

Version: 1.4

Datum: 08. 08. 2005

Autor: Michael Horvath

michael.horvath@onb.ac.at

Datei: ONB_PDF-Empfehlungen_1-4

Empfehlungen zum Erzeugen archivierbarer Dateien im Format PDF

Einführung

Das von der Firma Adobe entwickelte Portable Document Format PDF, dessen Spezifikation öffentlich und frei zugänglich vorliegt⁽¹⁾, hat in den vergangenen Jahren breite Akzeptanz gefunden. Das Format bietet den Vorteil, die innere Struktur und das äußere Erscheinungsbild digitaler Dokumente auch über technologische Grenzen hinweg beschreiben zu können.

PDF ist jedoch nicht vorrangig für Zwecke der Langzeitarchivierung digitaler Inhalte entwickelt worden. Die Spezifikation des Formats erlaubt eine Fülle an multimedialen, interaktiven und anderen Komponenten, deren langfristige Erhaltung nicht oder nur mit unverhältnismäßig hohem Aufwand zu gewährleisten wäre.

Die folgenden Empfehlungen der Österreichischen Nationalbibliothek hinsichtlich der Ablieferung von digitalen Dokumenten im Format PDF stellen daher Einschränkungen des Formats dar, welche die langfristige Archivierbarkeit von digitalen Dokumenten unterstützen sollen. Sie orientieren sich vor allem an dem zur Zeit in Ausarbeitung befindlichen ISO-Standard ISO/CD 19005-1 „Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF (PDF/A)“⁽²⁾, der auf der Version 1.4 des PDF-Formats basiert.

Die Empfehlungen sollten so weit eingehalten werden, wie es dem Ablieferer eines digitalen Dokuments technisch und organisatorisch möglich ist, wobei die Wahrscheinlichkeit, den Inhalt des Dokuments in der vom Autor beabsichtigten Form langfristig benutzbar halten zu können, umso größer ist, je besser dieses den Empfehlungen entspricht.

Hinweise

Wenn in der Folge von einem **PDF-Anzeigeprogramm** die Rede ist, so wird damit eine Software bezeichnet, die das PDF-Format zumindest lesen und anzeigen kann, z. B. der über das Internet kostenlos zum Download angebotene Adobe Reader. **PDF-Schreibprogramm** wird eine Software genannt, die das Erzeugen von PDF-Dateien ermöglicht, z. B. Adobe Acrobat Distiller oder AFPL GhostScript.

Weiters werden im vorliegenden Dokument zwei Arten von Empfehlungen unterschieden, die durch folgende Symbole gekennzeichnet sind:

- ▶ Die Einhaltung der Empfehlung trägt wesentlich zur langfristigen Archivierbarkeit eines digitalen Dokuments bei.
- > Die Einhaltung der Empfehlung unterstützt die langfristige Archivierbarkeit eines digitalen Dokuments, sie ist jedoch nicht unbedingt erforderlich.

Empfehlungen

Dateiformat

- ▶ Die Datei muss gemäß der von Adobe veröffentlichten Spezifikation für das Format PDF in Version 1.4 erstellt sein. Dateien, die in höheren Versionen des Formats erstellt worden sind, enthalten möglicherweise Komponenten, die für eine langfristige Archivierung nicht geeignet sind.

Konvertierung

- ▶ Wird die Datei von einem anderen Dateiformat, z. B. Microsoft Word, in PDF konvertiert, dann muss dabei ein Programm verwendet werden, das die Erstellung von Dateien in Version 1.4 von PDF erlaubt, etwa das von Adobe entwickelte Programm Acrobat Distiller in Version 5.0 oder höher oder das frei verfügbare AFPL GhostScript in Version 7.0 oder höher.

Detaillierte Angaben zu Einstellungen in ausgewählten PDF-Schreibprogrammen, die die praktische Umsetzung der vorliegenden Empfehlungen bestmöglich unterstützen, wird die Österreichische Nationalbibliothek jeweils in einem eigenen Dokument publizieren.

Dateiprofile

Profile bieten die Möglichkeit, eine PDF-Datei bei deren Erzeugung auf einen bestimmten Verwendungszweck hin zu optimieren. Einzelne Profile schließen einander nicht aus, so dass eine Datei auch unter Berücksichtigung von mehr als einem Profil erzeugt werden kann.

