

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 möglichst originalgetreue Repräsentation in Form einer PDF/A-Datei aufzubewahren, um diese dauerhaft lesbar zu halten. Aus diesem Grund wurden in der vorliegenden Handreichung die Eigenschaften der einzelnen PDF/A-Versionen 1, 2, 3 und 4 sowie deren jeweilige Konformitätsstufen (a = „accessible“, b = „basic“, u = „unicode“, e = „engineering“ und f = „flat“) im Hinblick auf die Tauglichkeit für eine dauerhafte Aufbewahrung ausgewertet sowie Empfehlungen für deren Einsatz ausgesprochen.

In der vorliegenden Handreichung wird bewusst von „(Text-)Datei“ anstatt von „Dokument“ gesprochen, da bei der elektronischen Aktenführung Dokumente oftmals nur logische Container darstellen, welche wiederum mehrere Dateien enthalten können, die deren Inhalt darstellen. Somit entsprechen die digitalen (Text-)Dateien den analogen Dokumenten.

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de

Inhalt

1. Zusammenfassung: Archivische Wertung bzw. Empfehlung	3
2. Auswertung der PDF/A-Versionen	5
2.1. Vorbemerkung	5
2.2 PDF/A-1	6
2.3 PDF/A-2	8
2.4 PDF/A-3	9
2.5 PDF/A-4	11
2.6 Gegenüberstellung der PDF/A-Versionen	13
3. Auswertung der PDF/A-Konformitätsstufen	14
3.1 Die PDF/A-Konformitätsstufen	14
3.2 Anwendung der PDF/A-Konformitätsstufen	15
4. Wie erkennt man eine PDF/A-Datei?	17
5. Konvertierung	18
6. Validierung	21
7. Barrierefreiheit durch PDF/UA	22
8. Verwendete Quellen	23

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de

1. Zusammenfassung: Archivische Wertung bzw. Empfehlung

Die Konvertierung von (Text-)Dateien in PDF/A dient in erster Linie ihrer layoutstabilen und rechtssicheren Aufbewahrung und somit der Nachvollziehbarkeit des Verwaltungshandelns. Bei der nachfolgenden Auswertung handelt es sich um eine archivische Einschätzung der Eignung der verschiedenen PDF/A-Versionen und -Konformitätsstufen, verbunden mit einer Empfehlung für deren Anwendung. Das Bundesarchiv besteht hierbei ausdrücklich nicht auf eine Konvertierung von Dateien in PDF/A vor einer Langzeitspeicherung oder Abgabe an das Archiv. Es spricht jedoch Empfehlungen aus, die den Behörden und Institutionen helfen sollen, Dateien über die Zeit der gesetzlich vorgeschriebenen Aufbewahrungsfrist lesbar und verkehrsfähig zu halten. Die Konvertierung und entsprechende Qualitätssicherung liegen in der Verantwortung der schriftgutführenden Stelle – das Bundesarchiv steht hierbei beratend zur Verfügung.

3

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. Eine Konvertierung in die PDF/A-Versionen 3 und 4 ist aus archivfachlicher Sicht nicht zu empfehlen, sofern die eingebetteten Dateien nicht PDF/A-konform sind oder in einem anderen langzeitspeicherfähigen Format vorliegen. Eingebettete proprietäre Dateiformate gefährden die uneingeschränkte dauerhafte Lesbarkeit der (Text-)Datei, welche ggf. durch sehr aufwändige Techniken wie z.B. Emulation wieder lesbar gemacht werden muss, wobei deren Integrität und Layoutstabilität nicht gewährleistet werden kann. Zudem können die eingebundenen Dateiformate per Formaterkennung nicht analysiert werden, was ggf. eine spätere Migration erschwert. 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

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

Extrahierbarkeit und Strukturierbarkeit des Textes die Barrierefreiheit bestmöglich gewährleistet. Die Stufe „a“ ist vor allem bei der Konvertierung von genuin digitalen (Text-)Dateien 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 einer Texterkennung mittels OCR 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 deren Inhalten verhindert. Die Anwendung der Konformitätsstufen „e“ und „f“ ist nur bei der PDF/A-Version 4 möglich und bedingt sich durch den beabsichtigten Zweck.

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

4

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 oder PDF/A-4 ist nur unter bestimmten Voraussetzungen sinnvoll. Eine Rücksprache mit dem zuständigen Archiv ist in diesen Fällen dringend angeraten.

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de

2. Auswertung der PDF/A-Versionen

2.1. Vorbemerkung

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

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

- aus gescannten Vorlagen
- über eine direkte Konvertierung der Ursprungsdaten
- über einen Export aus dem Programm, mit dem die Ursprungsdatei 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.

