






PDF/A – der Standard für die Langzeitarchivierung



- Was spricht für den PDF/A Standard?
- Was ist der PDF/A Standard?
- Was bedeuten PDF/A-1a, PDF/A-1b, PDF/A2?
- Wie wird der PDF/A Standard eingesetzt?
- Ist PDF/A die Lösung für Langzeitarchivierung?

Inhaltsverzeichnis

| | | |
|---|--|---|
| | Inhaltsverzeichnis | 2 |
|  | Hintergrund | 3 |
| | Einleitung | 3 |
| | Was spricht für den PDF/A Standard? | 3 |
|  | Der PDF/A Standard | 5 |
| | Die Ziele von PDF/A | 5 |
| | Was PDF von PDF/A unterscheidet | 5 |
| | PDF/A, A-1a, A-1b, A-2, „Babylon“ | 5 |
|  | Nutzung des PDF/A Standards | 7 |
| | Woher bekomme ich ein Exemplar? | 7 |
| | An wen richtet sich die Norm? | 7 |
| | Welche Tools sind erhältlich? | 7 |
| | PDF/A als Bestandteil eines umfassenden Konzepts für Langzeitarchivierung. | 7 |
|  | Fazit | 8 |
| | PDF/A – der Archivierungsstandard | 8 |
| | Wie reagiert der Markt? | 8 |
| | PDF/A als langfristige Strategie | 8 |
|  | Informationen | 9 |
| | PDF/A Competence Center | 9 |
| | PDF Tools AG | 9 |





Hintergrund

PDF/A – der Standard für Langzeitarchivierung

Einleitung

Die Internationale Organisation für Normung (ISO) hat am 28. September 2005 einen neuen Standard für das regelkonforme Archivieren elektronischer Dokumente definiert – die offizielle Bezeichnung lautet:

ISO-19005-1 - Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A-1).

Der Standard resultiert aus einer mehr als 36 Monate dauernden Zusammenarbeit zwischen Unternehmen und Organisationen auf der ganzen Welt.

Der Startschuss für diese Initiative fiel im Mai 2002 in den USA. Erklärtes Ziel war die Erschaffung eines standardisierten Formats für elektronisch archivierte Dokumente. Beteiligt waren die AIIM (Association for Information and Image Management), die NPES (National Printing Equipment Association) und die Verwaltung der US-amerikanischen Gerichte. Im Oktober 2002 fand das Kick-off Meeting statt. Bekannte PDF Hersteller nahmen daran teil: Adobe Systems, Library of Congress, Surety Inc., Quality Associates Inc., Appligent, Merck, EMC, PDF Sages, und NARA (National Archives & Records Administration). Später kamen unter anderem Xerox, Honeywell, EDS und Glaxo Smith Kline hinzu.

Die Gründer des Projekts stellten eine erste Version zusammen und reichten ihren Vorschlag bei der ISO ein, um diesen als internationalen Standard registrieren zu lassen. Das Projekt wurde von der ISO dem technischen Komitee TC 171 (Document Management Applications) zugeteilt. Dieses Komitee ist zusammengesetzt aus 15 Mitgliedsstaaten, deren Vertreter jeweils eine Stimme haben. Ergänzt wird das Komitee von einer Beraterkommission, welche weitere 21 Länder repräsentiert. Der Standard wurde in mehreren Etappen verbessert, bis er im September 2005 endgültig freigegeben wurde.

2005 wird PDF/A zum internationalen Standard erhoben.

Namhafte Organisationen und Hersteller sowie professionelle Anwender waren daran beteiligt.

Was spricht für den PDF/A Standard?

