



SLUB

Wir führen Wissen.

Erstellung PDF/A-konformer Dokumente: Erläuterung häufiger Fehlermeldungen

Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden
Dresden, 2015



Dieses Material steht unter der Creative-Commons-Lizenz Namensnennung 3.0 Deutschland. Um eine Kopie dieser Lizenz zu sehen, besuchen Sie:
<http://creativecommons.org/licenses/by/3.0/de/>

1	Einleitung	3
2	Fehlermeldungen	5
2.1	CIDSet in Schrift-Untergruppe fehlt	5
2.2	CIDSet in Schrift-Untergruppe ist unvollständig	6
2.3	CharSet für Typ 1 Schrift fehlt oder ist unvollständig	6
2.4	Verwendet OpenType-Schrift	7
2.5	Schrift ist nicht gültig	7
2.6	Glyphen fehlen in eingebetteter Schrift	8
2.7	Weiteninformationen für ausgegebene Glyphen inkonsistent	9
2.8	Verwendet JPEG2000-Kompression	9
2.9	Transparenz verwendet	10
2.10	XMP-Eigenschaft im Erweiterungsschema weder vordefiniert noch definiert	12
2.11	XMP-Eigenschaft ist vordefiniert, aber nicht entsprechend der Definition verwendet	12
2.12	„EmbeddedFiles“-Eintrag im Namens-Verzeichnis (Dictionary)	12
2.13	Enthält EF-Eintrag für eingebettete Dateien	14
3	Zusammenfassung	15
4	Literatur	15

Abbildungsverzeichnis:

Abbildung 1:	„CIDSet in Schrift-Untergruppe fehlt“	5
Abbildung 2:	„CIDSet in Schrift-Untergruppe fehlt“	5
Abbildung 3:	„CharSet für Typ 1“	6
Abbildung 4:	„OpenType-Schrift“	7
Abbildung 5:	„Schrift ist nicht gültig“	8
Abbildung 6:	„Glyphen fehlen in eingebetteter Schrift“	8
Abbildung 7:	„Weiteninformationen für ausgegebene Glyphen inkonsistent“	9
Abbildung 8:	„JPEG2000-Kompression“	10
Abbildung 9:	„Transparenz verwendet“	11
Abbildung 10:	„Transparenz verwendet“	11
Abbildung 11:	„Anlage löschen“	13
Abbildung 11:	„Anlage löschen bestätigen“	13

1 Einleitung

PDF/A¹ ist ein spezieller Standard des PDF-Formats, das bestimmte Dokumenteigenschaften verlangt, um eine Wiedergabe des Dokuments auch in einigen Jahrzehnten garantieren zu können:

„Ziel des PDF/A-Standards ist, dass PDF-Dokumente erstellt werden können, deren visuelles Erscheinungsbild über die Zeit erhalten bleibt“ (Drümmer, Oettler and Seggern, 2007, p.9).

Um die Garantie einhalten zu können, müssen PDF/A-konforme Dateien bestimmte Kriterien erfüllen:

„Der PDF/A Standard erlaubt keine Objekte, die in der PDF 1.4 Spezifikation, die von Adobe mit ACROBAT 5 veröffentlicht wurde, noch nicht vorgesehen waren“ (Drümmer, Oettler and Seggern, 2007, p.78).

„Verbotene“ Inhalte sind zum Beispiel Transparenz von Bildern (bei PDF/A-1), Metadaten, die nicht XMP-konform sind, Bilder in JPEG2000 Kompression und viele mehr. Zudem müssen alle im Dokument verwendeten Schriften in diesem eingebettet sein. Durch diese strikten Anforderungen ist eine PDF/A-konforme Datei nicht immer einfach zu erstellen. Mehrere Faktoren erschweren die Einhaltung des PDF/A-Standards, als auch die Erstellung von Anleitungen, bzw. Beschreibungen:

- Verwendung unterschiedlicher Programme zur Erstellung der Quelldokumente
 - ACROBAT PDFMAKER für WORD
 - ADOBE INDESIGN
 - LATEX WITH HYPERREF PACKAGE
 - MICROSOFT WORD 2007
 - MICROSOFT WORD 2010
 - PDFTEX
 - PSCRIPT5.DLL
 - QUARKXPRESS
 - TEX
- Verwendung unterschiedlicher Programme zur Erstellung des PDF/A-Dokuments
 - ACROBAT DISTILLER (6.0, 7.0, 7.0.5, ...)
 - ACROBAT PDFWRITER (5.0, ...)
 - ADOBE PDF LIBRARY (9.0, 9.9, ...)
 - DVIPS + GHOSTSCRIPT
 - PDFTEX (1.40.12, 1.40.14, ...)
- Nicht jede Software erstellt PDF/A-Dokumente nach den gleichen Regeln
- Die unterschiedlichen Prüfprogramme liefern in einigen Fällen unterschiedliche Ergebnisse. Dies erschwert die Abgabe korrekter PDF/A-Dokumente.
 - JHOVE²
 - 3-HEIGHTS™ PDF VALIDATOR³
 - ADOBE PROFESSIONAL PRO / PREFLIGHT⁴

1 <http://de.wikipedia.org/wiki/PDF/A>

2 <http://sourceforge.net/projects/jhove/>

3 <http://www.pdf-tools.com/pdf/pdfa-online-pruefen.aspx>

4 http://help.adobe.com/de_DE/acrobat/using/WS58a04a822e3e50102bd615109794195ff-7b81.w.html

Die Erfahrungen zeigen, dass das PREFLIGHT Modul in ADOBE PROFESSIONAL PRO sehr zuverlässig arbeitet und auch aussagekräftige Informationen zu den enthaltenen Fehlern liefert. Eine Zusammenstellung der möglichen Fehlerquellen und zugehörige Lösungsvorschläge findet sich in PDF/kompakt (Drümmer, Oettler and Seggern, 2007, p.66ff.).