5

Alle PDF/A-Versionen wurden so spezifiziert, dass man mit Hilfe der vorgegebenen Eigenschaften (Text-)Dateien dauerhaft aufbewahren und layoutstabil repräsentieren kann. So ist allen Versionen gemein, dass sämtliche Ressourcen wie zum Beispiel Bilder, Grafiken, Schriftzeichen etc. in die PDF/A-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 Dateimetadaten im XMP-Format. Weiterhin sind in allen vier PDF/A-Versionen interaktive Elemente sowie die Verwendung von Passwörtern verboten. Das Einbetten digitaler Signaturen oder die Verwendung von Hyperlinks stellen für die dauerhafte Aufbewahrung der (Text-)Dateien 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.

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

Nachfolgend soll auf die Unterschiede der einzelnen PDF/A-Versionen eingegangen werden, um deren Verwendbarkeit für die dauerhafte Aufbewahrung von (Text-)Dateien aufzuzeigen.

2.2 PDF/A-1

Eigenschaft	Archivische Wertung
basiert auf PDF 1.4	negativ
alle externen Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dateimetadaten (XMP-Format)	positiv
PDF-Ebenen sind verboten	negativ
transparente Elemente sind verboten	negativ
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	(negativ)
Einbettung von digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	positiv

6

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.

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 und somit der Nachvollziehbarkeit des Verwaltungshandelns widerspricht. Diese Eigenschaft ist daher aus archiverischer Sicht kritisch zu betrachten, da im Behördenschriftgut verstärkt mit solchen Ebenen gearbeitet wird, welche folglich darstellbar sein müssen.

Ansprechpartner:

Auch ist die Darstellung transparenter Elemente (wie z.B. Wasserzeichen, Grafiken) nicht erlaubt, was ebenfalls bei der Konvertierung von (Text-)Dateien, welche solche transparenten Elemente enthalten, zu Informationsverlusten führt, da die transparenten Pixel durch farbige Pixel ersetzt werden und Grafiken oder Wasserzeichen somit Text überdecken können, welcher dadurch unleserlich wird.

Die Eigenschaft, dass einige Bildkompressionsverfahren wie z.B. LZW und 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 aus archivischer Sicht als kritisch zu betrachten sind, zumal hierbei die Integrität der Datei beeinträchtigt wird, auch wenn die Datenverluste für den Menschen meist kaum wahrnehmbar sind. Daher wird empfohlen, vorgangsrelevante Bilder und Grafiken in ein für die dauerhafte Aufbewahrung geeignetes Bildformat zu konvertieren und diese Repräsentation mit zu verakten, sollte sich für eine Konvertierung in PDF/A-1 entschieden werden.

7

Dass bei PDF/A-1 konvertierten Dateien die Einbettung weiterer Dateien unzulässig ist, ist in solchen Fällen negativ zu bewerten, in denen die Einbettung weiterer Dateien als Bestandteil der Ursprungsdatei zwingend notwendig ist und es somit zu Informationsverlusten kommen kann. In Fällen, in denen keine Einbettung weiterer Dateien vorgesehen ist, ist diese Eigenschaft zu vernachlässigen.

Die PDF/A-Version 1 kann in den Konformitätsstufen a und b dargestellt werden.

Ansprechpartner:

2.3 PDF/A-2

Eigenschaft	Archivische Wertung
basiert auf PDF 1.7	positiv
alle externen Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dateimetadaten (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
Einbettung von digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	positiv

8

Die PDF/A-Version 2 basiert auf der PDF-Version 1.7 und bietet hinsichtlich der Langzeitaufbewahrung von (Text-)Dateien 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 bei einer Bildkomprimierung erreicht werden. Die archivische Wertung als „(negativ)“ liegt darin begründet, dass alle Kompressionsverfahren durch das Herunterrechnen der Auflösung zu Informationsverlusten führen, welche die Integrität der Datei beeinträchtigen können, auch wenn die Datenverluste für den Menschen meist kaum wahrnehmbar sind. Im Vergleich zu Version 1 können PDF/A-1 und 2 konforme Dateien eingebunden werden. Durch die Bedingung der

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

PDF/A-Konformität der eingebundenen Dateien ist auch deren Langzeitspeicherfähigkeit gewährleistet. Zugehörige Dateien anderer Formate sind zwar als Anhang ersichtlich, doch werden diese Dateien nicht tatsächlich angehängt und können deshalb weder geöffnet noch gespeichert werden.

Durch die in PDF/A-2 erlaubte Eigenschaft des „optional content“ können zudem diverse Elemente ein- und ausgeblendet werden, was eine eindeutige visuelle Reproduktion ermöglicht.