Praktisch jedes Land verwendet ein eigenes Format zur Archivierung von Dokumenten. Gängige Formate für die Archivierung wie Papier, Mikrofilm und Mikrofiche lassen sich grundsätzlich lange archivieren und sind langfristig reproduzierbar. Die Vorteile der digitalen Technologie fehlen allerdings: Große Dokumente lassen sich nicht schnell und unkompliziert um den Globus schicken. Zudem ist es praktisch unmöglich, klassisch archivierte Dokumente nach spezifischen Informationen abzusuchen. Viele Institutionen wählten für ihre ersten elektronischen Archive das TIFF Format aus. Einerseits ist die Reproduzierbarkeit bei diesem Format langfristig garantiert. Andererseits ist TIFF als Bildformat etabliert und kann unkompliziert und schnell an global vernetzte Institutionen verschickt werden. Die freie Suche nach Informationen bleibt aber auch im TIFF Format problematisch.

Das elektronische Archiv verdanken wir innovativen Technologien.

PDF bietet mehr Vorteile als das TIFF Format.

Deshalb hat man begonnen, das PDF Format in Betracht zu ziehen. PDF ist aus mehreren Gründen die attraktivste Alternative:

- PDF umfasst strukturierte Objekte (Text, Vektorgraphik, gerastertes Bild). Diese ermöglichen eine effiziente Suchabfrage im gesamten Datenarchiv. TIFF ist indessen ein gerastertes Format. Um darin eine umfassende Volltextsuche überhaupt zu ermöglichen, muss TIFF im Vorfeld durch eine Texterkennungssoftware (OCR) aufbereitet werden.
- PDF lässt sich kompakt und effizient komprimieren. Im Vergleich zur entsprechenden TIFF Datei braucht PDF nur einen kleinen Teil des Speichervolumens. Zusätzlich ist die Qualität fast immer besser. Besonders für den elektronischen Datenverkehr – E-Mail Anhang, Dateitransfer (FTP) usw. – ist die geringere Dateigrösse äusserst vorteilhaft.
- Metainformationen (Verfasser, Thematik, Inhalt, Schlagwörter, Datum der Entstehung bzw. Neuauflage, Verlag usw.) werden direkt in einem standardisierten Format (XMP) in die PDF Datei eingebettet. Daher lässt sie sich automatisch, systematisch und ohne manuelle Arbeit weiterverarbeiten.
- Das PDF Format ist in der Regel so konzipiert, dass es nicht an ein bestimmtes Gerät (bzw. Rasterauflösung, Farbsystem, usw.) gebunden ist. Erst bei der Wiedergabe wird der Seiteninhalt im Betrachter oder Drucker dargestellt; dieser Prozess heisst Rendering. Dadurch passen sich die PDF Dokumente an die technologische Entwicklung der Ausgabegeräte (Drucker, Bildschirm, Scanner usw.) an und sind auch Jahre nach ihrer Erstellung aktuell.

In den vergangenen fünfzehn Jahren hat Adobe Systems, die Urheberin des PDF Formats, insgesamt acht Versionen ihres PDF Reference Manual veröffentlicht. Das PDF Format wurde bei jeder Überarbeitung um neue Funktionen erweitert; bestehende Funktionen wurden überarbeitet. Darum war es unumgänglich, einen dauerhaften, beständigen und international gültigen Standard für die Langzeitarchivierung zu entwickeln, welcher auf dem PDF Format von Adobe aufbaut. Als Ergebnis dieser Entwicklung entstand der Standard PDF/A.



Der PDF/A Standard

Die Ziele von PDF/A

Die ISO Norm 19005 definiert ein Dateiformat basierend auf PDF, genannt PDF/A. Das Format bietet einen Mechanismus, der elektronische Dokumente auf solche Weise darstellt, dass das visuelle Erscheinungsbild über lange Zeit erhalten bleibt, unabhängig von Werkzeugen und Systemen zur Herstellung, Speicherung und Reproduktion. Dieser Standard legt weder die Methode noch den Sinn und Zweck der Archivierung fest. Definiert wird gleichsam eine Norm für elektronische Dokumente, die garantieren soll, dass das Dokument auch in Zukunft originalgetreu dargestellt werden kann. Deshalb darf das Dokument weder direkt noch indirekt auf eine externe Quelle verweisen. Ein Beispiel dafür wäre ein externes Bild oder eine nicht im Dokument selbst eingebettete Schrift. PDF/A ist als Reihe ausgelegt, die mehrere Standards umfasst. Bisher ist lediglich PDF/A-1 (ISO 19005 Part 1) in Kraft gesetzt worden.