- ▶ **Tagged PDF** schreibt einer PDF-Datei eine definierte interne Struktur vor, um deren Inhalte mit geeigneten Werkzeugen auslesen und für andere Zwecke weiterverarbeiten zu können. Mögliche Anwendungen sind die Übertragung von Inhalten in Dateiformate wie XML, HTML oder RTF. Dateien sind so zu erstellen, dass sie Tagged PDF entsprechen.
- > **Linearisiertes PDF** ist entwickelt worden, um die Seiten einer PDF-Datei innerhalb von Netzwerkumgebungen, z. B. im World Wide Web, möglichst rasch anzeigen zu können. Dieses Profil beeinträchtigt die langfristige Archivierbarkeit einer Datei nicht, seine Verwendung beim Generieren von PDF-Dateien wird empfohlen.

Metadaten

- > Wie im Entwurf zum PDF/A- Standard gefordert, soll eine archivierbare Datei möglichst selbstdokumentierend sein, was unter anderem durch die Speicherung von Metadaten auf Basis der von Adobe entwickelten eXtensible Metadata Platform – XMP⁽³⁾ in der Datei selbst gewährleistet wird. Zur Umsetzung dieser Forderung wird empfohlen, alle über das verwendete PDF-Schreibprogramm eingebbaren Metadaten-Felder wie Autor, Titel etc. vor Speicherung der Datei zu befüllen.

Verschlüsselung

- ▶ Der lesende oder schreibende Zugriff auf eine Datei darf keinerlei Beschränkungen unterliegen, insbesondere dürfen keine Passwörter verwendet werden, um die Anwendung bestimmter Funktionen auf die Datei zu unterbinden. Damit ist z. B. gewährleistet, dass die Datei zu Zwecken der Archivierung in andere Formate übertragen werden kann, wenn dies zur Erhaltung ihres Inhalts notwendig ist.

Authentisierung

- > Die Verwendung von Verfahren zum Nachweis der Authentizität einer PDF-Datei, beispielsweise von digitalen Signaturen, ist erlaubt. Die dauerhafte Erhaltung dieser Merkmale kann jedoch nicht garantiert werden.

Text

- ▶ Jede im Text der Datei verwendete Schriftart muss in die Datei eingebettet sein, die Datei muss also neben dem eigentlichen Text auch die grafischen Beschreibungen aller darin verwendeten Schriftarten enthalten. Diese Maßnahme gewährleistet, dass PDF-Anzeigeprogramme die Datei stets in der vom Autor/von der Autorin beabsichtigten Form darstellen können, ohne auf Ersatz-Schriftarten zurückgreifen zu müssen, welche die visuelle Darstellung der Datei verändern könnten.

- ▶ Auch die von Adobe zur Verfügung gestellten Standard-Schriftarten (**Base-14-Schriften**: Courier, Courier-Bold, Courier-BoldOblique, Courier-Oblique, Helvetica, Helvetica-Bold, Helvetica-BoldOblique, Helvetica-Oblique, Symbol, Times-Bold, Times-BoldItalic, Times-Italic, Times-Roman, Zapf-Dingbats) sind immer einzubetten.

- > Um die Größe einer Datei möglichst gering zu halten, ist die Einbettung von Untergruppen von Schriftarten möglich, also von Teilmengen des verfügbaren Zeichenvorrats, solange jede Untergruppe alle Zeichen der Schriftart enthält, die in der Datei tatsächlich verwendet werden. Details zur Vorgehensweise bei der Einbettung von Schriften in PDF-Dateien werden in den jeweiligen Dokumenten zu Einstellungen in PDF-Schreibprogrammen zu finden sein.

- > Grundsätzlich sollten nur öffentlich verfügbare Schriftarten verwendet werden, die keinerlei von den Inhabern an deren Rechten festgelegten Beschränkungen unterliegen. Die mit den Microsoft-Office-Produkten ausgelieferten Schriftarten können z. B. ohne Einschränkungen verwendet werden.

- > Möglichst alle in der Datei enthaltenen Zeichen sollen in maschinenlesbar codierter Form vorliegen und nicht als digitalisierte Bilder von Zeichen, damit sie für Suchfunktionen zur Verfügung stehen.

- > Für Textabschnitte, die Zeichen enthalten, welche über den Zeichensatz ISO-8859-1 für US-amerikanische und westeuropäische Sprachen (ISO-Latin-1)⁽⁴⁾ hinausgehen, soll der Zeichensatz UNICODE⁽⁵⁾ verwendet werden. Ein Autor/eine Autorin soll Texte also entweder in ISO-Latin-1 oder in UNICODE codieren, was konkret durch die Auswahl entsprechender Schriftarten realisiert wird, z. B. Arial Unicode MS.