Die PDF/A-Version 2 kann in den Konformitätsstufen a, b und u vorliegen.

2.4 PDF/A-3

Eigenschaft	Archivische Wertung
basiert auf PDF 1.7	positiv
alle externen Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dateimetadaten (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 digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	positiv

Die PDF/A-Version 3 basiert ebenfalls auf der PDF-Version 1.7. Der Unterschied zwischen PDF/A-2 und 3 besteht darin, dass in Version 3 Dateien beliebigen Formats eingebunden werden können,

Ansprechpartner:

die nicht PDF/A-1 und 2 konform sein müssen. Dies ist nur sinnvoll, wenn neben der langfristig lesbaren PDF/A-Version einer (Text-)Datei 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 aus archivischer Sicht als negativ zu bewerten, da neben der Vielzahl der enthaltenen Dateiformate in solchen Dateien deren Lesbarkeit nicht auf Dauer garantiert werden kann. Eingebundene Dateien, welche nicht im PDF(/A)-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 eine sehr aufwändige Emulation wieder lesbar gemacht werden. Unter diesem Aspekt ist eine Konvertierung in PDF/A-3 nur unter bestimmten Voraussetzungen für eine Langzeitspeicherung sinnvoll und sollte im Zweifel vorher mit dem zuständigen Archiv besprochen werden.

10 Mit der Version 3 ist ebenfalls eine Bildkomprimierung mit JPEG2000 erlaubt. Die archivische Wertung als „(negativ)“ liegt darin begründet, dass alle Kompressionsverfahren durch das Herunterrechnen der Auflösung zu Informationsverlusten führen, welche die Integrität der Datei beeinträchtigen können, auch wenn die Datenverluste für den Menschen meist kaum wahrnehmbar sind.

Durch die in PDF/A- 3 erlaubte Eigenschaft des „optional content“ können zudem diverse Elemente ein- und ausgeblendet werden, was eine eindeutige visuelle Reproduktion ermöglicht.

Die PDF/A-Version 3 kann in den Konformitätsstufen a, b und u vorliegen.

Ansprechpartner:*Referat B5**E-Mail: behoerdenberatung@bundesarchiv.de*

2.5 PDF/A-4

Eigenschaft	Archivische Wertung
basiert auf PDF 2.0	positiv
alle externen Ressourcen (Bilder, Grafiken, Schriftzeichen) müssen in der Datei eingebettet sein	positiv
fester Standard für Farbprofile (ICC-Farbprofile)	positiv
fester Standard für Dateimetadaten (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 erlaubt (nur inaktiv, nicht automatisch)	negativ
Verwendung von Passwörtern ist verboten	positiv
Einbindung von Dateien beliebiger Dateiformate ist erlaubt	(negativ)
Einbettung von digitalen Signaturen in Übereinstimmung mit den PAdES (PDF Advanced Electronic Signatures) möglich	positiv
Verwendung von Verweisen (Hyperlinks) möglich	positiv

11

Der Unterschied zwischen PDF/A-3 und PDF/A-4 besteht darin, dass PDF/A-4 auf PDF 2.0 basiert, der Einsatz von Javascript unter bestimmten Voraussetzungen erlaubt ist und es nicht mehr die Konformitätsstufen „a“, „b“ und „u“, sondern nur noch die beiden Konformitätsstufen „e“ und „f“ gibt.

Wie bei PDF/A-3 ist es auch bei PDF/A-4 möglich, sowohl PDF/A-konforme als auch nicht-konforme Dateien einzubinden. Letzteres stellt eine Herausforderung im Hinblick auf die langfristige Lesbarkeit der eingebundenen Dateien dar, da diese ggf. weder mit einem PDF-Viewer angezeigt werden können, noch gewährleistet werden kann, dass die betreffenden Formate in der Zukunft mit Hilfe der zur Verfügung stehenden Programme lesbar gemacht werden können. Vor allem im Hinblick auf die Langzeitspeicherung und Abgabe an ein Archiv sollte sich die abgebende Stelle in diesem Fall mit dem zuständigen Archiv in Verbindung setzen, um die notwendige

Ansprechpartner:

Dokumentation des eingesetzten Systems und der Dateiformate zu sichern, damit diese ggf. mit Hilfe der Emulation wieder lesbar gemacht werden können.

Mit der Version 4 ist ebenfalls eine Bildkomprimierung mit JPEG2000 erlaubt. Die archivische Wertung als „(negativ)“ liegt darin begründet, dass alle Kompressionsverfahren durch das Herunterrechnen der Auflösung zu Informationsverlusten führen, welche die Integrität der Datei beeinträchtigen können, auch wenn die Datenverluste für den Menschen meist kaum wahrnehmbar sind.