PDF/A Dateien sind selbstbeschreibend.

Alle Informationen, welche dazu benötigt werden, um das Dokument anzuzeigen, sind eingebettet.

PDF/A-1 auf der PDF Referenz 1.4. PDF/A-2 wird auf dem ISO-Standard für PDF 1.7 (ISO-32000) basieren.

Was PDF von PDF/A unterscheidet

Das PDF Format gewährleistet keine langfristige Reproduzierbarkeit, auch nicht die völlige Unabhängigkeit von der Software und dem Wiedergabegerät. Um beide Prinzipien zu garantieren, musste der bestehende PDF Standard eingeschränkt und zugleich erweitert werden. Es war von Anfang an klar, dass PDF/A-1 auf einer bestehenden PDF Version aufgebaut werden muss, um die Akzeptanz bei einem möglichst breiten Publikum zu erreichen. Als Grundlage für den PDF/A-1 Standard wählte das zuständige ISO Komitee TC 171 die Adobe PDF Referenz 1.4 aus.

Diese Referenz wurde von Adobe in ihrem Produkt Acrobat 5 implementiert. PDF/A-1 muss als Standard sowohl sämtlichen Anforderungen dieser Referenz als auch gewissen technischen Grenzen von Acrobat 5 entsprechen. Die ursprüngliche PDF Referenz und die ISO Norm 19005-1 ergeben zusammen den eigentlichen PDF/A-1 Standard. Die ISO Norm 19005-1 hält nur die Differenzen zur PDF Referenz fest. Die PDF Referenz 1.4 ist demnach die zentrale Grundlage, um den PDF/A-1 Standard zu verstehen.

Gewisse Funktionen von PDF 1.4, wie zum Beispiel Transparenz oder die Wiedergabe von Ton und Video, sind beim PDF/A-1 Standard nicht erlaubt. Gewisse Optionen von PDF 1.4 sind in PDF/A-1 obligatorisch: so müssen zum Beispiel alle verwendeten Schrifttypen im Dokument eingebettet sein. Der PDF/A-1 Standard macht also im Wesentlichen nichts anderes, als einzelne Eigenschaften der PDF Referenz 1.4 zu präzisieren bzw. festzulegen, ob diese absolut notwendig, empfohlen, eingeschränkt oder unerlaubt sind.

PDF/A, A-1a, A-1b, A-2 „Babylon“

Die PDF/A-1 Norm ist in zwei Übereinstimmungsgrade unterteilt: PDF/A-1a und PDF/A-1b. PDF/A-1a (Level A Conformance) definiert die Übereinstimmung mit allen Anforderungen des PDF/A-1 Standards.

In PDF/A-1 gibt es zwei Übereinstimmungsgrade.

PDF/A-1a erfüllt alle Anforderungen. PDF/A-1b erfüllt die Mindestanforderungen.

Die Mindestanforderungen zur Übereinstimmung mit PDF/A-1 ist in PDF/A-1b (Level B Conformance) enthalten. Die PDF/A-1b Anforderungen genügen in der Regel für eine eindeutige Wiedergabe über einen langen Zeitraum hinweg.

PDF/A-1a weicht von PDF/A-1b vor allem in Bezug auf die Anforderungen bezüglich Barrierefreiheit (Paragraph 508 des US Rehabilitation Act) ab.