Grafik

- ▶ Enthält die Datei grafische Darstellungen, dann müssen diese auf geräteunabhängig spezifizierten Farbräumen wie sRGB (Standard Red – Green – Blue)⁽⁶⁾ basieren, um eine voraussagbare Ausgabe zu gewährleisten. Für nähere Angaben zur praktischen Umsetzung dieser Empfehlung sei auf die Dokumente zu Einstellungen in PDF-Schreibprogrammen verwiesen.
- ▶ Der als visueller Effekt in Version 1.4 von PDF eingeführte sogenannte Transparenzschlüssel, der es ermöglicht, einander überlappende oder dreidimensional wirkende Grafiken zu erzeugen, darf generell nicht verwendet werden.

Einbindungen

- ▶ Alle für die Darstellung des Dokuments notwendigen Inhalte müssen in der Datei selbst enthalten sein, so dass kein Laden von Datenströmen aus externen Quellen erforderlich ist. Durch die so erzielte strukturelle Unabhängigkeit des Dokuments bleibt die Automatisierbarkeit der damit verbundenen Arbeitsflüsse, etwa der Übertragung auf einen anderen Speicherort oder in ein anderes Dateiformat, gewährleistet.
- ▶ Die Datei darf kein eingebundenes Objekt enthalten, dessen Darstellung ein externes Anwendungsprogramm erfordern würde. Auf diese Weise wird sichergestellt, dass ein Benutzer/eine Benutzerin ausschließlich ein PDF-Anzeigeprogramm benötigt, um auf den gesamten Inhalt der Datei zugreifen zu können.

Audio/Video

- ▶ Aus den im vorigen Absatz angeführten Gründen darf eine Datei auch weder Audio- noch Video-Datenströme enthalten.

Verknüpfungen (Links)

- > Interne Verknüpfungen, die auf Sprungmarken innerhalb der Datei, wie etwa Überschriften, verweisen, sind zulässig.
- > Externe Verknüpfungen, die auf Sprungmarken außerhalb der Datei verweisen, beispielsweise Hyperlinks zu Ressourcen im Internet, sollen nach Möglichkeit so aufgebaut sein, dass alle symbolischen Adressen dieser Marken, also Dateipfade, Uniform Resource Locators (URL) oder Persistent Identifiers, im Text der Datei enthalten sind und ohne zusätzliche Maßnahmen am Bildschirm angezeigt bzw. auf Papier ausgedruckt werden können. Ein Beispiel: Statt „[Website der ÖNB](#)“ mit dahinter liegendem Link sollte geschrieben werden: „Website der ÖNB (<http://www.onb.ac.at/>)“.

Kommentare

- ▶ Die Verwendung von Kommentaren ist zulässig, solange diese keine Audio- oder Video-Datenströme und keine sonstigen Dateianhänge enthalten.

Ausführbare Aktionen

- ▶ Die Datei darf keine Aufrufe von ausführbaren Anweisungsfolgen, wie z. B. Scripts, beinhalten, insbesondere darf weder innerhalb von Feldern in Formularen noch an anderer Stelle Javascript eingebunden sein.

Referenzen

- (1) "PDF Reference — Third Edition — Adobe Portable Document Format — Version 1.4"
<http://partners.adobe.com/public/developer/en/pdf/PDFReference.pdf>
- (2) PDF/A: [http://www.aiim.org/documents/standards/ISO_19005-1_\(E\).doc](http://www.aiim.org/documents/standards/ISO_19005-1_(E).doc)
bzw. als kostenpflichtiger ISO-Standard:
"ISO/PRF 19005-1 — Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)"
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=38920>
- (3) XMP (Adobe eXtensible Metadata Platform):
<http://www.adobe.com/products/xmp/>
- (4) ISO-Latin-1: <http://anubis.dkuug.dk/JTC1/SC2/WG3/docs/n411.pdf>
bzw. als kostenpflichtiger ISO-Standard:
"ISO/IEC 8859-1:1998 — Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1"
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=28245>
- (5) UNICODE: <http://www.unicode.org/versions/Unicode4.1.0/>
bzw. als kostenpflichtiger ISO-Standard:
"ISO/IEC 10646:2003 — Information technology — Universal Multiple-Octet Coded Character Set (UCS)"
<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39921>
- (6) sRGB: <http://www.srgb.com/srgb.html> bzw. als IEC-Standard:
"IEC 61966-2-1 – Ed. 1.0 – Bilingual — Multimedia systems and equipment — Colour measurement and management — Part 2-1: Colour management — Default RGB colour space — sRGB"
<http://webstore.iec.ch/webstore/webstore.nsf/artnum/025408?opendocument>