

Bessere Langzeitarchivierung mit Hilfe des PDF/A-Standards

ARCHIVIERUNG. Wo viel gedruckt wird, da sind viele Dokumente vorhanden, die oft auch archiviert sein wollen – wofür zumeist natürlich der Kunde selbst sorgen muss. Für die Langzeitarchivierung von digitalen Dokumenten bietet der ISO-Standard PDF/A eine sichere Basis. Nach einer erfolgreichen internationalen PDF/A-Konferenz zeigt das PDF/A Competence Center auf der Drupa Strategien zur Archivierung und Datenorganisation mit PDF/A auf.

Der Standard PDF/A wurde Ende 2005 von der ISO unter ISO 19005-1 verabschiedet. Mittlerweile gibt es zahlreiche Produkte und ein großes Anwenderinteresse – vor allem im deutschsprachigen Raum und in Europa. In den USA scheint sich der Standard dagegen deutlich langsamer zu etablieren, obwohl die amerikanische Organisation AIIM die Standardisierung des Formats auf den Weg gebracht hatte.

Da jetzt die erforderlichen Werkzeuge, nämlich der Standard und entsprechende Software-Tools zur Verfügung stehen, ziehen viele Unternehmen den Umstieg auf PDF/A in Erwägung oder sind gar bereits bei der Vorbereitung entsprechender Workflows. Im Gegensatz zur Druckindustrie, in der sich mit PDF/X schon lange ein PDF-Standard etabliert hat, sind ISO-standardisierte PDF-

Workflows den meisten Unternehmen, Behörden und Archiven noch eher fremd.

ARCHIVIERUNG. In großen Industrieunternehmen, für Behörden, Banken, Versicherungen, Publisher oder reine Archive stellt sich schon immer die Frage der geeigneten Archivierungsformate für große Dokumentenmengen. Traditionelle Archivierungsmethoden wie Papier und Mikrofilm sind umständlich und schwerfällig. Zudem lassen sich die Dokumente nur schlecht wieder auffinden oder verschicken. Deshalb haben schon vor einigen Jahren digitale Methoden wie eine Archivierung als TIFF-Dokument Einzug gehalten. Doch auch TIFF ist nicht unbedingt ein ideales Format, da es sich ohne OCR nicht durchsuchen lässt und außerdem recht viel Speicherplatz benötigt.



Auf der ersten internationalen PDF/A-Konferenz in Amsterdam diskutierten Experten und Anwender Strategien und technische Aspekte der PDF/A-Integration.

PDF dagegen bietet bekanntermaßen kompakte Speicherung, Durchsuchbarkeit und bessere Darstellungsqualität. Aus diesem Grunde wurde PDF/A bereits 2002 zur Standardisierung vorgeschlagen.

WAS IST PDF/A? Das standardisierte Format PDF/A stellt eine Untermenge von PDF dar, die eine authentische Darstellung der Daten auch nach vielen Jahrzehnten garantiert. Damit wird der Anwender unabhängig von bestimmten Herstellern und kann sicher sein, dass die Daten alles enthalten, um sie erneut darzustellen.

Der PDF/A-Standard beschreibt einen Satz von Regeln, was diese PDF-Daten enthalten müssen oder nicht enthalten dürfen. So müssen beispielsweise wichtige Elemente wie Schriften und Farbmanagementinformationen vorhanden sein, um die Darstellung genau reproduzieren zu können. Andererseits sind Dinge, die eine eindeutige Darstellung verhindern, verboten, etwa Transparenzen, Ebenen, Movies oder Javascript. Natürlich darf die Sicherheitseinstellung das Öffnen der PDF-Datei nicht verhindern. (Für Informationen über PDF/A ist die Web Site des PDF/A Competence Centers unter www.pdfa.org eine gute Quelle. Eine umfassende Einführung bietet das Buch »PDF/A kompakt«, ISBN 978-3-9811648-0-0).

PDF/A-1A UND PDF/A-1B. Der Standard definiert zwei verschiedene Varianten von PDF/A. PDF/A-1b garantiert die Mindest-



Das Thema Metadaten, hier präsentiert von Thomas Merz, Geschäftsführer der PDFiib GmbH, war für viele Besucher in Bezug auf Datenorganisation von großem Interesse.