- PDF/A-1a garantiert, dass die Textinhalte extrahierbar sind und die logische Struktur des Dokuments sowie der natürliche Lesevorgang des integrierten Textmaterials bestehen bleiben. Die Textextraktion ist hauptsächlich dann von Interesse, wenn Dokumente auf mobilen Geräten (z.B. PDA) angezeigt oder gemäss dem Paragraph 508 des US Rehabilitation Act visualisiert werden sollen. Dabei muss sich die Darstellung des Textes dem eingeschränkten Bildschirm anpassen, indem er neu angeordnet wird (re-flow). Diese Funktionalität bezeichnet man auch als Tagged PDF.
- PDF/A-1b gewährleistet, dass Text und restliche Seiteninhalte eindeutig wiedergegeben werden, ist aber keine Garantie dafür, dass der eingebettete Text auch verständlich und lesbar ist. Es steht dem Erzeuger einer PDF/A-1b konformen Datei jedoch frei, den Text in lesbarer Form einzubetten, auch wenn die höheren Anforderungen gemäss dem oben erwähnten Paragraph 508 nicht erfüllt sind.

Für gescannte Dokumente ist in der Regel die Übereinstimmung mit PDF/A-1b völlig ausreichend, auch wenn sie mittels OCR so überarbeitet wurden, dass eine Volltextsuche möglich ist.

Zur Zeit arbeitet das technische Komitee an einem neuen Teil des Standards: ISO 19005-2 (PDF/A-2). PDF/A-2 wird entwickelt, um den erweiterten Funktionsumfang der PDF Referenz 1.7 zu berücksichtigen. Inzwischen ist PDF 1.7 selbst normiert, d.h. PDF/A-2 wird nicht mehr auf der PDF Referenz von Adobe basieren, sondern auf der neuen ISO Norm 32000-1 (PDF 1.7).

Entgegen der Gewohnheit mit der PDF Referenz, löst PDF/A-2 den bestehenden Standard PDF/A-1 nicht ab, sondern wird ewig mit diesem koexistieren. PDF/A-2 konforme Betrachter müssen also gleichzeitig auch PDF/A-1 konforme Dokumente korrekt anzeigen können.

PDF/A-1a wurde entworfen, um die Anforderungen bezüglich Barrierefreiheit gemäss Abschnitt 508 des US Rehabilitation Act zu erfüllen.

PDF/A-1b genügt für die eindeutige visuelle Wiedergabe von Dokumenten.

PDF/A-2 soll die bestehende Norm nicht ersetzen, sondern neuere PDF Funktionen einbeziehen.



Nutzung des PDF/A Standards

Woher bekomme ich ein Exemplar?

Der PDF/A-1 Standard ISO 19005-1 wird direkt auf der ISO Website (www.iso.org) vertrieben. Man kann ein Exemplar in Papierform beziehen oder auch elektronisch als PDF. Wie alle anderen ISO Standards unterliegt das Dokument dem Schutz durch Copyright. Darum ist das Anbieten von Gratis-Kopien über das Netz illegal. Der PDF/A-1 Standard ist nur auf Englisch verfügbar.

An wen richtet sich die Norm?

Die PDF/A Norm hat das Ziel, Archivierungsmethoden zu optimieren. Der Standard an sich ist rein technischer Natur. Aus diesem Grund ist er lediglich für Spezialisten mit fundiertem Know-how bezüglich Seitenbeschreibungssprachen wie PostScript und PDF vollständig verständlich. Das Hauptdokument selbst ist klein; der Umfang der Grundlagendokumente dagegen sehr gross. Die PDF Referenz 1.4 beinhaltet alleine fast eintausend Seiten – alle mit der Referenz verknüpften Informationen wie Schrift- und Kompressionsformate, XML Spezifikation, ICC Farbprofile, Digitale Signaturen, RFCs usw. sind dabei noch nicht eingerechnet. Weiter muss festgehalten werden, dass der Standard allein keine langfristige Archivierung garantieren kann. Eine Strategie für die unternehmensweite Archivierung zu erarbeiten, ist in der Regel das Resultat eines umfassenden Projekts. Es wird empfohlen, mit Experten zusammenzuarbeiten, welche die Anforderungen des PDF/A Standards verstehen und anwenden können. Nur so kann eine konsistente Strategie entstehen, welche die langfristigen Ziele für die Aufbewahrung von Dokumenten festlegt.