Die beste Lösung besteht darin, dass schon im Originaldokument (MS WORD, ADOBE IN DESIGN...) darauf geachtet wird, dass alle PDF/A-Eigenschaften erfüllt werden. Dadurch umgeht man aufwändige Konversionen. Vor allem kostenfreie PDF-Erstellungssoftware haben Probleme damit, alle Inhalte eines Dokuments PDF/A-konform zu konvertieren.

Auch wenn der PDF/A-Standard noch keine breite Anwendung erfährt, fordern viele Bibliotheken (und andere Einrichtungen, die die Verantwortung für eine dauerhafte Bereitstellung von Dokumenten tragen) dieses Format, um eine dauerhafte Bereitstellung zu ermöglichen. Der Trend zum PDF/A-Standard wird sich daher mit Sicherheit verstärken und **eine korrekte Erstellung der Quelldateien wird somit immer wichtiger**.

Auf der SLUB-Website finden sich weitere Informationen zur Langzeitarchivierung digitaler Dokumente:

- <http://www.slub-dresden.de/service/pflichtmedien-abliefern/technische-standards-fuer-die-ablieferung-von-netzpublikationen/>
- <http://www.slub-dresden.de/ueber-uns/digitale-langzeitarchivierung/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/>

Um eine PDF/A-konforme Erstellung der Quelldateien zu erleichtern, werden im Folgenden die häufigsten auftretenden Fehlermeldungen von PREFLIGHT aufgelistet und beschrieben, wie diese vermieden werden können.

Das schon zitierte Dokument PDF/A kompakt (Drümmer, Oettler and Seggern, 2007) enthält Erläuterungen zu einigen Inhalten in PDF/A-konformen Dateien. Das Studium dieser Erläuterungen kann zum besseren Verständnis der Fehlerbeschreibungen beitragen, zum Beispiel:

- Transparenz (Drümmer, Oettler and Seggern, 2007, p.42)
- Schriften (Drümmer, Oettler and Seggern, 2007, p.46)
- Metadaten (Drümmer, Oettler and Seggern, 2007, p.48)
- Interaktive PDF-Dateien (Drümmer, Oettler and Seggern, 2007, p.52)

2 Fehlermeldungen

2.1 CIDSet in Schrift-Untergruppe fehlt

Ursache

Diese Meldung weist darauf hin, dass eine Schrift nicht korrekt eingebettet ist. Dies könnte in einigen Jahren dazu führen, dass das Dokument nicht mehr dargestellt werden kann, weil die Schrifttypen nicht mehr auf den Computern vorhanden sind. Ähnlich verhält es sich mit proprietären Schriften, die nur durch Lizenzierung genutzt werden dürfen. Dies widerspricht natürlich dem Vorhaben der Langzeitarchivierung, die es ermöglichen soll, Dokumente auch nach Jahrzehnten authentisch wiedergeben zu können.

Es muss unbedingt darauf geachtet werden, dass alle Schriften eingebettet sind. Dies bezieht sich auch auf Aufzählungszeichen und ähnliche Inhalte! Preflight weist in dem Prüfungsprotokoll auch auf alle Stellen hin, an der ein Fehler auftritt. Hier ein Beispiel:

- auf einem A 4 Blatt dargestellt werden,
- maximal 3 Farben enthalten,
- kein Gemälde sein,
- nicht zwingend die „700“ enthalten,
- einen Bezug zu Großsteinberg herstellen,
- auf der Rückseite den Namen, die Anschrift und Telefonnummer des Einreichers enthalten.

Abbildung 1: „CIDSet in Schrift-Untergruppe fehlt“

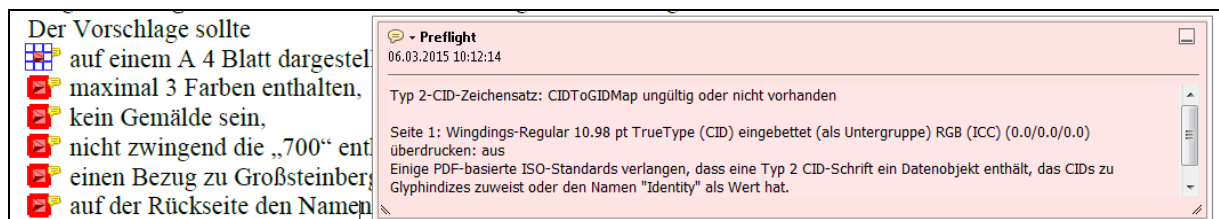


Abbildung 2: „CIDSet in Schrift-Untergruppe fehlt“

Bei folgenden Schriften traten häufiger Fehler mit dem CID-Zeichensatz auf:

- SymbolMT
- Wingdings-Regular

Ob Schriften eingebettet werden können, hängt auch von der PDF-Version ab, die erstellt wird. Folgenden Webseiten bieten weiterführende Informationen zu den unterschiedlichen Schrifttypen:

- <http://de.wikipedia.org/wiki/TrueType>
- <http://de.wikipedia.org/wiki/OpenType>
- <http://de.wikipedia.org/wiki/PostScript-Fontformate>
- <http://www.prepressure.com/pdf/basics/fonts>

Lösung

„Um dieses Problem zu beheben, muss die PDF-Datei unter Verwendung einer anderen Schrift, oder der gleichen, vollständigen Schrift neu erzeugt werden. Man kann auch die unvollständige Schrift, dann aber mit Char-Set verwenden“ (Drümmer, Oettler and Seggern, 2007, p.68).

