

Microsoft Word oder PowerPoint Dateien nach PDF/A-1b konvertieren

Um das verbreitete Dateiformat PDF von Adobe Systems für die Langzeitarchivierung zu verwenden wurde der ISO Standard PDF/A festgelegt. Word oder PowerPoint Dateien werden häufig nach PDF/A konvertiert bevor sie an digitale Archive abgeliefert werden. Im Gegensatz zu gebräuchlichen PDF Formaten verzichtet das PDF/A Format auf Funktionen, die nicht als langzeitauglich betrachtet werden. PDF/A speichert insbesondere alle Informationen innerhalb der Datei, die zur einwandfreien Wiedergabe benötigt werden (z.B. alle genutzten Zeichensätze). Wir empfehlen die Konformitätsstufe PDF/A-1b zu verwenden, welche auf eindeutige visuelle Reproduktion ausgelegt ist und keine eingebetteten Dateien erlaubt.

Das vorliegende Dokument ist in drei Abschnitte unterteilt. Im ersten Abschnitt geben wir eine Anleitung um aufwendige wissenschaftliche Dokumente nach PDF/A-1b zu konvertieren. Im zweiten Abschnitt beschreiben wir verschiedene alternative Methoden. Im dritten Abschnitt fassen wir unsere Tests zu den verschiedenen Methoden zusammen.

1. Anleitung

Um ein aufwendiges wissenschaftliches Dokument aus MS Word oder PowerPoint nach PDF/A-1b zu konvertieren, empfehlen wir URLs von Hyperlinks im Text explizit anzugeben und folgende Methode zu verwenden¹:

- a) In MS Word oder MS PowerPoint unter dem Menüpunkt „Datei“ wählen Sie die Schaltfläche „Drucken“.
- b) Unter „Drucker“ wählen Sie „Adobe PDF“ aus. (Falls das Feld „Adobe PDF“ fehlt, ist auf ihrem Computer vermutlich nur Adobe Reader installiert. Sie könnten entweder Adobe Acrobat Standard oder Professional installieren 2 oder Methode 3 zur Konvertierung verwenden.)
- c) Wählen Sie das Feld „Druckereigenschaften“ an. Im Tab „Adobe PDF Einstellungen“ wählen Sie in „Standardeinstellungen“ den Dropdown Menüpunkt „PDF/A-1b: 2005 (RGB)“ aus.
- d) Wählen Sie „Drucken“ um die PDF Datei zu erzeugen.

Sie sollten die Qualität der erzeugten PDF Datei sorgfältig visuell überprüfen. Achten Sie dabei insbesondere auf Formeln, Hyperlinks, Sonderzeichen, Umlaute, spezielle Fonts, Textschreibfehler, Auswählen und Suchen im Text, Kommentare, Tabellen, Farben, transparente Objekte, Vektorgraphiken sowie mehrfache Zeichenebenen.

2. Beschreibung verschiedener Konvertierungsmethoden

Es wurden vier Methoden getestet um PDF/A-1b Dateien aus MS Word oder MS PowerPoint Dateien zu erzeugen (MS Office Professional Plus 2010, Version 14.0). Dabei kommt Methode 3 ohne kostenpflichtige Software aus. Die Methoden 1, 2 und 4 benötigen die kostenpflichtigen Produkte Adobe Acrobat Standard oder Adobe Acrobat Professional²:

- 1) Sie können mit Adobe Acrobat Professional konvertieren. Dies geschieht in folgenden Schritten: Starten Sie Adobe Acrobat Professional und öffnen Sie die Word oder PowerPoint Datei mit Adobe Acrobat Professional. Dann wählen Sie im Menü „Datei“ den Menüpunkt „Speichern als...“ und wählen „PDF/A“. Im nächsten Fenster wählen Sie Dateinamen und Speicherort sowie die Schaltfläche „Einstellungen...“. Im nächsten Fenster wählen Sie „PDF/A-1b“ und „sRGB.“³
- 2) Ein Word oder PowerPoint File kann mit dem Acrobat PDFMaker als PDF/A-1b gespeichert werden. Dazu Word oder PowerPoint starten und das Menü „Acrobat“ auswählen, dann Grundeinstellungen, PDF/A-1b wählen (einmalig für jede Installation). Dann erzeugen Sie das PDF Dokument entweder im Menü „Acrobat“ mit dem Menüpunkt „PDF erstellen“ oder im Menü „Datei“ mit dem Menüpunkt „Als Adobe PDF speichern...“.⁴
- 3) Ein File kann in Word oder PowerPoint als PDF/A-1b gespeichert werden. Dazu unter dem Menü „Datei“, „Speichern unter...“ auswählen, dann Format PDF/A auswählen. Unter „Optionen...“ selektieren Sie „ISO19005-1-kompatibles PDF/A“ (ISO19005-1 ist PDF/A-1; das erzeugte File müsste also mindestens PDF/A-1b erfüllen).⁵
- 4) Ein File kann in Word oder PowerPoint zu einem PDF Drucker geschickt werden. Dies ist die empfohlene Methode. Sie wird oben im Abschnitt „Anleitung“ beschrieben.

