

PDF/A für die Langzeitarchivierung?

Experten diskutieren kontrovers über das Für und Wider eines nicht mehr ganz neuen Standards.

Von **Frank Zscheile***

Seit viereinhalb Jahren ist Portable Document Format/Archive (PDF/A) das normierte Format für die Langzeitarchivierung. Mit seiner zunehmenden Verbreitung als ISO-Standard sind in letzter Zeit Missverständnisse aufgekommen. Einige Experten sehen die Entwicklung in

Deutschland kritisch und dadurch den Nutzen für die Anwender gefährdet. Unsicherheit herrscht auch darüber, welche Dokumente am besten in dem Format gespeichert werden sollten und welche nicht. Zuletzt trafen sich auf der CeBIT zu diesem Thema PDF/A-Verfechter und ECM-Herstell-

er, um bei einer Podiumsdiskussion unter Moderation von Unternehmensberater Bernhard Zöller die Sinnhaftigkeit von PDF/A zu diskutieren. Dabei wurde unter anderem eines deutlich: Die Technik bietet heute bereits mehr Funktionen, als Anwender in der Praxis derzeit zu nutzen gewillt sind.

Das bringt PDF/A

PDF/A ist der ISO-Standard 19005 für die Langzeitarchivierung im Portable Document Format. Er stellt eine eingegrenzte Variante von PDF dar, ein standardisiertes Profil zur Verwendung von PDF in der Langzeitarchivierung. Der Standard gibt detailliert an, welche Inhalte erlaubt sind und welche nicht. Durch diese und andere Vorschriften soll eine langfristige Lesbarkeit der Dokumente garantiert sein – und zwar unabhängig davon, mit welcher Anwendungssoftware und auf welchem Betriebssystem sie ursprünglich erstellt wurden. Allerdings ist die Erstellung solcher Dokumente mit erhöhtem Aufwand verbunden: Während der Anwender für das Erzeugen herkömmlicher PDFs aus zahllosen Freeware-Tools wählen kann, benötigt er für den ISO-Standard verlässliche Werkzeuge, die immer mehr Hersteller anbieten. So genannte Validatoren, die vor der Einstellung des Dokuments in das elektronische Archiv prüfen, ob es sich wirklich um ein PDF/A handelt, sorgen für eine einheitliche Qualität.

Akzeptanz von PDF/A ist in Europa größer als in den USA

Die Schaffung von PDF/A im Jahr 2002 ging zwar auf amerikanische Initiative zurück. Laut Harald Grumser, Vorstandsvorsitzender des PDF/A Competence Center, eines internationalen Interessenverbands, hat sich PDF/A aber „in Europa viel dynamischer entwickelt als in den USA. Dort hinkt die Marktentwicklung bislang gut zwei Jahre hinterher.“ Als CEO des Herstellers Compant berät Grumser vor allem Großkunden bei Archivierungs- und Druck-Output-Projekten. „Die Frage nach dem zu

Karl-Heinz Mosbach,
ELO:



„Viele Unternehmen archivieren immer noch Standard-PDFs.“

wählenden Format wird dort gar nicht mehr diskutiert“, sagt er. „PDF/A hat sich unmerklich in den Köpfen festgesetzt, was auch mit der wachsenden Bedeutung des Themas Compliance zu tun hat: Wer auf einen ISO-Standard setzt, der ist auf der sicheren Seite.“

Einschlägige Gesetze werden auch hierzulande kommen

In Deutschland ist PDF/A als Standard zur Langzeitarchivierung bislang nicht zwingend vorgeschrieben. Anhänger des Formats gehen aber davon aus, dass dies nur noch eine Frage weniger Jahre ist.

In den USA stehen derzeit eine Reihe neuer Gesetzesentwürfe an, die wohl rasch zu einer stärkeren Verbreitung des Standards führen werden. Carsten Heiermann, Geschäftsführer des Softwareanbieters Luratech Europe GmbH und Mitglied im PDF/A Competence Center: „Welche Compliance-Vorschriften zum Beispiel US-amerikanische Atomkraftbetreiber einhalten müssen, bestimmt die Nuclear Regulatory Commission (NRC), und die hat PDF/A im Jahr 2009 als Standard für alle Dokumente gesetzt, die bei Regierungsbehörden und Kontrollkom-

missionen eingereicht werden.“ Auch in europäischen Ländern werden entsprechende Gesetze erlassen.