2.2 CIDSet in Schrift-Untergruppe ist unvollständig

Ursache

Dieser Fehler tritt häufig mit 2.1 CIDSet in Schrift-Untergruppe fehlt und aus ähnlichen Gründen auf, weshalb auf eine ausführliche Beschreibung verzichtet wird.

Lösung

„Um dieses Problem zu beheben, muss die PDF-Datei unter Verwendung einer anderen Schrift, oder der gleichen, vollständigen Schrift neu erzeugt werden. Man kann auch die unvollständige Schrift, dann aber mit Char-Set verwenden“ (Drümmer, Oettler and Seggern, 2007, p.68).

2.3 CharSet für Typ 1 Schrift fehlt oder ist unvollständig

Ursache

In diesem Fall ist die Benennung des Fehlers sehr aufschlussreich, denn:

„Wenn eine Schrift im Typ1-Format unvollständig eingebettet wird, muss diese eine Liste der eingebetteten Buchstaben enthalten, um nach PDF/A konvertiert werden zu können.“ (Drümmer, Oettler and Seggern, 2007, p.68)

Dieser Fehler tritt häufig mit 2.4 Verwendet OpenType-Schrift und 2.5 Schrift ist nicht gültig auf. Das folgende Beispiel zeigt, dass unter Umständen auch nur Bestandteile eines Textes diesen Fehler verursachen. Dies könnte auf Formatierungsfehler im Quelldokument zurückzuführen sein.

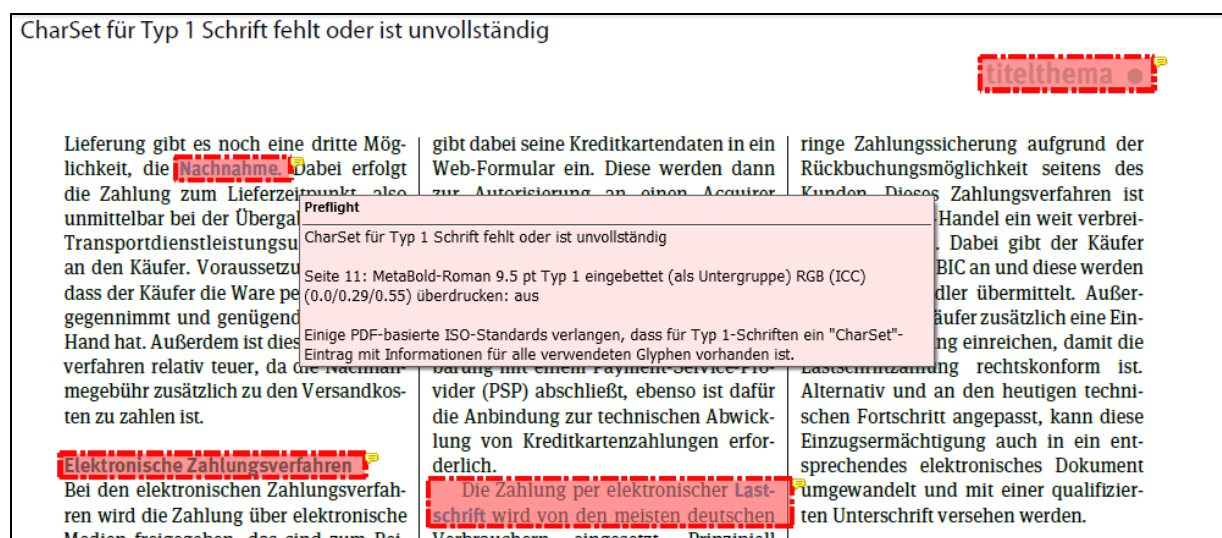


Abbildung 3: „CharSet für Typ 1“

Lösung

„Um dieses Problem zu beheben, muss die PDF-Datei unter Verwendung einer anderen Schrift, oder der gleichen, vollständigen Schrift neu erzeugt werden. Man kann auch die unvollständige Schrift, dann aber mit CharSet verwenden“ (Drummer, Oettler and Seggern, 2007, p.68).

Es ist ersichtlich, dass dies am besten in dem Quelldokument durchgeführt wird.

2.4 Verwendet OpenType-Schrift

Ursache

Dieser Fehler tritt häufig mit 2.3 CharSet für Typ 1 Schrift fehlt oder ist unvollständig und 2.5 Schrift ist nicht gültig auf und tritt aus ähnlichen Gründen auf, weshalb auf eine ausführliche Beschreibung verzichtet wird.

Es soll hier aber als Beispiel dieselbe Stelle aus dem Report gezeigt werden, um zu verdeutlichen, dass die Fehler an den gleichen Textstellen auftreten. Somit würden, wenn möglich, mehrere Fehler „mit einer Klappe“ korrigiert.