Die Eigenschaft, dass Java-Script in PDF/A-4-Dateien erlaubt ist, steht einer eindeutigen visuellen Reproduzierbarkeit entgegen, welche somit die Authentizität und Integrität der Dateien beeinträchtigt.

12 Die PDF/A-Version 4 kann ausschließlich in den Konformitätsstufen e und f vorliegen.

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

2.6 Gegenüberstellung der PDF/A-Versionen

Eigenschaft	PDF/A-1	PDF/A-2	PDF/A-3	PDF/A-4
Einbettung aller externer Ressourcen	✓	✓	✓	✓
fester Standard für Farbprofile	✓	✓	✓	✓
Dateimetadaten im XMP-Format	✓	✓	✓	✓
PDF-Ebenen erlaubt	✗	✓	✓	✓
Transparente Elemente erlaubt	✗	✓	✓	✓
Interaktive Elemente verboten	✓	✓	✓	✓
Bildkomprimierung erlaubt	✓	✓	✓	✓
Java-Script verboten	✓	✓	✓	✗
Verwendung von Passwörtern verboten	✓	✓	✓	✓
Einbindung weiterer Dateien erlaubt	✗	✓	✓	✓
Einbettung digitaler Signaturen erlaubt	✓	✓	✓	✓
Verwendung von Hyperlinks erlaubt	✓	✓	✓	✓

Ansprechpartner:

3. Auswertung der PDF/A-Konformitätsstufen

3.1 Die PDF/A-Konformitätsstufen

Bei den verschiedenen PDF/A-Versionen können folgende Konformitätsstufen verwendet werden:

- *a* = „accessible“ (PDF/A-1, PDF/A-2, PDF/A-3)
- *b* = „basic“ (PDF/A-1, PDF/A-2, PDF/A-3)
- *u* = „unicode“ (PDF/A-2, PDF/A-3)
- *e* = „engineering“ (PDF/A-4)
- *f* = „flat“ (PDF/A-4)

Die Eigenschaften der jeweiligen Konformitätsstufen werden in nachfolgender Tabelle gegenübergestellt:

Eigenschaft	a	b	u
eindeutige visuelle Reproduzierbarkeit gewährleistet	ja	ja	ja
Textinhalte müssen extrahierbar sein	ja	optional	optional
Struktur muss die natürliche Leseabfolge eindeutig abbilden	ja	optional	optional
hohe Anforderungen an verwendete Schriftzeichen	ja	nein	nein
Barrierefreiheit wird bestmöglich umgesetzt	ja	optional	optional
Textinhalte in Unicode	ja	optional	ja

Die Konformitätsstufen „e“ und „f“ basieren auf der Konformitätsstufe „u“ mit folgenden Abweichungen:

- *f* – Einbettung PDF/A-konformer und nicht konformer (2D-)Dateien erlaubt
- *e* – Einbettung von 3D-Inhalten erlaubt

Ansprechpartner:

3.2 Anwendung der PDF/A-Konformitätsstufen

Betrachtet man die jeweiligen Eigenschaften ist die Entscheidung für eine bestimmte Konformitätsstufe von der digitalen Ursprungsdatei und dem beabsichtigten Verwendungszweck abhängig.

a = „accessible“

Diese Konformitätsstufe kommt bei den PDF/A-Versionen 1, 2 sowie 3 vor und sollte bevorzugt bei der Konvertierung von genuin digitalen Ursprungsdateien angewendet werden. Eine Anwendung dieser Konformitätsstufe bei digitalisierten Dokumenten ist nur möglich, wenn eine Texterkennung mittels OCR stattgefunden hat und in die PDF/A-Datei eingebettet wurde. Neben der Darstellung des Textes in Unicode, bleibt die Struktur der (Text-)Datei und die natürliche Lesereihenfolge erhalten, sodass die (Text-)Datei mit Screenreader-Programmen vorgelesen werden kann und somit eine bestmögliche Barrierefreiheit bietet.

15

b = „basic“

Diese Konformitätsstufe kommt bei den PDF/A-Versionen 1, 2 sowie 3 vor und eignet sich bevorzugt bei der Konvertierung von digitalisierten Unterlagen. Auch kann hier eine Texterkennung mittels OCR mit entsprechender Einbettung in die PDF/A-Datei erfolgen, ist bei dieser Konformitätsstufe jedoch nicht verpflichtend umzusetzen. Da die für eine barrierefreie (Text-)Datei grundlegenden Anforderungen nur optional sind und nicht zwingend angewendet werden müssen, eignet sich diese Konformitätsstufe nur unter gewissen Voraussetzungen für den Einsatz barrierefreier (Text-)Dateien.

