

Empfehlungen des Bundesarchivs zur Anwendung der verschiedenen PDF/A-Versionen

Bei der dauerhaften Aufbewahrung von digitalen Unterlagen empfiehlt das Bundesarchiv, diese zum einen im Ausgangsformat und zum anderen soweit notwendig als originalgetreue Repräsentation in Form eines PDF/A-Dokuments aufzubewahren, um diese dauerhaft lesbar zu halten. Aus diesem Grund wurden die Eigenschaften der einzelnen PDF/A-Versionen 1, 2 und 3 sowie deren jeweiligen Konformitätsstufen (a = „accessible“, b = „basic“, u = „unicode“) im Hinblick auf die Tauglichkeit für eine dauerhafte Aufbewahrung ausgewertet.

Archivische Wertung bzw. Empfehlung

Gemäß der nachfolgenden Auswertung ist die PDF/A-Version 2 der Version 1 vorzuziehen, da hier durch die Möglichkeit Ebenen und Transparenz abzubilden, wesentlich geringere Informationsverluste auftreten. Die PDF/A-Version 3 ist aus archivfachlicher Sicht nicht zu empfehlen, da die Eigenschaft beliebige Dateiformate einbinden zu können die uneingeschränkte dauerhafte Lesbarkeit der Datei gefährdet und die eingebundenen Dateien ggf. durch sehr aufwändige Techniken wie z. B. Emulation wieder lesbar gemacht werden müssen, wobei deren Integrität und Layoutstabilität nicht gewährleistet werden kann. Zudem können die eingebundenen Dateiformate per Formaterkennung nicht analysiert werden, was ggf. eine Migration erschweren würde. Weiterhin müssen ggf. Spezialapplikationen zur Darstellung der eingebundenen Dateien angewendet werden, was eine Abhängigkeit der Programme von der Nutzungsumgebung bedeuten kann.

Hinsichtlich der Konformitätsstufen, sollte die Stufe „a“ bevorzugt werden, da diese über die Extrahierbarkeit und Strukturierbarkeit des Textes die Barrierefreiheit bestmöglich gewährleistet. Die Stufe „a“ ist vor allem bei der Konvertierung von genuin digitalen Dokumenten anzuwenden. Sollte diese Stufe nicht umsetzbar sein, so ist die Stufe „u“ der Stufe „b“ vorzuziehen, da die Textinhalte in Unicode vorliegen müssen und der Text somit durchsuchbar und extrahierbar bleibt. Diese Stufe kann u. a. bei digitalisierten Dokumenten mit OCR-Texterkennung angewendet werden. Die Barrierefreiheit wird hierbei zumindest optional gewährleistet. Die Stufe „b“ garantiert lediglich eine eindeutige visuelle Reproduzierbarkeit ohne die Inhalte dabei durchsuchbar zu halten, was Recherchen nach Dokumentinhalten verhindert.

Das Bundesarchiv empfiehlt die folgenden PDF/A-Versionen und Konformitätsstufen in priorisierter Reihenfolge:

1. PDF/A-2a
2. PDF/A-2u
3. PDF/A-2b

4. PDF/A-1a
5. PDF/A-1b

Eine Konvertierung zu PDF/A-3 ist nicht zu empfehlen.

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)
Telefon: 03018/7770-105
E-Mail: behoerdenberatung@bundesarchiv.de

Auswertung der PDF/A-Versionen

Vorbemerkung

Vorwegzunehmen ist, dass je nach Ausgangsdokument bzw. Verwendungszweck ggf. entsprechend andere PDF/A-Versionen angewandt werden müssen, da diese jeweils für bestimmte Anforderungen optimiert sind.

Dabei lassen sich PDF/A-Dokumente auf unterschiedlichen Wegen erzeugen:

- aus gescannten Vorlagen
- über eine direkte Konvertierung der Ursprungsdaten
- über einen Export aus dem Programm, mit dem das Ursprungsdokument erstellt wurde
- über einen Zwischenschritt, der aus einer PDF-Datei schließlich PDF/A erzeugt
- über Druckausgabeformate bzw. Druckdatenströme wie GDI, PCL, Post-Script, AFP und XPS
- und im Massenverfahren unter Verwendung eines gesonderten Konvertierungswerkzeugs