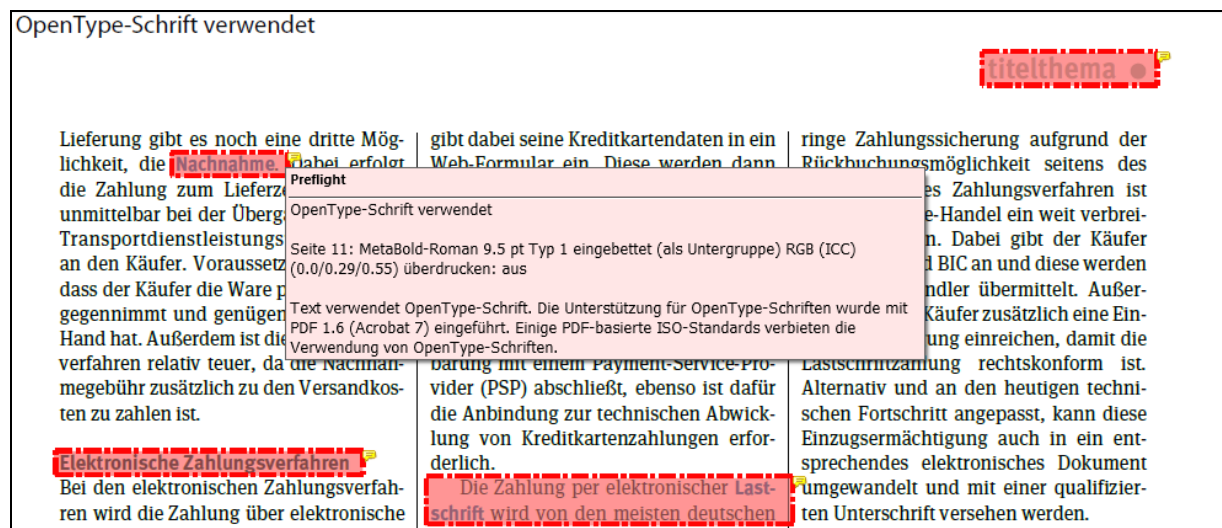


Abbildung 4: „OpenType-Schrift“

2.5 Schrift ist nicht gültig

Ursache

Dieser Fehler tritt häufig mit 2.3 CharSet für Typ 1 Schrift fehlt oder ist unvollständig und 2.4 Verwendet OpenType-Schrift auf und tritt aus ähnlichen Gründen auf, weshalb auf eine ausführliche Beschreibung verzichtet wird.

Es soll hier aber als Beispiel dieselbe Stelle aus dem Report gezeigt werden, um zu verdeutlichen, dass die Fehler an den gleichen Textstellen auftreten. Somit würden, wenn möglich, mehrere Fehler „mit einer Klappe“ korrigiert.



Abbildung 5: „Schrift ist nicht gültig“

2.6 Glyphen fehlen in eingebetteter Schrift

Ursache

Dies ist ein weiteres Problem mit Schriftzeichen. Hier besteht der Fehler darin, dass eine eingebettete Schrift nicht alle Zeichen enthält und somit diese Zeichen nicht dargestellt werden können.

Das Beispiel zeigt, dass dieses Problem auch nur in wenigen Textstellen auftreten kann. Trotzdem ist durch diese Stellen das ganze Dokument nicht PDF/A-konform.

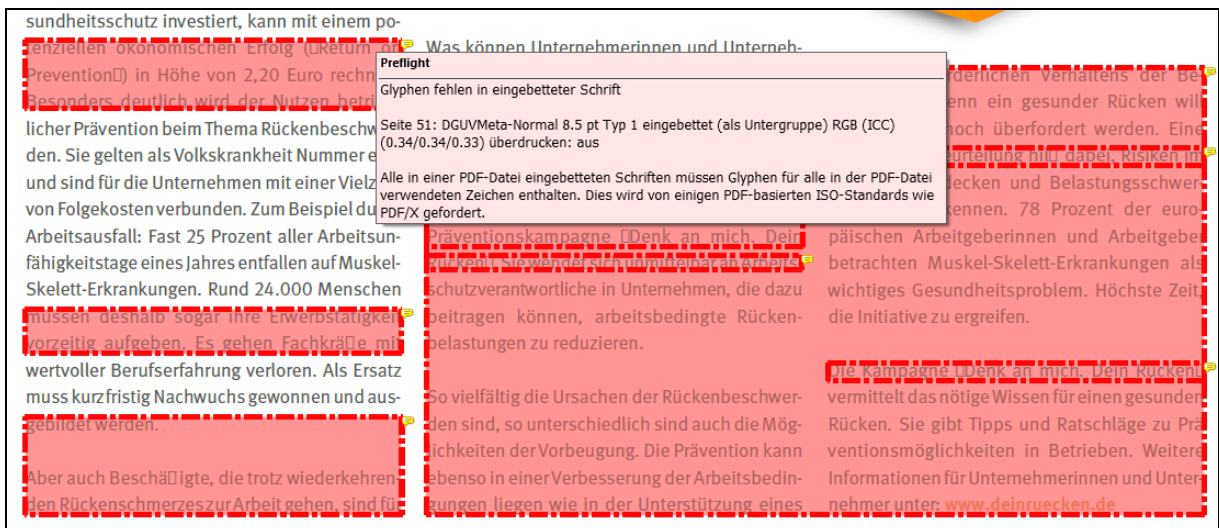


Abbildung 6: „Glyphen fehlen in eingebetteter Schrift“

Lösung

„Um dieses Problem zu beheben, muss die PDF-Datei neu erzeugt werden“ (Drümmer, Oettler and Seggern, 2007, p.72).