Ansprechpartner:

u = „unicode“

Diese Konformitätsstufe kommt bei den PDF/A-Versionen 2 sowie 3 vor und eignet sich vor allem bei digitalisierten Dokumenten, deren Inhalte durchsuchbar sein müssen, da bei dieser Konformitätsstufe alle Textinhalte in Unicode vorliegen. Voraussetzung hierfür ist eine Texterkennung mittels OCR, welche in das PDF/A eingebettet werden kann. Da die für eine barrierefreie (Text-)Datei grundlegenden Anforderungen nur optional sind und nicht zwingend angewendet werden müssen, eignet sich diese Konformitätsstufe nur unter gewissen Voraussetzungen für den Einsatz barrierefreier (Text-)Dateien.

e = „engineering“

16

Diese Konformitätsstufe kommt ausschließlich bei der PDF/A Version 4 vor und erfüllt zunächst alle Eigenschaften der Konformitätsstufe „u“. Zudem erlaubt die Konformitätsstufe „e“ die Einbettung von 3D-Inhalten in eine PDF/A-4-Datei. Die Anwendung der Konformitätsstufe „e“ eignet sich daher v.a. bei in PDF/A-4 konvertierten Dateien, bei denen 3D-Inhalte eingebettet werden müssen. Da die für eine barrierefreie (Text-)Datei grundlegenden Anforderungen bei dieser Konformitätsstufe nur optional sind und nicht zwingend angewendet werden müssen, eignet sich die Konformitätsstufe „e“ nur unter gewissen Voraussetzungen für den Einsatz barrierefreier (Text-)Dateien.

f = „flat“

Diese Konformitätsstufe kommt ausschließlich bei der PDF/A Version 4 vor und erfüllt zunächst alle Eigenschaften der Konformitätsstufe „u“. Zudem erlaubt die Konformitätsstufe „f“ die Einbettung von PDF/A-konformen und nicht-konformen Inhalten. Die Anwendung der Konformitätsstufe „f“ eignet sich daher v.a. bei PDF/A-4-Dateien, bei denen 2D-Inhalte eingebettet werden müssen. Da

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

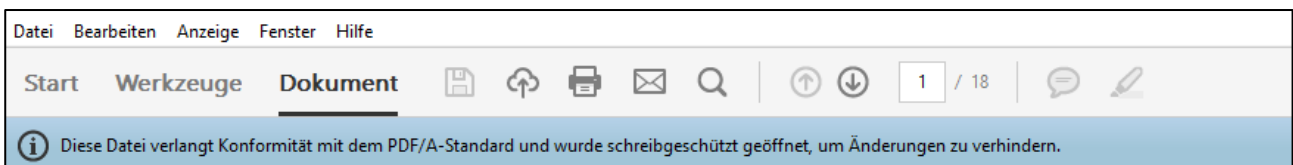
die für eine barrierefreie (Text-)Dateien grundlegenden Anforderungen bei dieser Konformitätsstufe nur optional sind und nicht zwingend angewendet werden müssen, eignet sich die Konformitätsstufe „f“ nur unter gewissen Voraussetzungen für den Einsatz barrierefreier (Text-)Dateien.

4. Wie erkennt man eine PDF/A-Datei?

Eine PDF/A-Datei lässt sich weder in den Dateieigenschaften, noch an der Dateiendung erkennen, da auch hier die Endung wie bei einer PDF-Datei „.pdf“ lautet. Dennoch gibt es verschiedene Wege, eine PDF/A-Datei als solche zu identifizieren.

a) Anzeige im PDF-Reader

Je nach PDF-Reader kann eine entsprechende Anzeige auf das Vorliegen einer PDF/A-Datei hinweisen, wie im nachfolgenden Beispiel zu sehen ist:



b) Auslesen des Dateiformates in einem Texteditor

Es besteht weiterhin die Möglichkeit, das Dateiformat in einem einfachen Texteditor auszulesen. Hierzu öffnet man die Datei in einem Texteditor und gibt in das Suchfeld den Begriff „pdfaid“ ein. Die angezeigten Informationen geben Aufschluss über die PDF/A-Konformität („A“) sowie über die PDF/A-Version („2“).

```
<pdfaid:part>2</pdfaid:part><pdfaid:conformance>A</pdfaid:conformance>
```

Ansprechpartner:

c) Formaterkennung über ein Tool

Zahlreiche Tools, die aus Gründen des Wettbewerbsvorteils an dieser Stelle nicht genannt werden können, ermöglichen eine gezielte Formaterkennung mit Ausgabe des jeweiligen Ergebnisses.