Betsy Fanning von der AIIM, die den ISO-Standard PDF/A auf den Weg brachten, äußert sich erfreut über die gute Akzeptanz des Standards vor allem in Europa.

anforderungen, so dass die visuelle Reproduktion des Dokumentes gewährleistet ist. PDF/A-1a dagegen stellt höhere Anforderungen. Hier muss die logische Dokumentstruktur gespeichert werden. Das erfordert Tagged PDF, das den logischen Aufbau und den Textzusammenhang beschreibt. Vorteil dieser Variante ist, dass sich die Dateien nicht nur darstellen, sondern die Inhalte extrahieren und wiederverwerten lassen. Das ist zum Beispiel für die in den USA bereits gesetzlich verankerte und auch in der EU für offizielle Dokumente vorgeschriebene Barrierefreiheit (Accessibility) erforderlich, also

beispielsweise um den Text für Blinde vorlesen zu können.

PDF/A UND PDF/X. Der in der Druckindustrie verbreitete Austauschstandard PDF/X hat über die ISO-Standardisierung hinaus einiges mit PDF/A gemeinsam. Das ist nicht weiter verwunderlich, da beide für eine eindeutige Reproduzierbarkeit sorgen sollen. Zunächst einmal: Ein Dokument kann durchaus gleichzeitig beiden Standards genügen, also gleichzeitig PDF/A- und PDF/X-konform sein. Natürlich ist der Anwenderkreis eines Austauschformates wie

PDF/X ein anderer als der eines Archivformates, aber es ist durchaus denkbar, dass Druckdaten archiviert werden sollen.

Auch wenn sich die Standards weitgehend überschneiden, gibt es doch gewisse Unterschiede. So sind beispielsweise Überfüllungseinstellungen in einer PDF/A-Datei nicht erforderlich, dafür sind im Gegensatz zu PDF/X bestimmte Metadaten obligatorisch. Zudem gibt es bei der Beschreibung der Farbdaten einige Unterschiede (Vergl. Tabelle auf Seite 76).

PDF/A-WORKFLOWS. Es gibt mittlerweile zahlreiche Softwareanbieter, die Produkte für einen PDF/A-Workflow im Angebot haben. Das sind zum einen Produkte zur Erstellung von PDF/A und zur Konvertierung in dieses Format, zum anderen solche zum Anzeigen und Validieren von PDF/A. Eine Übersicht wichtiger Produkte ist unter <http://www.pdfa.org/doku.php?id=pdfa:produkte> zu finden.

Doch trotz zahlreicher verfügbarer Produkte fehlt es vielen Unternehmen an PDF-Know-how und Erfahrung, um die Umstellung ihres Workflows anzugehen.

DAS PDF/A COMPETENCE CENTER.

Bei der Einrichtung von PDF/A-Workflows und der Vermittlung des nötigen Wissens hilft seit 2006 das PDF/A Competence Center. Eine technische Arbeitsgruppe arbeitet außerdem Richtlinien für die Implementation des Standards aus. Einen Eindruck davon verschaffen die bereits veröffentlichten TechNotes (vergl. <http://www.pdfa.org/doku.php?id=pdfa:techdok>)

Das 2006 gegründete PDF/A Competence Center ist eine unabhängige Herstellerver-

I AM... MARKLESS

RELIABLE
FAST
INVISIBLE
INNOVATIVE

Q.I. Press Controls stellt das mRC vor. Die erste markenlose Registerregelung für Rollendruckmaschinen.

Durch das mRC in Kombination mit dem IDS Farbregelsystem bietet Q.I. Press Controls eine vollkommen markenlose Lösung für Farbregelsysteme und Registersteuerungen.

Sind Sie an Makulaturreduzierung und Qualitätsverbesserung interessiert? Bitte besuchen Sie uns auf der Drupa oder auf unserer Website www.qipc.com.