2.7 Weiteninformationen für ausgegebene Glyphen inkonsistent

Dieser Fehler tritt häufig mit 2.6 Glyphen fehlen in eingebetteter Schrift auf und tritt aus ähnlichen Gründen auf, weshalb auf eine ausführliche Beschreibung verzichtet wird.

Das Beispiel zeigt, dass dieses Problem auch nur in wenigen Textstellen auftreten kann. Trotzdem ist durch diese Stellen das ganze Dokument nicht PDF/A-konform.

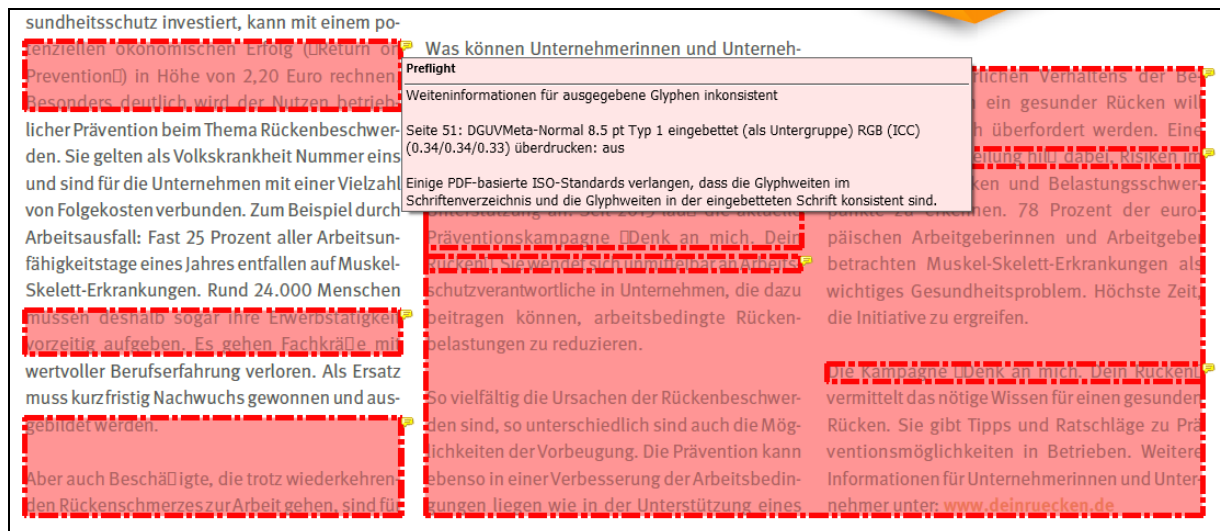


Abbildung 7: „Weiteninformationen für ausgegebene Glyphen inkonsistent“

2.8 Verwendet JPEG2000-Kompression

Ursache

Bilder in Dokumenten, die durch JPEG2000-Kompression komprimiert wurden, sind in PDF/A-1-Dokumenten nicht erlaubt. Das folgende Beispiel zeigt, dass unter Umständen auch nur Bestandteile eines Bildes diesen Fehler verursachen.

Dies bezieht sich nur auf PDF/A-1! PDF/A-2 erlaubt transparente Elemente (<http://de.wikipedia.org/wiki/PDF/A>). Es ist jedoch möglich, dass dann andere Inhalte dem PDF/A-2-Standard nicht entsprechen und es einfacher wäre, den PDF/A-1-Standard einzuhalten.