Die PDF/A-Norm (ISO 19005-1) ist auf der ISO Homepage erhältlich: www.iso.org

PDF/A ist eine rein technische Norm und benötigt das Wissen von Experten, um umgesetzt werden zu können.

Welche Tools sind erhältlich?

Seit 2006 sind verschiedene Werkzeuge auf dem Markt, mit welchen man PDF/A Dokumente erzeugen, verarbeiten und auch prüfen kann. Die Version 8 von Adobe Acrobat beinhaltet die entsprechenden Werkzeuge. Für Office 2007 bietet Microsoft ein Add-in an, welches man separat herunterladen kann. Damit lassen sich PDF/A konforme Dokumente direkt aus der Officepalette heraus anfertigen. Da es bereits unzählige Produkte zur Erzeugung von PDF/A Dokumenten gibt, ist es unterdessen unumgänglich, die jeweiligen Resultate hinsichtlich einwandfreier PDF/A Konformität zu überprüfen.

PDF/A als Bestandteil eines umfassenden Konzepts für Langzeitarchivierung

Die PDF/A Norm als solche ist ein Bestandteil einer umfassenden Lösung. Die Norm selbst gewährleistet noch keine Langzeitarchivierung oder Reproduktionsparameter. Sie ist auch nicht für jedes Projekt die optimale Lösung. PDF/A definiert die spezifischen Anforderungen für elektronische Dokumente, damit diese langfristig archiviert werden können. Falls ein Archiv aufgebaut werden soll, das dem PDF/A Standard entspricht, müssen weitere Aspekte in Betracht gezogen werden. Das betrifft unter anderem die konzerneigenen Standards und Prozesse, das Qualitätsmanagement, vertrauenswürdige Datenquellen sowie dezidierte Anforderungen, die auf den spezifischen Zweck der Anwendung zugeschnitten sind. Insbesondere der Transfer bestehender Papier- oder TIFF-Archive zu einem PDF/A konformen Archiv erfordert eine sorgfältige Planung.

PDF/A ist ein Bestandteil einer umfassenden Archivierungsstrategie.

Die Norm allein garantiert keine Langzeitarchivierung, ist aber eine wesentliche Voraussetzung dafür.



Fazit

PDF/A – der Archivierungsstandard

PDF/A ist der Standard für das Archivieren elektronischer Dokumente. Das PDF Format ist global weit verbreitet. Es wird sowohl im öffentlichen als auch im privaten Bereich für vielfältige Zwecke verwendet. Die PDF/A Norm ist das perfekte Instrument, um die Langzeitarchivierung und Reproduzierbarkeit der Dokumente langfristig zu gewährleisten.

Der PDF/A Standard beeinflusst auch die zukünftige Entwicklung von PDF selbst. Adobe wird unabhängig davon die Entwicklung neuer Funktionen weiterführen. Das betrifft beispielsweise 3-dimensionale Modelle oder XFA für dynamische PDF Formulare. Diese Entwicklungen werden umgekehrt den PDF/A Standard beeinflussen.

Wie reagiert der Markt?

Es ist nicht damit zu rechnen, dass PDF/A Produkte den Markt überschwemmen. Das Wissen, das es braucht, um die Technologie des PDF/A Standards zu verstehen, ist beträchtlich und spezifisch. Dazu kommt, dass der Anwender höhere Qualitätsansprüche an eine Software stellt, die einem Standard entsprechen soll. Die ersten Applikationen sind seit 2006 auf dem Markt. Die Nachfrage betrifft vor allem die PDF/A konforme Herstellung von Dokumenten, die PDF/A-Konformität (Validierung) und die einfache Umwandlung bestehender PDF Dokumente in PDF/A Dokumente.