Doch wann ist es überhaupt sinnvoll, Dokumente oder E-Mails samt Attachments in PDF/A zu archivieren? Kann es sogar schädlich sein? Dies wollte Bernhard Zöllner von den Diskussionsteilnehmern wissen. Jürgen Biffar, Vorstand der Docuware AG, betrachtet das Format für gescannte Dokumente als überflüssig. Es bringe keine Vorteile: „PDF/A wurde definiert, um sicherzustellen, dass sämtliche Zusatzinformationen, die zur präzisen Darstellung erforderlich sind, zum Beispiel alle verwendeten Schriftarten, in die Datei eingebettet sind. Bei gescannten Dokumenten gibt es aber keine derartigen Zusatzinformationen. Man darf nicht vergessen, dass die Erzeugung von PDF/A-Dokumenten mehr Aufwand bedeutet und bestimmte Techniken erfordert.“ Der Trend gehe dahin, immer mehr farbige Dokumente zu scannen und dafür PDF zu wählen. Biffar weiter: „Man hat bisher primär JPGs erzeugt, was aber in der Handhabung komplizierter war, weil man pro Seite eine Datei hat. Heute kann man bequem aus farbigen gescannten Dokumenten normale PDFs er-

Harald Grumser,
Compart:



„Wer auf einen ISO-Standard setzt, der ist auf der sicheren Seite.“

Carsten Heiermann,
Luratech:



„Keine unterschiedlichen Formate für die Archivierung.“

zeugen, die eine vernünftige Größe haben und alle Anforderungen an die Langzeitarchivierung erfüllen.“ Als Format für die Langzeitarchivierung von computererzeugten Dokumenten befürwortet Biffar jedoch PDF/A.

Die Branche lässt ihren Anwendern die Wahl

ECM-Hersteller überlassen ihren Kunden daher die Wahl des Formats. Bei ELO können Benutzer standardmäßig entscheiden, ob sie im PDF/A-Format archivieren möchten oder nicht. „Eine reine Einstellungssache“, so Karl-Heinz Mosbach, Geschäftsführer der ELO Digital Office GmbH. Seine Beobachtung: Unternehmen, die Compliance-orientiert denken, entscheiden sich für PDF/A. Dies empfiehlt der ELO-Chef auch seinen Kunden und stellt gleichwohl fest, dass die Mehrzahl noch immer Standard-PDFs archiviert, weil Langzeitaufbewahrung kein drängendes Thema für sie ist.

PDF/A also nur dort, wo es wirklich sinnvoll ist? „Für Dokumente, die nicht langzeitarchiviert werden müssen, reicht das normale PDF teilweise aus“, erklärt Mosbach. ▶

Wohin werden Ihre Daten Sie bringen?



Es heißt: „Der Weg ist das Ziel.“ Wenn Ihre Daten Ihr Unternehmen unterstützen, können Sie beides genießen. Bessere Daten verbessern Geschäftsprozesse, lassen Sie Wege entdecken, um Ihre Umsätze schnell zu erhöhen, Kosten zu senken und Risiken zu minimieren.

Innovative Datenmanagement-Lösungen von DataFlux erlauben es Ihnen, Daten in ein strategisches Gut zu verwandeln. Fach- und IT-Anwender können gemeinsam daran arbeiten, die Basis für ein agileres und geschickt agierendes Unternehmen zu legen. DataFlux konnte mit seinen Datenqualitäts-, Datenintegrations- und Stammdatenmanagement-Lösungen schon zahlreiche Preise gewinnen. Unsere Lösungen ermöglichen auch in Ihrem Unternehmen bessere und schnellere Entscheidungen – basierend auf genauen und zuverlässigen Daten.

Wenn Ihr Unternehmen von besseren Daten gelenkt wird, eröffnen sich Ihnen unendlich viele Möglichkeiten.



info.de@dataflux.com

► Wer aber sichergehen will, dass sein Dokument auch nach zehn Jahren noch lesbar ist, sollte sich aufgrund von Spezifikationen wie der Font-Einbettung für PDF/A entscheiden. Dass die Dateien durch die enthaltenen Fonts zu groß sind, sei ein unzutreffendes Gerücht, erklärt Compart-CEO Grumser: „Eine einseitige Schwarzweiß-Datei umfasst als Tagged Image File Format (Tiff) rund 70 KB, als PDF/A ist sie nur unwesentlich größer. Die Frage stellt sich außerdem angesichts der immer geringeren Preise für Speicherplatz nicht mehr.“

Einheitliche Archive – leichte Migrationen

Unterschiedliche Formate für die Archivierung also, je nach Art des Dokuments? Eine Strategie, vor der es PDF/A-Verfechtern wie Luratech-Chef Heiermann graust: „Die Idee hinter dem Ganzen ist doch eine Formatvereinheitlichung im elektronischen Archiv. Spätestens wenn eine Migration ansteht, wird es mit vielen unterschiedlichen Formaten zu kompliziert. Man kann nicht bei



Bernhard Zöller,
Zöller & Partner:

„Es gibt wenige Ausnahmen, bei denen PDF/A zu Informationsverlust führt.“

jeder Datei klären, ob sie noch einmal zehn Jahre in der jetzigen Form aufbewahrt werden kann.“ Er rät Anwendern daher, beim Archivformat keinen Unterschied zwischen gescannten oder elektronisch erzeugten, farbigen oder schwarzweißen Dokumenten zu machen. „PDF/A ist ein normierter Container, bei dem man bleiben kann, wenn sich kein signifikanter Nachteil daraus ergibt.“

Normierter Container für gemischte Inhalte

Neben der Volltext-Durchsuchbarkeit liegt der Vorteil des Formats vor allem darin, dass es sich für alle Dokumentenarten nutzen lässt, und das ist gut, wenn etwa ein Dokument aus gemischten Inhalten erstellt

werden soll, zum Beispiel aus elektronisch erzeugten und gescannten Dokumenten. Laut Heiermann gibt es daher kaum Gründe, an den alten Techniken festzuhalten.