Alle PDF/A-Versionen wurden so spezifiziert, dass man mit Hilfe der vorgegebenen Eigenschaften digitale Dokumente dauerhaft aufbewahren und layoutstabil repräsentieren kann. So ist allen Versionen gemein, dass sämtliche Ressourcen wie zum Beispiel Bilder, Grafiken, Schriftzeichen etc. in die Datei eingebettet sein müssen, um diese unabhängig von der verwendeten Nutzungsumgebung originalgetreu darstellen zu können. Zudem werden Farben geräteneutral definiert und über ICC-Farbprofile wiedergegeben. Auch gibt es einen festen Standard zur Darstellung der Dokumentmetadaten im XMP-Format. Weiterhin sind in allen drei PDF/A-Versionen interaktive Elemente, Java-Script und die Verwendung von Passwörtern verboten. Das Einbetten digitaler Signaturen oder die Verwendung von Hyperlinks stellen für die dauerhafte Aufbewahrung der digitalen Dokumente hingegen kein Problem dar. Ob ein externer Link zu einem gültigen Ziel führt, ist für die Einhaltung des PDF/A-Standards nicht relevant. Die Links sollten jedoch so aufgebaut sein, dass Dateipfade, URLs oder Persistent Identifier als solche ausgeschrieben werden.

Nachfolgend soll auf die Unterschiede der einzelnen PDF/A-Versionen eingegangen werden, um deren Verwendbarkeit für die dauerhafte Aufbewahrung digitaler Dokumente aufzuzeigen.

PDF/A-1

Eigenschaft	Archivische Wertung
basiert auf PDF 1.4	negativ
alle Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dokumentmetadaten (XMP-Format)	positiv
PDF-Ebenen sind verboten	negativ
transparente Elemente sind verboten	negativ

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)
Telefon: 03018/7770-105
E-Mail: behoerdenberatung@bundesarchiv.de

interaktive Elemente sind verboten	positiv
einige Bildkompressionsverfahren (LZW, JPEG2000) sind verboten, Kompressionsverfahren wie JBIG2 und JPEG sind erlaubt	negativ
Java-Script ist verboten	positiv
Verwendung von Passwörtern ist verboten	positiv
Einbindung von ganzen Dateien ist unzulässig	positiv
Einbettung von digitalen Signaturen möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	neutral

PDF wurde 2008 in der Version 1.7 als ISO-Standard zertifiziert. PDF/A-1 basiert jedoch auf der PDF-Version 1.4 was bedeutet, dass die Eigenschaften welche zwischen diesen beiden Versionen liegen nicht umgesetzt werden können. Die PDF/A-Version 1 kann in den Konformitätsstufen a und b dargestellt werden, welche auf Seite 5 genauer erläutert werden.

In der PDF/A-Version 1 ist die Darstellung von Ebenen verboten, was bei der Konvertierung jedoch zu Informationsverlusten (z.B. bei überblendeten Verfügungen, Unterschriften etc.) führt. Diese Eigenschaft ist daher aus archivischer Sicht kritisch zu betrachten, da im Behördenschriftgut verstärkt mit solchen Ebenen gearbeitet wird, welche folglich darstellbar sein müssen. Auch ist die Darstellung transparenter Elemente (wie z. B. Wasserzeichen) nicht erlaubt, was ebenfalls bei der Konvertierung von Dokumenten, welche solche transparenten Elemente enthalten, zu Informationsverlusten führt. Die Eigenschaft, dass einige Bildkompressionsverfahren wie z. B. LZW, JPEG2000 verboten, andere Kompressionsverfahren wie JBIG2 und JPEG hingegen erlaubt sind, ist als kritisch zu betrachten, da man mit der JPEG2000-Kompression im Vergleich zur Dateigröße eine bessere Qualität erzeugen könnte. Generell führen jedoch alle Kompressionsverfahren durch das Herunterrechnen der Auflösung zu Informationsverlusten, welche als kritisch zu betrachten sind, zumal hierbei die Integrität beeinträchtigt wird. Daher wird empfohlen bei vorgangsrelevanten Bildern und Grafiken diese in einem für die dauerhafte Aufbewahrung geeigneten Bildformat zu konvertieren und diese Repräsentation mit zu verakten.