5. Konvertierung

Allgemeine Hinweise zur Konvertierung

Die Konvertierung in PDF/A kann je nach verwendeter PDF/A-Version und Konformitätsstufe zu Konvertierungsfehlern und folglich zu Informationsverlusten führen. Das Bundesarchiv empfiehlt, eine Konvertierung in PDF/A direkt nach der Fertigstellung eines Dokuments durch die zuständige Bearbeiterin bzw. den zuständigen Bearbeiter durchführen zu lassen, damit dieser in einer anschließenden Qualitätssicherung der konvertierten Datei das Ergebnis prüfen und Konvertierungsfehler unmittelbar beheben kann. Da dieses Vorgehen in der Praxis jedoch nicht immer umgesetzt werden kann, greifen manche Behörden auf das Mittel der Massenkonzertierung zurück. Hierbei wird z.B. der Inhalt eines ganzen Vorgangs nach dessen Abschluss (z.d.A.-Verfügung) in PDF/A konvertiert. Eine solche Massenkonzertierung spart Zeit, geht jedoch meist mit zahlreichen Konvertierungsfehlern und Informationsverlusten einher. Eine Qualitätssicherung im Anschluss des Konvertierungsvorgangs ist daher immer notwendig. Zudem sollte das Ausgangsformat zusammen mit dem konvertierten Format aufbewahrt werden, um bei Konvertierungsfehlern die Datei zu einem späteren Zeitpunkt erneut lesen und ggf. konvertieren zu können. Zudem wird an dieser Stelle darauf hingewiesen, dass sich nicht alle Dateiformate ohne weiteres in PDF/A konvertieren lassen. Eine Konvertierung von Excel-Dateien in PDF/A kann z.B. zur Folge haben, dass bei einer automatischen Konvertierung nur das erste Tabellenblatt konvertiert wird, wodurch die Informationen auf weiteren Tabellenblättern verloren gehen oder Informationen, die über Seitenbegrenzungen hinaus dargestellt werden, bei der Konvertierung abgeschnitten werden und ggf. verloren gehen. Zudem ist eine Konvertierung von TIFF-Dateien in PDF/A nicht

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

notwendig, da dieses Format für eine Langzeitspeicherung ebenfalls geeignet ist. Aus diesem Grund empfiehlt das Bundesarchiv eine Liste derjenigen Dateiformate anzulegen, welche in dem jeweiligen E-Akte-System abgelegt werden dürfen und nach entsprechender Abwägung für jedes Dateiformat ein Zielformat für die Konvertierung festzulegen. Das Bundesarchiv steht hierbei beratend zur Verfügung.

Konvertierung nach PDF/A

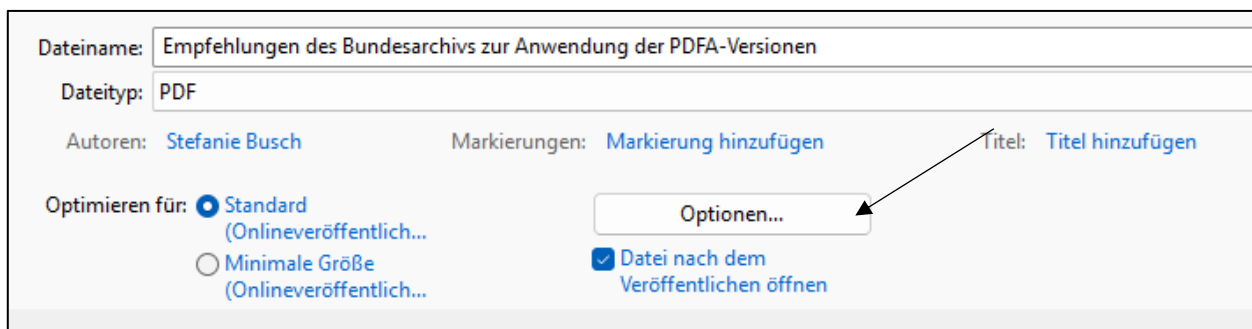
Bei der Konvertierung von (Text-)Dateien in PDF/A sind die oben genannten Eigenschaften der verschiedenen PDF/A-Versionen zu beachten, um das geeignete Zielformat zu bestimmen.