Q.I. PRESS
CONTROLS

See you at

print media messe
drupa
Stand 17A56

Gemeinsamkeiten und Unterschiede zwischen den beiden Formaten PDF/A und PDF/X

	PDF/A	PDF/X
Ziel	Langzeitarchivierung mit garantierter Reproduzierbarkeit der visuellen Darstellung (PDF/A-1b) und Erhalt der Dokumentstruktur (PDF/A-1a)	Austauschformat für die grafische Industrie
Farben	Output-Intent nur dann erforderlich, wenn geräteabhängige Prozessfarbräume verwendet werden	Output-Intent erforderlich
Schriften	Einbettung erforderlich, Ausnahme unsichtbarer Text	Einbetten der Fonts erforderlich
Überfüllungseinstellung	nicht erforderlich	erforderlich
Seitenparameter	nicht erforderlich	erforderlich
Transparenz	unzulässig	unzulässig
Ebenen	unzulässig	unzulässig
Metadaten	grundlegende XMP-Metadaten erforderlich	möglich
Dokumentstruktur	für PDF/A-1a erforderlich	nicht erforderlich
Aktionen	unzulässig	unzulässig
Javascript	unzulässig	unzulässig
Audio- und Videodaten	unzulässig	unzulässig
Digitale Signaturen	zulässig	zulässig
Passwortschutz und Verschlüsselung	unzulässig	unzulässig

einigung. Ziel des internationalen Verbandes ist die Förderung des Informations- und Erfahrungsaustausches auf dem Gebiet der Langzeitarchivierung gemäß ISO 19005: PDF/A. Mittlerweile sind darin etwa 70 Mitglieder vertreten, darunter weltweite Software-Hersteller, Berater und Anwender.

Auf der vom PDF/A Competence Center organisierten Internationalen PDF/A-Konferenz am 10. und 11. April in Amsterdam versammelten sich über 200 Teilnehmer aus Europa, USA und Japan und diskutierten über Technik und Anwendung des Standards. Dabei kamen neben Experten und Beratern auch Anwender wie Airbus, die Rhätische Bahn oder das Europäische Patentamt zu Wort (weitere Informationen unter www.pdfa.org).

Auf der Drupa 2008 ist das PDF/A Competence Center in Halle 7, Stand 7.0i vertreten.

VALIDIERUNG UND DIE TESTSUITE. Eine der Fragen, die im Rahmen der Konferenz immer wieder auftraten, war: Wie stelle ich eigentlich sicher, ob es sich bei meinen Dokumenten tatsächlich um echtes, standardkonformes und zukunftssicheres PDF/A handelt?

Es gibt hier zwei Vorgehensweisen. Die eine ist, die Dokumente einzeln zu prüfen, also zu validieren, um sicherzustellen, dass sie standardkonform sind. Gerade, wenn Dokumente aus unterschiedlichsten Quellen ins Archiv gelangen, wird man nicht darum herumkommen, sie einzeln zu prüfen. Plant das Unternehmen dagegen einen immer gleichen Workflow zur Erzeugung eigener Dokumente, ist es auch möglich, diesen so zu konzipieren und abzusichern, dass nicht bei jedem Dokument erneut geprüft werden muss. Doch in beiden Fällen braucht der Anwender eine Referenz, also die Validierung seiner Daten oder seines Workflows. Es gibt auf dem Markt eine Reihe von Validierungsprodukten. Doch leider stimmen diese Produkte bisher nicht in ihrer Interpretation des Standards überein, so dass zum Teil Dokumente von einem Validierer abgelehnt oder als richtig akzeptiert werden, während ein anderer Validierer zum umgekehrten Ergebnis kommt.

Damit Anwender eine Sicherheit gewinnen können, dass alle Hersteller gleichen Richtlinien folgen und damit die Validierer richtig prüfen, hat die Technical Working Group des PDF/A Competence Center eine Testsuite entwickelt. Diese so genannte »Isartor-

Suite« enthält zahlreiche Testdokumente, die jeweils einzelne Aspekte des Standards prüfen. Diese Testsuite wurde zur internationalen PDF/A-Konferenz erstmalig vorgestellt und soll demnächst auf www.pdfa.org veröffentlicht werden.

XMP-METADATEN. Ein weiterer wichtiger Aspekt für die Dokumentenarchivierung ist deren Wiederauffindbarkeit. Nützlich an PDF ist, dass es sich durchsuchen lässt, sofern es Text enthält. Doch viele Archive arbeiten schon seit langem mit Metadaten, die Informationen über das Dokument enthalten und dessen Wiederauffindbarkeit verbessern.