3. Tests zu den verschiedenen Konvertierungsmethoden

Um Vor- und Nachteile dieser vier Methoden zu testen wurden fünf Word Dateien und fünf PowerPoint Dateien ausgewählt. Diese Dateien enthielten jeweils aufwendige wissenschaftliche Publikationen mit Figures, Gleichungen und speziellen Schriftzeichen. Die vier Methoden wurden folgendermassen beurteilt:

- Es wurde visuell überprüft ob die druckbaren Eigenschaften originalgetreu abgebildet wurden. Dabei wurden insbesondere Figures, Formeln, spezielle Fonts und Text geprüft.
- Das Funktionieren von Endnoten-Links und dem Text unterlegten Hyperlinks wurde überprüft. Dieser Punkt wurde als weniger wichtig beurteilt, weil solche impliziten Links in Publikationen selten vorkommen.
- Rechenzeit und Dateigrösse
- Die Gültigkeit der erzeugten PDF/A-1b Datei wurde mit Adobe Preflight überprüft (Ctrl+Shift+X in Adobe Professional XI).

Die Erhaltung von Tags oder Metadaten wurde nicht beurteilt. Ebenso wurde die Erhaltung von Kommentaren unterhalb der PowerPoint Seiten nicht überprüft.

Unsere Testresultate zeigen, dass man Methode 4 benutzen sollte (Tabelle 1). Mit dieser Methode wurden bei den visuellen Kontrollen keine Unterschiede zum Original gefunden und alle erzeugt

Dateien waren auch wirklich konform mit dem PDF/A-1b Standard. Zudem erzeugte diese Methode die kleinsten Files, und die Rechenzeit dauerte auch für komplexe Dokumente (bis zu 40 Seiten) nur einige Sekunden.

Tabelle 1: Resultate für die vier Konvertierungsmethoden

Kriterien	1) In Acrobat öffnen und konvertieren	2) Acrobat PDFMaker	3) Save As PDF	4) Adobe PDF Drucker (Distiller)
Visuelle Überprüfung	Keine Fehler, jedoch teilweise ungenügende Bildqualität, sowie teilweise Text, der weder such- noch kopierbar war.	Ein File konnte nicht zuverlässig konvertiert werden.	Für vier der 10 Dateien waren einzelne Bilder nicht brauchbar, weil transparente Objekte undurchsichtig wurden.	Keine Fehler
Links und Verweise	Keine Fehler	Keine Fehler	Endnoten-Links gebrochen	Unterlegte Hyperlinks gebrochen, Endnoten-Links gebrochen
Adobe Preflight Überprüfung der Kompatibilität mit PDF /A-1b	Keine Fehler	Vier der 10 Files waren keine gültigen PDF/A-1b Dateien.	Eine der zehn Dateien war keine gültige PDF/A-1b Datei.	Keine Fehler
Grösse des PDF/A-1b Files	Bis vier Mal grösser als Original	Bis fünf Mal grösser als Original	Bis sechs Mal grösser als Original	Bis zwei Mal grösser als Original
Rechenzeit (Intel Core i7, 3.4 GHz)	Dauerte bis zu einigen Minuten	Dauerte bis zu einigen Minuten.	Dauerte höchstens wenige Sekunden	Dauerte höchstens wenige Sekunden

¹ Weitere Instruktionen auf folgender Webseite (Stand 29. 1. 2015): http://blogs.adobe.com/acrolaw/2007/01/pdfa_in_action . Acrobat Distiller konvertiert dabei von PostScript (*.ps) nach PDF. Auch auf dem kostenlosen OpenOffice 3 für Windows (Oracle) können Sie mit Distiller nach PDF/A-1b konvertieren.

² Adobe Acrobat Professional kann von ETH-Angehörigen bei <https://ides.ethz.ch> bestellt werden. Bitte kontaktieren Sie Ihren Informatik Support. Siehe auch Adobe Acrobat XI : Produktvergleich, http://www.adobe.com/ch_de/products/acrobat/product-comparison.html

³ Folgende Webseite gibt weitere Instruktionen (Stand 29/01/2015): <http://blogs.adobe.com/acrolaw/2011/05/using-save-as-to-to-conform-to-pdf>

⁴ Folgende Webseite gibt weitere Instruktionen (Stand 29/01/2015): <http://helpx.adobe.com/acrobat/using/creating-pdfs-pdfmaker-windows.html>

⁵ Folgende Webseite gibt weitere Instruktionen (Stand 29/01/2015): <https://support.office.com/en-us/article/Save-as-PDF-d85416c5-7d77-4fd6-a216-6f4bf7c7c110#bm11>