PDF/A-2

Eigenschaft	Archivische Wertung
basiert auf PDF 1.7	positiv
alle Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dokumentmetadaten (XMP-Format)	positiv
PDF-Ebenen sind erlaubt	positiv
transparente Elemente sind erlaubt	positiv
interaktive Elemente sind verboten	positiv
Bildkomprimierung mit JPEG2000 erlaubt	negativ
Java-Script ist verboten	positiv
Verwendung von Passwörtern ist verboten	positiv
Einbindung von PDF/A-1 und 2 konformen Dateien ist erlaubt	positiv

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)

Telefon: 03018/7770-105

E-Mail: behoerdenberatung@bundesarchiv.de

Einbettung von Open Type-Fonts ist erlaubt	positiv
Einbettung von digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	neutral

Die PDF/A-Version 2 basiert auf der PDF Version 1.7 und bietet hinsichtlich der Langzeitaufbewahrung digitaler Dokumente bezüglich der Vermeidung von Informationsverlusten entscheidende Vorteile gegenüber der PDF/A-Version 1. So ist im Gegensatz zu PDF/A-1 die Darstellung von Ebenen und transparenten Elementen erlaubt, was die in Version 1 beschriebenen Informationsverluste bei der Konvertierung umgeht. Zudem kann mit der in Version 2 erlaubten JPEG2000-Komprimierung im Vergleich zur Dateigröße eine bessere Qualität erreicht werden, auch wenn diese wiederum verlustbehaftet ist. Im Vergleich zu Version 1 können PDF/A-1 und 2 konforme Dateien eingebunden werden. Zugehörige Dateien anderer Formate sind zwar als Anhang ersichtlich, doch werden diese Dateien nicht tatsächlich angehängt und können deshalb weder geöffnet noch gespeichert werden. In Version 2 dürfen zudem OpenType-Fonts eingebettet werden, welche zwar ein eingetragenes Warenzeichen von Microsoft sind, die Technik jedoch uneingeschränkt auf andere Betriebssysteme übertragen werden kann, womit einer dauerhaften Aufbewahrung solcher Dokumente nichts entgegensteht. Durch die in PDF/A-2 und 3 erlaubte Eigenschaft des „optional content“, können diverse Elemente ein- und ausgeblendet werden, was eine eindeutige visuelle Reproduktion hauptsächlich nur noch am Computer ermöglicht. Die PDF/A-Version 2 kann in den Konformitätsstufen a, b und u vorliegen, welche auf Seite 5 genauer erläutert werden.

PDF/A-3

Eigenschaft	Archivische Wertung
basiert auf PDF 1.7	positiv
alle Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dokumentmetadaten (XMP-Format)	positiv
PDF-Ebenen sind erlaubt	positiv
transparente Elemente sind erlaubt	positiv
interaktive Elemente sind verboten	positiv
Bildkomprimierung mit JPEG2000 erlaubt	negativ
Java-Script ist verboten	positiv
Verwendung von Passwörtern ist verboten	positiv
Einbindung von Dateien beliebiger Dateiformate ist erlaubt	negativ
Einbettung von Open Type-Fonts sind erlaubt	positiv
Einbettung von digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	neutral

Der Unterschied zwischen PDF/A-2 und 3 besteht darin, dass in Version 3 Dateien beliebiger Da-

Ansprechpartner:

teiformate eingebunden werden können, die nicht PDF/A-1 und 2 konform sein müssen. Dies ist nur sinnvoll, wenn neben der langfristig lesbaren PDF/A-Version eines Dokuments die ursprüngliche Datei mit maschinenlesbaren Daten erhalten werden soll, um deren Inhalte weiterverarbeiten zu können. Jedoch ist diese Eigenschaft hinsichtlich der dauerhaften Aufbewahrung digitaler Unterlagen als negativ zu bewerten, da neben der Vielzahl der enthaltenen Dateiformate in solchen Dokumenten deren Lesbarkeit nicht auf Dauer garantiert werden kann. Eingebundene Dateien, welche nicht im PDF-Format vorliegen, können demnach auch nicht im PDF-Viewer dargestellt werden und erfordern entsprechende Spezialapplikationen. Sollten die eingebetteten Dateiformate nicht in einem für die langfristige Aufbewahrung geeigneten Format vorliegen, müssen diese ggf. über die sehr aufwändige Emulation wieder lesbar gemacht werden. Die PDF/A-Version 3 kann in den Konformitätsstufen a, b und u vorliegen, welche nachfolgend genauer erläutert werden.