Diese Methode lässt sich auf PDF-Daten übertragen, indem XMP-Metadaten ins Dokument eingebettet werden. XMP-Metadaten kommen auch in der Druckindustrie zum Einsatz und werden von einigen Branchenverbänden empfohlen, etwa vom AdML-Konsortium für die Werbebranche oder vom International Press Telecommunications Council (IPTC) als »IPTC Core« für die Metadaten von Bildern und anderen Zeitungselementen. PDF/A-1 benötigt XMP zur Identifikation konformer Dateien und unterstützt zusätzlich benutzerdefinier-

te Metadaten mittels XMP-Extension-Schemas. Eine formale Beschreibung der Extension-Schemas muss im PDF/A-Dokument eingebettet sein, damit benutzerdefinierte Metadaten zu einem späteren Zeitpunkt zweifelsfrei interpretiert werden können (detaillierte Informationen zu diesem Thema unter <http://www.pdfliib.com/de/developer/xmp-metadata/>). Das bedeutet, Unternehmen und Archive müssen bei der Archivierung mit PDF/A nicht unbedingt auf vordefinierte Metadaten-Sets zurückgreifen, sondern können die Metadaten bei Bedarf nach eigenen Kriterien zusammenstellen.

AUSBLICK. Der ISO-Standard PDF/A ist dabei, sich weltweit durchzusetzen, wobei derzeit gerade in Europa besonderes Interesse an dem Archivstandard besteht. Mittlerweile sind zahlreiche Produkte auf dem Markt, die das Format unterstützen.

Nun müssen sich die Anwender daranwagen, solche Workflows zu etablieren. Und das will gut durchdacht und technisch sorgfältig geprüft sein: Denn das Archiv soll nicht nur in 100 Jahren noch wiedergegeben sein, sondern die Dokumente sollen dann immer noch indiziert, gefunden und eventuell wiederverwertet werden können.

Die Autorin Irina Leichsenring, Marketing PDFliib GmbH, arbeitet viele Jahre für Fachzeitschriften, unter anderem als Redakteurin bei der Page und dem Macintosh-Magazin Mac Profiler.

Softwarelösungen für PDF/A

Portfolio der PDFliib GmbH ■ Der PDF-Spezialist aus München bietet eine Reihe von Entwickler- und Serverprodukten an, mit denen sich unter anderem PDF/A-Workflows realisieren lassen. Die PDF/A-Unterstützung in aktuellen Produkten des Herstellers sieht folgendermaßen aus (weitere Infos unter www.pdfliib.com):

PDFliib-Familie: Die Bibliotheken PDFliib, PDFliib+PDI und PDFliib Personalization Server erlauben eine dynamische Erzeugung von PDF/A-1b und PDF/A-1a. Die PDFliib-Familie unterstützt XMP-Metadaten und benutzerdefinierte XMP-Extension-Schemas.

PDFliib Plop: Die Sicherheits- und Optimierungsfunktionen des Produktes erlauben es, PDF/A-konforme Eingabedaten nach der Bearbeitung wieder PDF/A-konform auszugeben. Mit Plop können XMP-Metadaten in vorhandene PDF/A-Dokumente aufgenommen werden. Da Plop 3.1 XMP-Extension-Schemas für PDF/A unterstützt, können diese in einem zweiten Bearbeitungsschritt nach der Erstellung mit anderen Produkten ohne XMP-Unterstützung in existierende Dokumente eingebracht werden.

PDFliib Plop DS: Die erweiterte Version Plop DS kann zusätzlich PDF/A-konforme Signaturen ins PDF/A einbringen.

PDFliib P-Cos: P-Cos bietet eine einfache Programmiermethode, um XMP-Metadaten aus PDF-Dokumenten zu extrahieren. P-Cos ist als eigenständiges Produkt verfügbar und darüber hinaus in alle anderen Produkte integriert.

PDFliib TET: Das Text Extraction Toolkit stellt ein patentiertes Verfahren zur Verfügung, Text aus existierenden PDF-Dokumenten zu extrahieren.

PDFliib TET PDF I-Filter: Der auf der TET-Engine basierende I-Filter dient zum Extrahieren von Text, Metadaten und Bookmarks, beispielsweise für die Indizierung in Microsoft Office Sharepoint Server, Exchange Server, SQL Server und anderen Produkten, welche die I-Filter-Schnittstelle unterstützen.

MINI-PRODUKTE GANZ SPEZIAL.



Beipackzettel, mit bis zu 14 Parallel-Faltungen, geleimt oder gelabelt, Lotterierprodukte. Die genaue und gekonnte Produktion im Einfach- oder Mehrfachnutzen macht wettbewerbsfähig.