

L^AT_EX und die Archivierbarkeit von PDF-Dateien

Vortrag bei DANTE 2015

Günter Partosch¹

18. 4. 2015

¹Gunter.Partosch@hrz.uni-giessen.de

- 1 Zusammenfassung
- 2 Einleitung
- 3 Eigenschaften von PDF/A
- 4 Anstehende Arbeiten
- 5 Einbetten von Fonts und anderer Ressourcen
- 6 Metadaten
- 7 Farbprofile
- 8 Zuordnung Glyphen zu Unicode
- 9 Abschaltung der LZW-Komprimierung
- 10 Ausgabe
- 11 Links

Archiv

 Günter
Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Mit Hilfe von pdf \LaTeX lassen sich mit relativ geringem Aufwand hochwertige PDF-Dateien erzeugen. Falls diese PDF-Dateien aber in Repositorien oder Dokumenten-Servern eingestellt werden sollen, sind zusätzliche Arbeiten notwendig.

Beispielsweise sollen solche Dateien mögliche Technologie- und Plattformwechsel schadlos überstehen, d. h. sie sollen auch nach mehreren Jahren identisch auf verschiedenen Rechnern reproduzierbar sein — was Schriften, Farben, Layout, usw. betrifft. Das bedeutet u. a. auch, dass alle verwendeten Fonts, Bilder, Farbprofile, Metadaten usw. in die PDF-Datei eingebettet werden müssen.

Im Vortrag wird gezeigt, was dazu in \LaTeX zu tun ist. Das Ergebnis ist eine PDF-Datei, die PDF/A-1b gehorcht.

Archiv

Günter
Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

- Ich bin eifriger \LaTeX -Nutzer – vor allem mit pdf\LaTeX .
- Mit seiner Hilfe generiere ich PDF-Dateien.
- Diese PDF-Dateien sind durchgängig von guter Qualität.
- Sie sind durchsuchbar und enthalten Bookmarks (Lesezeichen), anklickbare Links, Verzeichnisse, Verweise, usw.

- Schon seit einigen Jahren stellt die Universitätsbibliothek der Universität Gießen Abschlussarbeiten in die Gießener Elektronische Bibliothek (GEB, ein Repositorium) ein. – Sie werden archiviert.
- Solche Dateien sollten mögliche Technologie- und Plattformwechsel schadlos überstehen, d. h. sie sollen auch nach mehreren Jahren identisch auf verschiedenen Rechnern reproduzierbar sein – was Schriften, Farben, Layout, usw. betrifft.
- Das bedeutet u.a. auch, dass alle verwendeten Fonts, Bilder, Farbprofile, Metadaten usw. in die PDF-Datei eingebettet werden müssen. Dazu sind zusätzliche Arbeiten notwendig.
- Für Word-Dateien lassen sich diese Anforderungen kaum realisieren, wohl aber bei PDF-Dateien, die aus Word-Dokumenten (auch aus LibreOffice- oder OpenOffice-Dokumenten) erstellt wurden.
- Für diese Zwecke stellt die UB Gießen etablierte Workflows zur Verfügung.
- Für \LaTeX gibt es nichts Entsprechendes.

Archiv

Günter
Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Die erforderlichen PDF-Eigenschaften können im **PAC** (=PDF Accessibility Checker) [[xyMedia 2013](#)] bzw. **Acrobat Reader** eingesehen und im **Adobe Acrobat Professional** korrigiert werden. Hier einige skizzenhafte Hinweise für das Vorgehen im Adobe Acrobat Professional:

- **Metadaten:** *Datei* → *Eigenschaften* → *Beschreibung* → *zusätzliche Metadaten* → *Erweitert*
- **PDF/A-Konformität:** *Werkzeuge* → *Druckproduktion* → *Preflight*
- **Font-Einbettung:** *Datei* → *Eigenschaften* → *Schriften*
- **Farb-Management:** *Werkzeuge* → *Druckproduktion* → *Farben konvertieren*

Das gewünschte Verhalten der PDF-Datei kann in einer Reihe von PDF/A-Normen [[Wikipedia 2014](#)] beschrieben werden, beispielsweise:

- **PDF/A-1b** (Basic) conformance: basiert auf PDF 1.4; eindeutige visuelle Reproduzierbarkeit; alle verwendeten Bilder und Schriften in der Datei eingebettet; Farben mittels geeigneter Farbprofile definiert; Transparenz, JavaScript, LZW-Komprimierung und Verschlüsselung nicht erlaubt; eindeutige Beschreibung der Datei durch Metadaten
- **PDF/A-1a** (Accessible) conformance: eindeutige visuelle Reproduzierbarkeit; Text vollständig in Unicode darstellbar; inhaltliche Strukturierung des Dokuments (»tagged« PDF) → Barrierefreiheit
- **PDF/A-2b**: basiert auf PDF 1.7 (ISO 32000-1); JPEG-2000-Kompression, Transparenz und Ebenen erlaubt; sonst weitgehend wie PDF/A-1a
- **PDF/A-2a** zusätzlich zu PDF/A-2b: realisiert vollständig die Norm ISO 19005-2 (alle strukturellen und semantischen Eigenschaften einer PDF-Datei → Barrierefreiheit)
- **PDF/A-2u** zusätzlich zu PDF/A-2a: gesamter Text in Unicode dargestellt

Archiv

Günter Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Da derzeit pdf^LA^TE^X keine Strukturinformationen generieren und bestenfalls PDF 1.4 erzeugen kann, soll im Folgenden nur PDF/A-1b betrachtet werden. Das notwendige Procedere wird übrigens in [[River Valley 2010](#)] ausführlich beschrieben.