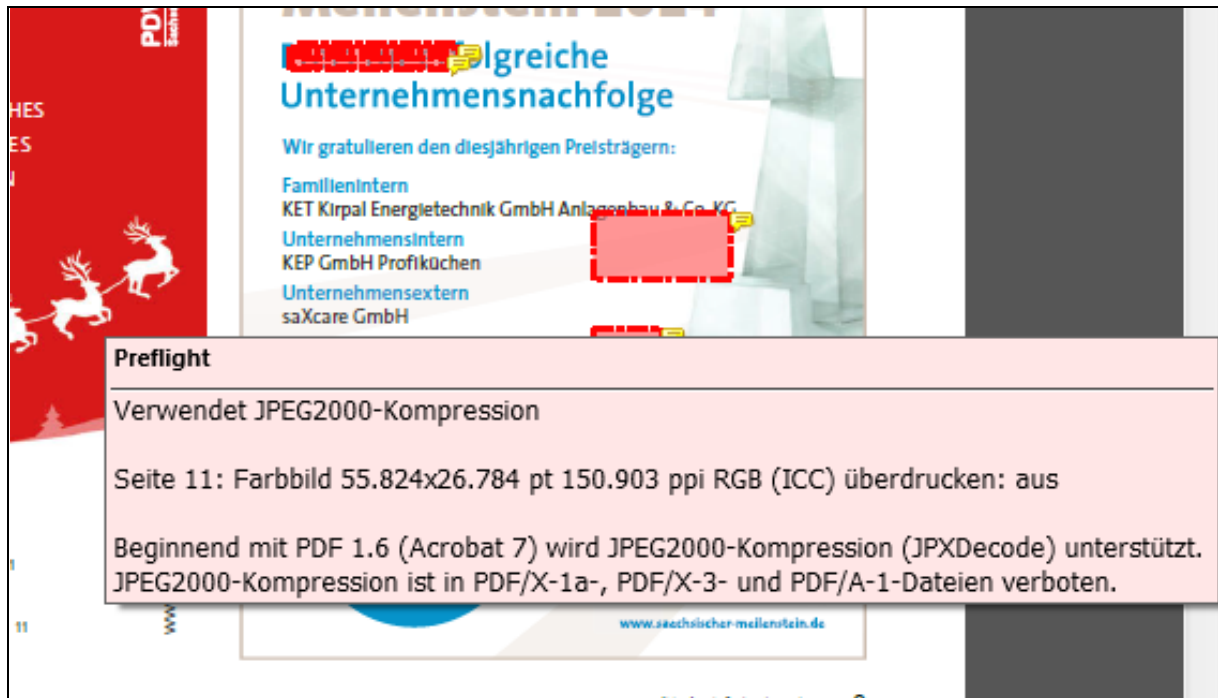


Abbildung 8: „JPEG2000-Kompression“

Lösung

„Acrobats PDF-Optimierung mit JPEG oder ZIP-Komprimierung (ohne Berechnung) verwenden, oder als PDF 1.4 sichern“ (Drümmer, Oettler and Seggern, 2007, p.78).

Diese Lösung ist jedoch nur möglich, wenn keine anderen Inhalte wie OpenType-Schrift eingebettet sind. Am besten ist es hier, wenn in das Quelldokument nur ZIP, LZW und JPEG komprimierte Bilder eingefügt werden.

2.9 Transparenz verwendet

Ursache

„In dieser PDF-Datei sind Objekte auf „transparent“ (durchscheinend) gesetzt. PDF/A verlangt, dass alle in einer PDF-Datei verwendeten Merkmale auf vorhersehbare Weise am Monitor und im Druck dargestellt werden können. Da dies für die Verrechnung von transparenten Objekten mit ihrem Hintergrund nicht gewährleistet ist, ist Transparenz in PDF/A-Dateien nicht erlaubt“ (Drümmer, Oettler and Seggern, 2007, p.79).

Dies bezieht sich nur auf PDF/A-1! PDF/A-2 erlaubt transparente Elemente (<http://de.wikipedia.org/wiki/PDF/A>). Es ist jedoch möglich, dass dann andere Inhalte dem PDF/A-2-Standard nicht entsprechen und es einfacher wäre, den PDF/A-1-Standard einzuhalten.

2 Review of Results for the Eigenvalue Problem of the 1-Laplace Operator

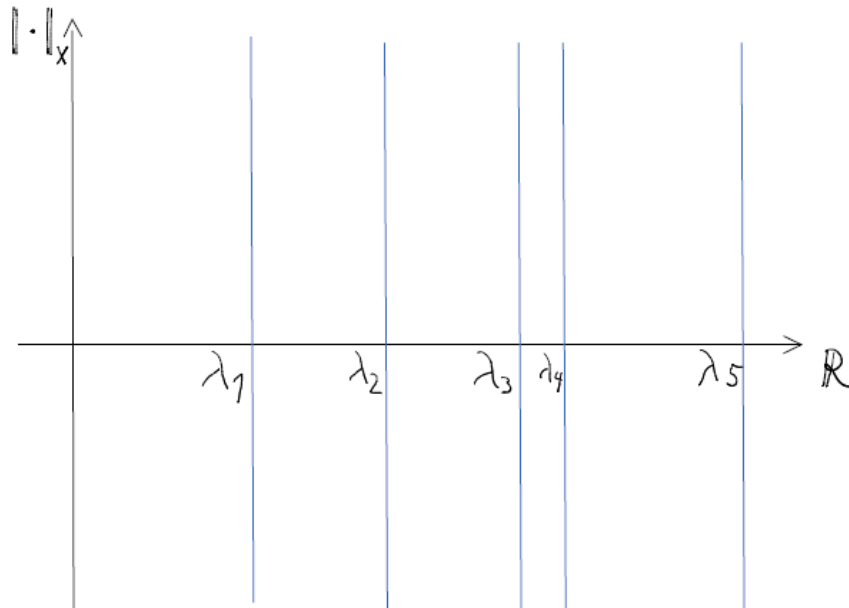


Figure 2.2: Structure of the eigensolutions (λ, u) .

