

## PDF 2.0 – Next Generation

Dokumentenmanagement, Archivierung, Dateiformate, ISO-Standard, PDF-Unterformate, PDF/A-Validator



Halle 1 B52

Dass industrielle Erzeugnisse mehr als 20 Jahre überleben, ist selten. Dass sie dies in der Informatik-Branche tun, ist umso erstaunlicher. Nicht einmal die Väter von PDF konnten sich im Juni 1993, als die erste Version von Acrobat auf dem Markt erschien, den einmaligen Erfolg ihres Dateiformats vorstellen. Nun arbeiten die Mitglieder der internationalen Standardisierungsorganisation ISO an der nächsten Generation des beliebten Dateiformats.

Seit Mitte 2008 ISO-Standard mit der genauen Bezeichnung ISO-32'000-1, „Document Management – Portable Document Format – Part 1: PDF 1.7“, wurde die sechste Ausgabe von Adobe's berühmter PDF Reference bis dato nicht wesentlich verändert, sondern nur in die ISO-Sprache übersetzt. Mit dem baldigen Erscheinen des zweiten Teils des Standards „Part 2: PDF 2.0“ wird dies nun anders: Die neue PDF-Version wurde vollständig durch die ISO-Mitglieder, genauer genommen durch das technische Komitee 171, Sub-Komitee 2, erarbeitet. Um dies auch nach außen sichtbar zu machen, hat die Hauptversionsnummer eine Zwei erhalten. Gegenwärtig ist der Standard im Zustand DIS (Draft International Standard), eine Stufe kurz vor der Publikation, über die am 16. September 2015 abgestimmt wird.

### Was bringt der neue Standard?

Die Liste der Änderungen umfasst mehr als 50 Einträge. Die wichtigsten Neuerungen und Verbesserungen betreffen die folgenden Themen:

- Verschlüsselung: Unverschlüsselte Verpackungen von verschlüsselten Dokumenten, 256-bit AES-Verschlüsselung, Unicode-Kennwörter
- Digitale Signaturen: Signaturen basierend auf dem CAdES-Standard, Zertifikate basierend auf elliptischen Kurven, Langzeit-Signaturen (LTV)

[www.pdf-tools.com](http://www.pdf-tools.com)

**Dr. Hans Bärffuss** ist Gründer und Geschäftsführer der **PDF Tools AG**, einer international erfolgreichen Softwareentwicklungs- und Vertriebsgesellschaft. Er ist Delegierter der Schweizerischen Normenvereinigung (SNV) bei der ISO und hilft bei der Standardisierung von Dateiformaten und digitalen Signaturen mit. Er ist einer der Initiatoren und Gründer der PDF Association eines Verbands zur Bekanntmachung und Förderung der ISO Standards PDF und PDF/A, und Chairman des Swiss Chapter und hält zahlreiche Fachvorträge auf Konferenzen und Seminaren und publiziert Fachartikel zum Thema digitale Dokumente.



- Annotationen: Projektionen, 3D, Rich Media
- Barrierefreiheit: Aussprache-Hinweise
- 3D: Unterstützung des neuen ISO-Standards „PRC“, Messungen im 3D-Raum
- Attributierung von Dokumententeilen

Im Komitee hatte man auch den Mut, einige Altlasten zu loszuwerden. Die wichtigsten sind:

- XFA Formulare: Adobes Formulartechnologie, basierend auf XML, hat vielen Anbietern große Bauchschmerzen verursacht
- Movie, Sound: Multimedia Inhalte entsprechen nicht dem Konzept eines portablen Dokumentenformats
- Überflüssige, redundante, veraltete oder nicht portable Informationen wie das Document Information Dictionary, welches durch XMP ersetzt wurde, nicht mehr zeitgemäße digitale Signaturen, Betriebssystem-abhängige Dateinamen und kaum verwendete Standards wie OPI (Open Prepress Interface)

Im neuen Teil des Standards wurden auch substantielle Überarbeitungen gemacht. Dies betrifft vor allem die folgenden Kapitel:

- Ausgabe auf einem Rastergerät (Rendering)
- Transparenz
- Digitale Signaturen
- Metadaten
- Zugänglichkeit und barrierefreie Dokumente (Tagged PDF and Accessibility Support)

Doch die zahlreichen Änderungen fordern ihren Preis. Sieben Jahre für die Erarbeitung des zweiten Teils sind ein wesentlich längerer Zeitraum als für frühere Versionen benötigt wurde – Adobe schaffte sieben Versionen in nur 15 Jahren. Und dies in einer hervorragenden Qualität. Dafür ist der zweite Teil des Standards nun bei den ISO-Mitgliedern breit abgestützt. Viele Textstellen sind präziser formuliert. Dies ermöglicht es der Industrie,

die Spezifikation besser zu verstehen, die Qualität der Implementationen zu erhöhen und damit die Interoperabilität zu verbessern. Man hofft, dass dadurch deutlich weniger „Bad PDF“ erzeugt werden.

### Welchen Einfluss hat die neue Version?

Für die wichtigsten Anwendungen von PDF – Archivierung (PDF/A), den Dokumentenaustausch (PDF/X), das Engineering (PDF/E) und die Zugänglichkeit (PDF/UA) – hat die ISO spezialisierte Unterstandards definiert, welche mehrheitlich auf dem ersten Teil des PDF-Standards basieren. Es ist abzusehen, dass auch diese Standards so weiterentwickelt werden, dass sie sich auf den zweiten Teil beziehen können.

Allerdings sollte man nicht davon ausgehen, dass nun der Überstandard „neu“ ist und die Unterstandards „alt“. Vielmehr ist die Entwicklung dieser Standards als Wechselwirkung zu verstehen. So sind viele Änderungen im zweiten Teil des PDF-Standards aus den Erkenntnissen der Arbeit an den Unterstandards hervorgegangen und in die Entwicklung eingeflossen. Und – anders als im PDF-Überstandard – besteht in den Unterstandards für PDF/X, PDF/E und PDF/UA kein großer Änderungsdruck, da diese Standards schon seit längerem eigenständig und unabhängig von Adobe laufend weiterentwickelt werden. Etwas anders sieht die Sache für PDF/A aus.

### Der Spezialfall Archivierung

Sobald die ersten PDF-Dateien der Version 2.0 eingesetzt werden, stellt sich sofort die Frage, wie diese Dateien normgerecht archiviert werden können. Darauf muss PDF/A eine Antwort geben können. Anders als bei den anderen Unterstandards von PDF entsteht für diese Anwendung ein gewisser zeitlicher Druck

und die Vielzahl der Änderungen ist auch nicht gerade die beste Voraussetzung für eine rasche Lösung. Zudem hat die PDF/A-Gemeinschaft noch mit anderen Problemen zu kämpfen, insbesondere mit dem Thema der Validierung.

Mit dem Prozess der Validierung wird geprüft, ob eine PDF-Datei einem bestimmten Standard entspricht. Diese Art der Prüfung ist vor allem für PDF/X- und PDF/A-Dateien verbreitet. Für die Archivierung ist eine Prüfung vital, da Verstöße gegen den Standard dazu führen können, dass die archivierten Dateien nach zehn oder mehr Jahren nicht mehr einwandfrei gelesen werden können.

## Unterschiede bei der Validierung von PDF/A-Dateien

Es gibt eine Reihe kommerzieller Softwareprogramme (Validatoren), welche die Konformität zu den verschiedenen Teilen des PDF/A-Standards prüfen. Da alle Teile des PDF/A-Standards auf bestimmten Überstandards wie PDF 1.4, PDF 1.7 usw. basieren, müssten die Validatoren auch die Konformität zu diesen Standards prüfen. Zudem interpretieren die Hersteller dieser Validatoren den Text der Standards oft unterschiedlich. Beides führt dazu, dass die Ergebnisse der Prüfung unterschiedlich ausfallen können, je nach eingesetztem Produkt. ►