PDF/A-Dateien lassen sich auf verschiedenen Wegen erzeugen. Die gängigsten Varianten sind dabei:

19 a) Abspeichern direkt als PDF/A

Um eine in „MS Word“¹ erstellte (Text-)Datei im PDF/A-Format zu speichern, geht man wie folgt vor:

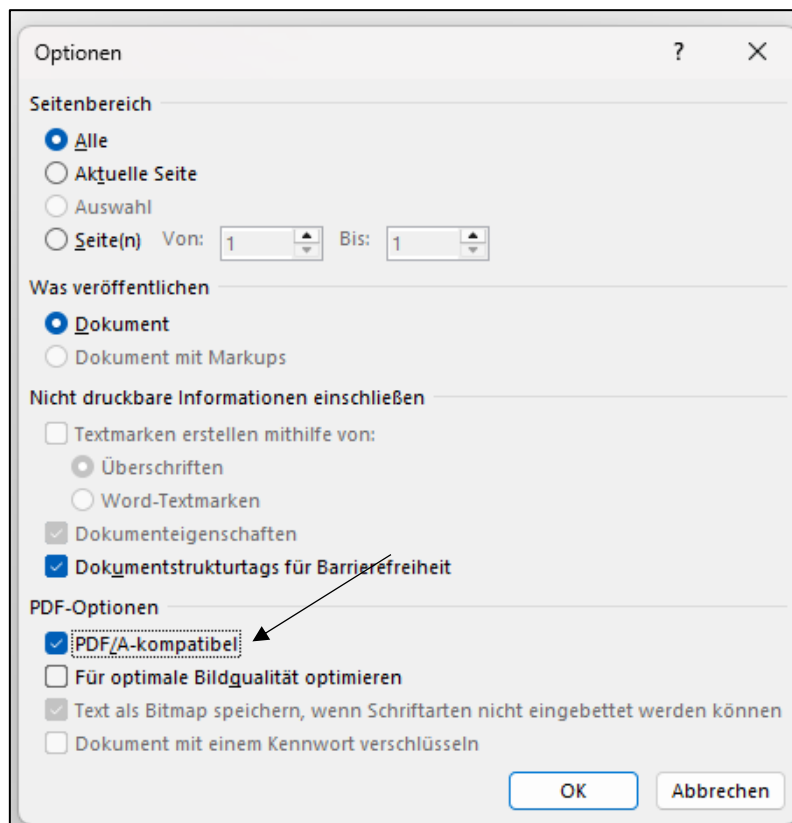
Datei → Speichern unter → Dateityp: PDF → Optionen...



¹ Die Nennung des Produktes erfolgt an dieser Stelle aus dem Grund, da „MS Word“ bereits als gängiges Textverarbeitungsprogramm in der Bundesverwaltung eingesetzt wird. Ein Wettbewerbsvorteil soll hierdurch nicht erzielt werden.

Ansprechpartner:

In den „Optionen“ kann man nun unter „PDF-Optionen“ das Feld „PDF/A-kompatibel“ anklicken. Somit wird bei der Speicherung eine PDF/A-Datei erzeugt.



20

Eine Auswahl, welche PDF/A-Version erzeugt werden soll, lässt sich an dieser Stelle nicht individuell festlegen und ist abhängig von der verwendeten „MS Word“-Version. In der Regel verwenden ältere „MS Word“-Versionen die PDF/A-Version „1“, die neueren „MS Word“-Versionen (Office 2013/2016/2019/365) verwenden die PDF/A-Version „3“.

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de

b) Konvertierung einer Datei über ein entsprechendes Tool

Verschiedene Tools, die aus Gründen des Wettbewerbsvorteils an dieser Stelle nicht genannt werden können, bieten die Möglichkeit, (Text-)Dateien in PDF/A zu konvertieren. In der Regel erhält man hierbei die Möglichkeit, die gewünschte PDF/A-Version sowie die gewünschte Konformitätsstufe auszuwählen.

6. Validierung

Die Validierung von PDF/A-Dateien durch ein entsprechendes Tool ist ein wichtiger Schritt in der Qualitätssicherung, 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-Dateien wird daher empfohlen.

21

Eine Validierung der PDF/A-Dateien sollte stattfinden:

- nach der Erstellung
- nach dem Empfang
- vor dem Versand bzw. der Verteilung
- vor der Langzeitspeicherung bzw. 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 des Validierungstools reparieren. Andernfalls muss die Datei mit modifizierten Einstellungen erneut erzeugt werden. Eine Übersicht der Eignung verschiedener

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de

Validierungstools bietet z.B. die Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST).²

