

Das [XPS \(XML Paper Specification\)-Format](#) wurde im Jahr 2003 von Microsoft eingeführt und 2009 als ECMA-388-Standard veröffentlicht. In Windows 7 und Vista sind standardmäßig ein XPS-Drucker (Microsoft XPS-Dokument-Generator) und ein XPS-Betrachter integriert. Die Verbreitung von XPS gegenüber PDF ist gegenwärtig noch deutlich geringer.

4. Bildformate (TIF, PJG)

Wenn man zu archivierende Dokumente im TIFF- oder JPG-Format, also als Bilddateien, ablegt, stehen die in ihnen enthaltenen Texte, anders als beim PDF-Format, zunächst nicht als separate Objekte zur Verfügung. Sie können also nicht durchsucht werden. Das Durchsuchen ist erst nach dem Einsatz von OCR (Optical Character Recognition) möglich. Aus diesem Grund bietet es sich an, bereits in anderen digitalen Formaten vorliegende Dokumente sofort in PDF beziehungsweise PDF/A umzuwandeln. Dokumente, die im Original in Papierform vorliegen, sollten aus Gründen der Einheitlichkeit nach dem Einscannen genauso abgelegt werden. Einige weitere Vorteile der Archivierung in PDF (PDF/A) sind: geringerer Speicherplatz, Speicherbarkeit von Metadaten wie Titel, Autor, Erstellungsdatum, Änderungsdatum, Thema, Schlüsselwörter.

5. Datenverwaltung in dynamischen Formularen: XFA

Das XFA-Format wurde ursprünglich von dem Unternehmen JetForm für die Verwendung in Web-Formularen entwickelt. JetForm ist im Jahr 2002 von Adobe aufgekauft worden. XFA wird heute auch für (dynamische) PDF-Formulare eingesetzt. Aus technischer Sicht ist es allerdings vollkommen unabhängig von PDF und PDF/A. XFA ist bisher kein offizieller, internationaler Standard. Es ist relativ einfach in DMS zu implementieren und bietet breite Möglichkeiten zur Weiterverarbeitung von Formulardaten.

6. Anwendungsbereiche

Einige mögliche [Anwendungsbereiche für die beschriebenen Archivformate](#) sind: allgemeine Archivsysteme als Überführung eines Papierarchivs in ein elektronisches Archiv, spezielle E-Mail-, Fax- und Berichte-Archive, Umwandlung jeglicher im Tagesgeschäft anfallenden elektronischen Korrespondenz (aus Textverarbeitung, Tabellenkalkulation, Präsentationen etc.). Speziell für das PDF- und PDF/A-Format sind zusätzlich zum Beispiel denkbar: auf Vorlagen basierende dynamische PDF-Erstellung. Und im PDF-, PDF/A -und XPS-Format lassen sich elektronische Rechnungen auch als signierte Dateien erstellen.

7. Ein Lösungsanbieter: Soft Xpansion

Das Unternehmen Soft Xpansion (www.soft-xpansion.de) bietet eine Reihe von Produkten zur Anzeige, Erstellung und Bearbeitung von PDF-, PDF/A - und XPS-Dateien sowie zum Einsatz des XFA-Formats. Dabei handelt es sich zum einen um Standardsoftware, die [Perfect PDF 7-Produktfamilie](#), sowie eine Lösung zur serverbasierten PDF-Erstellung, das [PDF Render Center](#). Zum anderen werden [Werkzeuge für Entwickler](#) von Dokumentenmanagement-Software angeboten. Soft Xpansion hat im Oktober 2011 die Version 8 seines PDF Xpansion Software Development Kit veröffentlicht.



Der Test bezieht sich auf Perfect PDF 7 Premium



Die Tests beziehen sich auf Perfect PDF 6 Premium



Die Tests beziehen sich auf Perfect PDF 5 Premium