Mit den ersten professionellen PDF/A Tools entstehen zugleich umfangreiche Projekte für den Aufbau PDF/A konformer Archive. Allerdings darf man zur Zeit noch nicht allzu hohe Erwartungen an die Funktionalität stellen. Wie so oft bei der Einführung eines neuen Standards, wird eine Vielzahl von Produkten auf den Markt kommen, die zwar PDF/A Konformität vorgeben, den Standard aber in Wirklichkeit nicht erfüllen. Experten-Gutachten für die Evaluation sind grundsätzlich zu empfehlen.

PDF/A als langfristige Strategie

Der PDF/A Standard ist keine Eintagsfliege. Schon seit Jahren besteht die Nachfrage nach einer standardisierten Norm für das Archivieren mittels PDF. Das Format wird bereits heute genau für diesen Zweck verwendet, auch wenn viele Anwender dafür spezifische Richtlinien definieren müssen. Die Tatsache, dass Microsoft dem Wunsch der Kundschaft nachgekommen ist, das direkte Erstellen eines PDF/A Dokuments aus der neusten Office Palette heraus zu ermöglichen, ist ein klares Signal: der international gültige Standard für Langzeitarchivierung PDF/A wird von Dauer sein.



Informationen

PDF/A Competence Center

2006 wurde das PDF/A Competence Center gegründet. Dieser internationale Verband hat zum Ziel, den Informations- und Erfahrungsaustausch auf dem Gebiet der Langzeitarchivierung gemäss ISO 19005 – PDF/A zu fördern. Führungskräfte der Unternehmen callas software GmbH, Compart Systemhaus GmbH, intarsys consulting GmbH, LuraTech Europe GmbH, PDF Tools AG, PDFlib GmbH und Seal Systems bilden den Vorstand. In weniger als zwei Jahren sind mehr als 85 Betriebe und Organisationen sowie zahlreiche Fachleute aus mehr als 20 Ländern dem Verband beigetreten.

www.pdfa.org

PDF Tools AG

Schon seit 1993 beschäftigen sich die Experten der PDF Tools AG mit der PDF Technologie. PDF Tools AG ist 2002 aus einem Spin-off entstanden und ist heute ein global führender Produzent von PDF Software für Kunden in allen Marktsegmenten.

Das Angebot der PDF Tools AG umfasst qualitativ hochwertige, client- und serverbasierte Softwareprodukte. Diese sind speziell für Entwickler, Integriatoren, Spezialisten für Kundenanpassungen und IT-Abteilungen konzipiert worden. Zehntausende von Unternehmen setzen die Produkte weltweit entweder direkt oder über ein globales Netzwerk von OEM Partnern ein. Die Tools können problemlos in Applikationsprogramme eingefügt werden.

Der CEO von PDF Tools AG, Dr. Hans Bärzfuss, ist ein weltweit anerkannter PDF Fachmann. Er ist sowohl ein Mitglied des ISO Komitees, welches sich für den PDF/A Standard verantwortlich zeigt, als auch Vizepräsident des PDF/A Competence Center, welches durch PDF Tools AG mitbegründet wurde.

PDF Tools AG hat seinen Hauptsitz in der Schweiz, in unmittelbarer Nähe von Zürich. Ebenfalls dort sind die Entwicklungsabteilung und das europäische Vertriebsteam angesiedelt. Für den Vertrieb in Nord- und Südamerika und im pazifischen Raum ist die Filiale in Kanada zuständig. Sämtliche Produkte können direkt über das Internet erworben werden. Kostenlose DemoverSIONen sind ebenfalls auf der PDF Tools AG Homepage verfügbar.

PDF Tools AG, Kasernenstrasse 1, 8184 Bachenbülach, Schweiz
Tel. +41 43 411 44 50, Fax +41 43 411 44 55, E-Mail pdfsales@pdf-tools.com
www.pdf-tools.com