7. Barrierefreiheit durch PDF/UA

Um barrierefreie (Text-)Dateien zu erzeugen, müssen bestimmte visuelle und technische Anforderungen an die Barrierefreiheit wie z.B. das Einhalten der Lesereihenfolge, die Verwendung von Tags, Beschriftungen von Grafiken usw. eingehalten werden. Um die erzeugten barrierefreien (Text-)Dateien layoutstabil zu halten, wird das Dateiformat PDF/UA verwendet, welches zudem eine Validierung auf Barrierefreiheit ermöglicht. „UA“ steht hierbei für „Universal Accessibility“ und hat nichts mit dem PDF/A-Format zu tun, bei dem das „A“ für „Archivable“ steht und somit in erster Linie Anforderungen an die Langzeitspeicherfähigkeit von (Text-)Dateien stellt. Das PDF/UA-Format ist somit nicht zwangsläufig langzeitspeicherfähig, da es den Einsatz von Eigenschaften zulässt, welche der Langzeitspeicherung entgegenstehen. PDF/UA ist daher nur in den Fällen für eine Langzeitspeicherung geeignet, wenn zusätzlich die in der vorliegenden Handreichung erläuterten Anforderungen an PDF/A umgesetzt werden. Wiederum lässt sich pauschal keine Aussage darüber treffen, ob eine PDF/A-Datei barrierefrei ist, da dies von der zu konvertierenden Ausgangsdatei abhängt. Sollte die Ausgangsdatei barrierefrei sein und in ein PDF/A-Format (mit der Konformitätsstufe „a“) konvertiert werden, so ist die Barrierefreiheit zumindest gegeben, muss jedoch nach der Konvertierung nochmal entsprechend geprüft und bestätigt werden.

Für die Langzeitspeicherung barrierefreier (Text-)Dateien empfiehlt das Bundesarchiv, diese zunächst gemäß den Anforderungen an die Barrierefreiheit zu erstellen und anschließend in PDF/UA zu konvertieren. Zum Abschluss ist die Datei hinsichtlich der Barrierefreiheit zu validieren.

² Koordinierungsstelle für die dauerhafte Archivierung elektronischer Unterlagen: PDF/A-Validatoren: https://kost-ceco.ch/cms/pdfa_validator_software_de.html; Zugriffsdatum: 12/23.

Ansprechpartner:

Referat B5

E-Mail: behoerdenberatung@bundesarchiv.de

Bei einem positiven Validierungsergebnis kann die Datei anschließend in PDF/A-2a konvertiert und anschließend entsprechend dem PDF/A-Standard validiert werden.

8. Verwendete Quellen

- axes4: Achtung Verwechslungsgefahr: PDF/A oder PDF/UA?; <https://www.axes4.com/de/blog/post/2023/achtung-verwechslungsgefahr-pdf-a-oder-pdf-ua>; Zugriffsdatum: 12/23.
- Barrierekompass: Barrierefreie PDF erstellen – wie geht das?; <https://barrierekompass.de/pdf/barrierefreie-pdf-erstellen-wie-geht-das.html>; Zugriffsdatum: 12/23.
- Bundesamt für Sicherheit in der Informationstechnik (BSI): Technische Richtlinie 03125. Beweiswerterhaltung kryptographisch signierter Dokumente. Anlage TR-ESOR-F: Formate. 2015.
- Bundesfachstelle Barrierefreiheit: Barrierefreie PDF; <https://www.bundesfachstelle-barrierefreiheit.de/DE/Fachwissen/Informationstechnik/Barrierefreie-PDF/barrierefreie-pdf.html?nn=fb65f4b1-fb95-4e43-9626-f85281d10e6d#doc0645f0cd-ef87-4699-9538-2475ea94a50dbodyText5>; Zugriffsdatum: 12/23.
- Compart: PDF-Formate; <https://www.compart.com/de/pdf-formate-a-1-2-3-e-ua-vt-x>; Zugriffsdatum: 12/23.
- 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.

Ansprechpartner:

- 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.
- PDFlib: Die PDF/A-Archivierungsstandards; <https://www.pdflib.com/de/pdf-know-how/pdfa/die-pdf-a-standards/>; Zugriffsdatum: 12/23.
- PDF Tools: PDF Standards. PDF/A-4 Übersicht; <https://www.pdf-tools.com/de/pdf-knowledge/pdf-a-4-overview/>; Zugriffsdatum: 12/23.
- SEALSYSTEMS – Enterprise Conversion Solutions: PDF/A-4 – ein Überblick; <https://www.sealsystems.de/blog/pdf-a-4-ein-ueberblick/>; Zugriffsdatum: 12/23.
- Seggern, Dietrich von (PDF/A association): PDF und PDF/A Validierung – im Brennpunkt Schriften. 2017.
- Technische Universität Dresden: Barrierefreie Dokumente. Anleitung zur Erstellung barrierefreier PDF-Dokumente mit Word. Dresden. 2020.
- 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.
- WebPDF: PDF/A – Das Format der Zukunft. Teil 4: PDF/UA; <https://www.webpdf.de/blog/pdfa-das-format-der-zukunft-teil-4-pdfua/>; Zugriffsdatum: 12/23.

Ansprechpartner:

Referat B5
E-Mail: behoerdenberatung@bundesarchiv.de