Nicht alle Formate überführbar

Wenige Ausnahmen bestätigen auch bei der Formattfreiheit die Regel: „Dateien aus Microsoft Project oder Audiofiles, die sich eben gar nicht oder nur mit Informationsverlust in Tiff oder PDF/A konvertieren lassen, sondern im eigenen proprietären Format aufbewahrt werden müssen“, warf Zöller ein. Ein weiteres Argument führt Docuware-Vorstand Biffar an: „Wir stellen in Projekten regelmäßig fest, dass unsere Kunden viele Dokumente und E-Mails bewusst im Originalformat aufbewahren, mit allen Problemen, die daraus entstehen, und dem Risiko, dass die Dateien möglicherweise in 20 Jahren nicht mehr lesbar sind. Wenn nämlich jemand ein Dokument auch künftig editieren will, dann gibt eine Konvertierung, in welches Format auch immer, keinen Sinn.“ Docuware stellt daher parallel Viewing-Techniken zur Verfügung, damit die Dokumente auch unabhängig von der ursprünglichen Anwendung lesbar sind. Für den PDF/A-Experten Heiermann ist dies unverantwortlich: „Auch solche Viewer veralten. Wenn ein Unternehmen heute seine Dokumente in einem Format archiviert, das in 20 Jahren womöglich nicht mehr zu öffnen ist, dann sollten die ECM-Hersteller ihre Kunden dies nicht einfach so tun lassen, sondern sie beraten und ihnen empfehlen zu konvertieren.“

Hybridarchivierung als gesunder Mittelweg

Ein gesunder Mittelweg könnte daher die Hybridarchivierung sein: Jede neue Version eines Dokuments wird zusätzlich als PDF/A gespeichert. Darauf scheinen sich die ECM-Hersteller einigen zu können, wie zum Beispiel ELO-Chef Mosbach, der seinen Kunden bereits die parallele Erzeugung von PDF- und Tiff-Dateien – neben der Beibehaltung eines proprietären Formats – empfiehlt. So nämlich lässt sich eines auf jeden Fall vermeiden: dass wichtige Firmenunterlagen irgendwann überhaupt nicht mehr lesbar sind. Auch Carsten Heiermann empfiehlt bei Dokumenten, die zum Beispiel in einer elektronischen Akte noch im „Lebenszyklus“ sind, eine Version im nativen Format zu speichern, um sie weiterbearbeiten zu können, sowie eine zweite als PDF/A.

Wie so oft nehmen Kunden auch hier nicht ohne weiteres an, was die Industrie ihnen bietet. So zum Beispiel die nützliche Eigenschaft von PDF/A, Metadaten in einem nor-



Jürgen Biffar,
Docuware:

„Viele Dokumente und E-Mails werden bewusst im Originalformat aufbewahrt.“

mierten Format abzubilden und über Validatoren prüfen zu können – bei herkömmlichen PDFs fehlt diese Eigenschaft. Diese hilfreiche Funktion läuft aber ins Leere, da, so Biffar, Anwender die Metadaten kaum benutzen, obwohl sie im Zuge der stets wachsenden Dokumentenarten und -mengen immer wichtiger werden. Denn sie stellen ein gutes Werkzeug zur schnelleren Bearbeitung oder Indexierung dar und erleichtern zudem den Austausch zwischen verschiedenen Anwendungen. (ue)

PDF/A-Neuerungen

Der Standard entwickelt sich weiter: Für Herbst 2010 ist der neue Standardteil **PDF/A-2** anvisiert. Er setzt auf Version 1.7 des PDF auf und unterstützt deshalb auch technische Neuerungen wie JPEG2000. Außerdem sollen sich damit **komplexere PDFs** mit mehreren Schichten transparent darstellen lassen. Eine Ablösung wird es jedoch nicht geben, Anwender sollen vielmehr selbst entscheiden, welche Version des Standards sie nutzen wollen. Allerdings ist die Erstellung solcher Dokumente mit erhöhtem Aufwand verbunden: Während der Anwender für das Erzeugen herkömmlicher PDFs aus zahllosen Free-ware-Tools wählen kann, benötigt er für den ISO-Standard verlässliche Werkzeuge, die immer mehr Hersteller anbieten. So genannte **Validatoren**, die vor der Einstellung des Dokuments in das elektronische Archiv prüfen, ob es sich wirklich um ein PDF/A handelt, sorgen für eine einheitliche Qualität.

*Frank Zscheile ist freier Journalist in München.