Hinzu kommen unrealistischen Vorstellungen einzelner Anwender, was ein Validator leisten soll und kann. So hat beispielsweise die Schweizerische Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) auf ihrer Webseite die Forderung aufgestellt, dass ein Validator feststellen soll, ob eine PDF/A-Datei mit einem bestimmten Programm und einem bestimmten Druckertreiber einwandfrei gedruckt werden kann. Motiviert war die Forderung durch einen Fehldruck, welcher, wie eine Analyse des Falls zeigte, durch einen Softwarefehler im druckenden Programm verursacht wurde. Sowohl die Mängel der Validatoren als auch die unrealistischen Vorstellungen einzelner Anwender können dazu führen, dass das PDF/A-Konzept in Frage gestellt wird. Deshalb wird der Ruf nach einem „Definitiven Validator“ immer lauter.

### Das VeraPDF-Projekt

VeraPDF ist ein Projekt der Open Preservation Foundation. Es wurde als Konsortium zwischen den Partnern PDF Association, Dual Lab, The Digital Preservation Coalition und Keep Solutions gebildet. Ziel des Projekts ist die Entwicklung eines PDF/A-Validators, welcher von PREFORMA ausgeschrieben war, PREservation FORMAts for Culture Information/E-Archives ist ein prä-kommerzielles Einkaufsprojekt (PCP), welches vom EU-Programm FP7-ICT mitgegründet wurde. Der Open Source Validator von PREFORMA soll für drei Dateiformatstandards – neben PDF sind dies TIFF und ein Video-Format – eingesetzt werden und die Langzeitarchivierung in Gedächtnisinstitutionen unterstützen. VeraPDF hat die Ausschreibung der ersten zwei Phasen für die Realisierung des PDF/A-Validators gewonnen. Die erste Phase beinhaltet im Wesentlichen die Spezifikation des Validators und ist abgeschlossen. Die zweite Phase, die Realisierung eines Prototyps, steht kurz vor dem Abschluss.

Die bisherige Erfahrung mit dem Projekt hat gezeigt, dass es

nicht so einfach ist einen PDF/A-Validator auf der grünen Wiese zu entwickeln. Es braucht dazu sehr viel Erfahrung mit PDF. Und einen der kommerziellen Hersteller von Validatoren damit zu beauftragen, war aus politischen Gründen nicht möglich. Der Einbezug der PDF Association hat den Vorteil, dass darüber die Erfahrungen der Hersteller einfließen und die Entwicklung des Validators breit abgestützt ist. Der Nachteil ist aber, dass die Entwicklung dadurch viel aufwändiger ist und länger dauert. Es fehlt auch das Geld für eine vollständige Implementation und man versucht, mit einer entsprechenden Softwarearchitektur (Plug-ins) das Konzept brauchbar zu machen.

Einer der Vorteile des Projekts gegenüber kommerziellen Validatoren soll sein, dass der Validator selbst „validiert“ werden kann. Dies soll dadurch ermöglicht werden, dass der gesamte Programmcode öffentlich gemacht wird. Tatsache ist, dass kaum jemand den Code wirklich anschaut. Deshalb versucht man, sich auf geeignete Testdateien zu konzentrieren. Es ist aber kaum überraschend, dass die Entwicklung der Testdateien mindestens so viel Zeit benötigt, wie die Entwicklung des Validators selbst. Die existierenden Testsuiten wie die Isartor Testsuite, welche durch die PDF Association entwickelt wurde, genügen bei weitem nicht.

### Fazit

Steht die Entwicklung der neuen PDF-Version kurz vor der Veröffentlichung, ist mit konkreten Ergebnisse aus dem VeraPDF-Projekt für die Validierung von PDF/A-Dokumenten noch nicht zu rechnen. Denn die voraussehbaren Erfahrungen mit diesem Vorhaben beginnen die Gemeinschaft zu spalten und erste Enttäuschungen machen sich breit. Viele Hoffnungen werden nicht oder erst sehr spät erfüllt werden. Die Anwender werden sich wohl noch für eine längere Zeit auf kommerzielle Validatoren verlassen müssen. ■