Schnelle Notizen Seite 1

Abbildung 9: „Transparenz verwendet“

2 Review of Results for the Eigenvalue Problem of the 1-Laplace Operator

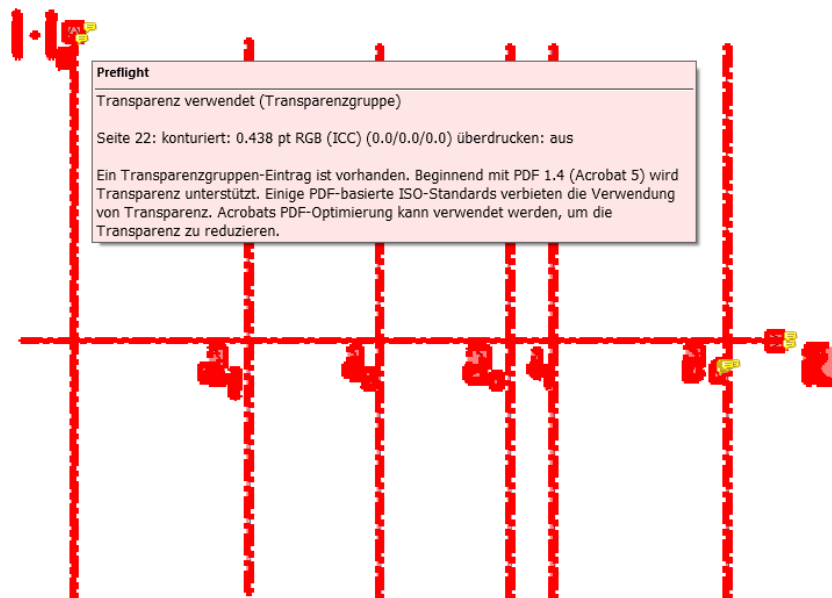


Figure 2.2: Structure of the eigensolutions (λ, u) .

Abbildung 10: „Transparenz verwendet“

Lösung

„ADOBE ACROBAT PROFESSIONAL (Version 6,7 oder 8) beinhaltet ein Modul zur Transparenz-Reduzierung, das Sie für das Entfernen der Transparenz verwenden können“ (Drümmer, Oettler and Seggern, 2007, p.79).

Eine weitere Möglichkeit besteht darin, das Objekt in einem Bildbearbeitungsprogramm erneut ohne Transparenz zu speichern. Außerdem ist es im Quelldokument natürlich möglich, das Objekt zu bearbeiten.

2.10 XMP-Eigenschaft im Erweiterungsschema weder vordefiniert noch definiert

2001 wurde Extensible Metadata Platform (XMP) Metadaten veröffentlicht und 2012 ist der Kernteil der XMP-Spezifikation als ISO-Standard ISO 16684-1 erschienen⁵.

Ursache

Die Benennung des Fehlers beschreibt die Ursache schon sehr gut: XMP-Eigenschaften, die erwartet werden sind nicht vordefiniert und nicht definiert.

Lösung

Diese Fehlermeldung wird in PDF/A kompakt (Drümmer, Oettler and Seggern, 2007) weder aufgeführt, noch erläutert. Obwohl sich viele XMP-Fehler durch eine Preflight-Konversion korrigieren lassen, kann es vorkommen, dass PREFLIGHT den Fehler nicht beheben kann. Der XMP-Validator⁶ liefert genauere Fehlerhinweise. Eine Korrektur des Fehlers ist jedoch nur individuell und wahrscheinlich nur durch eine erneute Erstellung der Datei möglich. Auch der 3-HEIGHTS™ PDF VALIDATOR kann aufschlussreiche Informationen liefern. Kompliziert wird eine Lösung, wenn ein Dokument mit mehreren Programmen erstellt werden muss (zum Beispiel ADOBE INDESIGN und ADOBE PROFESSIONAL PRO).

2.11 XMP-Eigenschaft ist vordefiniert, aber nicht entsprechend der Definition verwendet

Ursache

Dieser Fehler tritt häufig mit 2.10 XMP-Eigenschaft im Erweiterungsschema weder vordefiniert noch definiert und aus ähnlichen Gründen auf, weshalb auf eine ausführliche Beschreibung verzichtet wird.

2.12 „EmbeddedFiles“-Eintrag im Namens-Verzeichnis (Dictionary)

Ursache

Dieser Fehler tritt häufig in Zusammenhang mit 2.13 Enthält EF-Eintrag für eingebettete Dateien auf. Der Grund sind Anlagen, die der Datei beigelegt wurden, wie zum Beispiel „joboptions“-Dateien.

⁵ http://de.wikipedia.org/wiki/Extensible_Metadata_Platform

⁶ <http://www.pdflib.com/knowledge-base/xmp-metadata/free-xmp-validator/>

Lösung

Eine Korrektur ist mit ADOBE ACROBAT PROFESSIONAL möglich. Dazu werden folgende Schritte durchgeführt:

- Aufrufen der Datei
- *Anlagen* in der Leiste links aktivieren
- mit rechter Maustaste die Datei in Anlagen anklicken
- *Anlage löschen* anklicken