Konformitätsstufen

Die Anwendung der jeweiligen Konformitätsstufen ist abhängig vom Eingangsmaterial und dem beabsichtigten Verwendungszweck. Nachfolgend wird zwischen folgenden Konformitätsstufen unterschieden:

a = „accessible“

Eigenschaft
Textinhalte müssen extrahierbar sein
Struktur muss die natürliche Leseabfolge eindeutig abbilden
hohe Anforderungen an verwendete Schriftzeichen
Barrierefreiheit wird bestmöglich umgesetzt

Diese Konformitätsstufe sollte bevorzugt bei der Konvertierung aus digitalen Ursprungsdokumenten angewendet werden. Eine Anwendung dieser Konformitätsstufe bei digitalisierten Dokumenten ist nur möglich, wenn eine OCR-Texterkennung stattgefunden hat und in die PDF/A-Datei eingebettet wurde. Neben der Darstellung des Textes in Unicode, bleibt die Struktur des Dokumentes und die natürliche Lesereihenfolge erhalten, sodass das Dokument mit Screenreader-Programmen vorgelesen werden kann und somit bestmögliche Barrierefreiheit bietet.

b = „basic“

Eigenschaft
eindeutige visuelle Reproduzierbarkeit gewährleistet
Extrahierbarkeit von Textinhalten nur optional
Barrierefreiheit nur optional gewährleistet

Diese Konformitätsstufe eignet sich bevorzugt bei der Konvertierung von digitalisierten Unterlagen. Auch kann hier eine OCR-Texterkennung mit entsprechender Einbettung in die PDF/A-Datei erfolgen.

u = „unicode“

Eigenschaft
eindeutige visuelle Reproduzierbarkeit gewährleistet

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)
Telefon: 03018/7770-105
E-Mail: behoerdenberatung@bundesarchiv.de

alle Textinhalte müssen in Unicode vorliegen
Barrierefreiheit nur optional gewährleistet

Bei dieser Konformitätsstufe müssen alle Textinhalte in Unicode vorliegen, so dass der gesamte Text durchsucht und extrahiert werden kann.

Validierung

Die Validierung von PDF/A-Dateien ist ein wichtiger Schritt in der Qualitätssicherung der selbigen, da die Erzeugung valider PDF/A-Dateien eine große Herausforderung hinsichtlich der Fehleranfälligkeit bei der Konvertierung darstellt. Vor allem automatische Datenfelder oder interaktive Elemente bereiten bei der Konvertierung große Schwierigkeiten. Eine Validierung der PDF/A-Dokumente wird daher empfohlen.

Eine Validierung der PDF/A-Dateien sollte stattfinden:

- nach der Erstellung
- nach dem Empfang
- vor dem Versand bzw. der Verteilung
- vor der Archivierung
- nach bestimmten Arbeitsprozessen (z. B. Einfügen von Seiten in die PDF/A-Datei)

Sollten bei der Validierung Verstöße gegen die Regeln des PDF/A-Standards erkannt werden, lässt sich die Datei meist mit Hilfe einer entsprechenden Software reparieren. Andernfalls muss die Datei mit modifizierten Einstellungen als PDF-Dokument erneut erzeugt werden.

Verwendete Quellen

- Bundesamt für Sicherheit in der Informationstechnik (BSI): Technische Richtlinie 03125. Beweiswerterhaltung kryptographisch signierter Dokumente. Anlage TR-ESOR-F: Formate. 2015.
- Debenath, O.; Merzaghi, M.; Röthlisberger, C. (Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST)): PDF/A-2 und PDF/A-3: Was ist neu? Eine Studie der KOST. 2013.
- IANUS: IT-Empfehlung für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften. PDF-Dokumente. 2017
- Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder: Empfehlungen zur Anwendung von PDF/A bei der Aufbewahrung elektronischer Unterlagen in Behörden und Gerichten.
- Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): PDF/A: Validatoren. 2010.
- Oettler, Alexandra (PDF/A association): PDF/A kompakt 2.0. PDF für die Langzeitarchivierung. 2013.

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)
Telefon: 03018/7770-105
E-Mail: behoerdenberatung@bundesarchiv.de

- Seggern, Dietrich von (PDF/A association): PDF und PDF/A Validierung – im Brennpunkt Schriften. 2017.
- Vogeler, Bernhard; Lacher, Ingrid (zmi): Barrierefreie PDF-Dokumente aus Word-Quellen. Dokumente in Word erfassen und in barrierefreie PDF-Dokumente transformieren. 2015.
- Vogeler, Bernhard; Lacher, Ingrid (zmi): Barrierefreie PDF-Dokumente aus Word-Quellen. Weiterführende Aspekte. 2015.

Ansprechpartner:

Referat B5 (Übernahme Digitaler Unterlagen)
Telefon: 03018/7770-105
E-Mail: behoerdenberatung@bundesarchiv.de