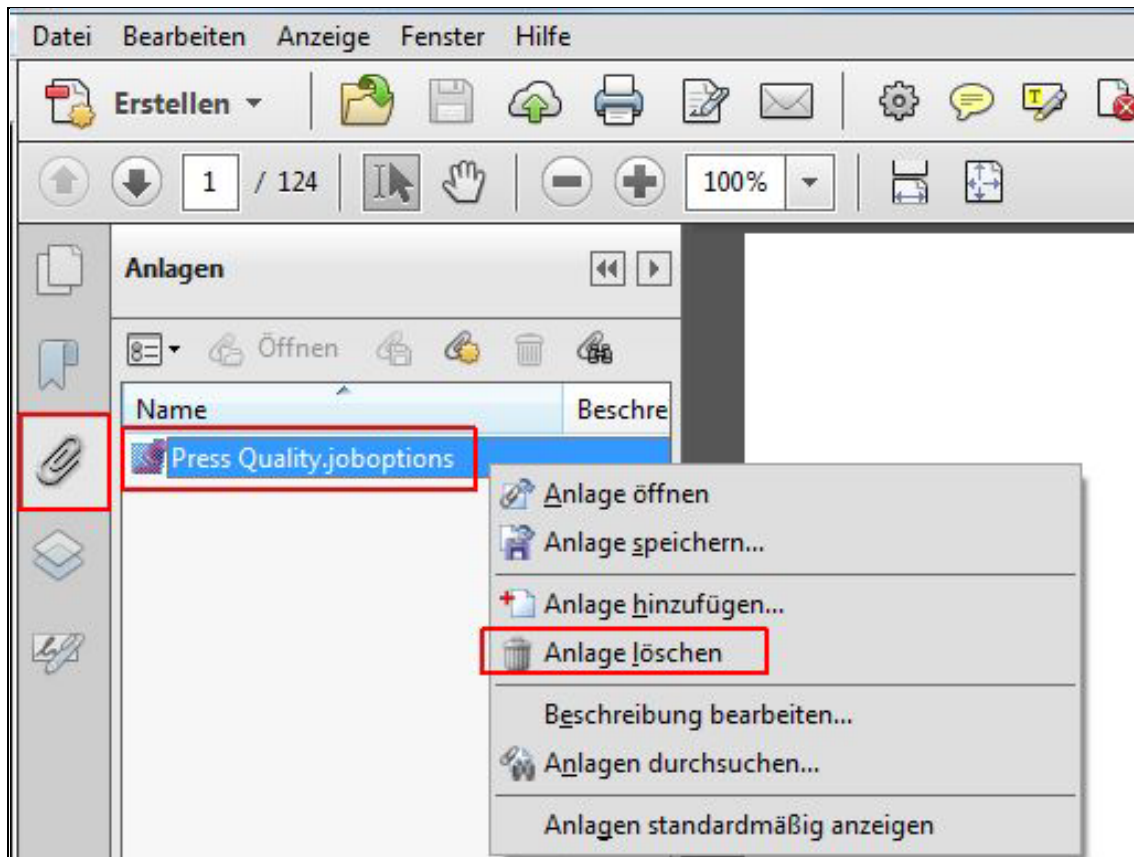


Abbildung 11: „Anlage löschen“

Es öffnet sich ein neues Fenster, in dem

- *Ja* bestätigt wird

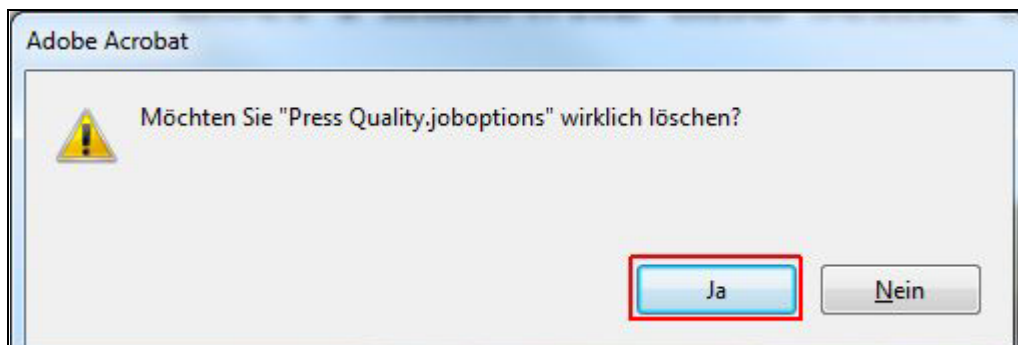


Abbildung 12: „Anlage löschen bestätigen“

Danach ist die Anlage gelöscht und die Fehlermeldung tritt nicht mehr auf.

2.13 Enthält EF-Eintrag für eingebettete Dateien

Ursache

Dieser Fehler tritt in Zusammenhang mit 2.12 „EmbeddedFiles“-Eintrag im Namensverzeichnis (Dictionary) auf. Die bisherige Erfahrung zeigt, dass eine Korrektur des EF-Eintrags auch diesen Fehler behebt.

Zusätzlich muss die Datei jedoch noch in ADOBE ACROBAT PROFESSIONAL nach PDF/A konvertiert werden, weil der EF-Eintrag der Anlage nicht durch das Löschen der Anlage entfernt wird.

3 Zusammenfassung

In diesem Dokument können nicht alle Fehler beschrieben werden, die auftreten, wenn eine PDF-Datei nach PDF/A-Standard gespeichert werden soll. Die aufgeführten Beispiele sollen jedoch in Verbindung mit PDF/A kompakt (Drümmer, Oettler and Seggern, 2007) zeigen, wie Fehler schon während der Erstellung des Originaldokuments vermieden werden. Zu beachten sind vor allem folgende Bestandteile:

- Bilder (Transparenz, Kompression, Farbraum)
- Schriften (zugelassene Schriftarten, Einbettung der Schriftarten mit allen Zeichen)
- Metadaten (korrekte XMP-Metadaten)
- Größe der Datei

4 Literatur

Drümmer, O., Oettler, A. and Seggern, D. von, 2007. *PDF/A kompakt: digitale Langzeitarchivierung mit PDF*. [online] Callas Software GmbH. Available at: <http://www.pdfa.org/wp-content/uploads/2011/08/PDFA_kompakt_pdfa1b.pdf>.