- Einbetten von Fonts und anderer Ressourcen
- Metadaten
- Farbprofile
- Zuordnung von Glyphen zu Unicode
- Abschaltung der LZW-Komprimierung

PDF/A-1b fordert, dass alle verwendeten Bilder und Schriften in der PDF-Datei enthalten sein müssen.

- Wenn Sie auf OpenType-Fonts verzichten, werden bei der PDF-Generierung mittels pdf \LaTeX üblicherweise alle Schriften eingebettet.
- Bei eingebetteten Graphiken werden die für die Darstellung notwendigen Daten nur dann integriert, wenn alle notwendigen Informationen bereits in den verknüpften Dateien enthalten sind.

Archiv

Günter Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Im Zusammenspiel von pdf \LaTeX und hyperref (mit korrekter Verwendung der Anweisung `\hypersetup`) sind einige Metadaten schon vereinbart.

Was jetzt noch zu tun ist, wird im folgenden Listing skizziert.

Ertüchtigung von \LaTeX für PDF/A-1b, skizziert

```

1 ...
2 \usepackage{hyperxmp} % XMP-Daten fuer die PDF-Datei
3 ...
4 \usepackage[pdftex,    % pdfTeX als Prozessor
5                       pdfa] % nimmt Ruecksicht auf PDF/A
6                       {hyperref} % Hypertextstrukturen ermoeeglichen
7 ...
8 \hypersetup{...}      % ggf. Vereinbarung weiterer Metadaten
9 ...

```

Alternative

```
\usepackage{xmpinclude}
```

Archiv

 Günter
 Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

- **Zeile 2:** Mit Hilfe des Pakets `hyperxmp` [Pakin 2014] können u. a. zusätzliche Metadaten in die PDF-Datei geschrieben werden. Weiterhin informiert es das Paket `hyperref` über weitere mögliche Metadaten-Angaben bei `\hypersetup`.
- **Zeile 4—6:** Das Paket `hyperref` [Rahtz et al 2012a, Rahtz et al 2012b] ermöglicht — zusammen mit `pdfLATEX` — Hypertextstrukturen in der Ausgabedatei.
 - Die Option `pdftex` deutet an, dass `pdfTEX` als Prozessor benutzt wird;
 - die Option `pdfa` bewirkt, dass die PDF-Datei das Label PDF/A erhält und verhindert weitgehend, dass PDF/A-Eigenschaften beim Generieren der PDF-Datei verletzt werden.
- **Zeile 8:** Aufruf des `hyperref`-Befehls `\hypersetup` mit zusätzlichen Metadaten; konkret könnte der Aufruf wie im folgenden Listing aussehen:

Archiv

Günter
Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Erweiterung des Aufruf von \hypersetup für zusätzlicher Metadaten

```

1 \hypersetup{% Setup fuer PDF-/Hypertext-Generierung + Metadaten
2 pdftitle           = {Titel ohne Umlaute},
3 pdfauthor          = {Autor(en)},
4 pdfsubject         = {Untertitel ohne Umlaute},
5 pdfkeywords        = {Schluesselwoerter},
6 pdflang            = de,
7 bookmarks          = true,
8 pdfdisplaydoctitle = true,
9 colorlinks         = true,
10 plainpages        = false,
11 %allcolors = black,
12 hypertexnames     = false,
13 pdfpagelabels     = true,
14 hyperindex        = true,
15 unicode            = true,

```

Erweiterung des Aufruf von \hypersetup für zusätzlicher Metadaten

```

16 pdfcaptionwriter = {Gunter Partosch},
17 pdfcontactaddress = {HRZ; Heinrich-Buff-Ring 44},
18 pdfcontactcity = {Giessen},
19 pdfcontactpostcode = {35392},
20 pdfcontactcountry = {Deutschland},
21 pdfcontactregion = {Hessen},
22 pdfcontactemail = {Gunter.Partosch@hrz.uni-giessen.de},
23 pdfcontactphone = {0641-99-13055},
24 pdfcontacturl = {http://www.staff.uni-giessen.de/partosch/}
25 pdfmetalang = {de},
26 }

```

Archiv

Günter
Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Archiv

Günter Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

- 2 `pdftitle={Titel}` wird bei den Metadaten eingetragen. Diese Eigenschaften lassen sich im Acrobat Reader oder Adobe Acrobat Professional über *Datei* → *Eigenschaften* → *Beschreibung* anzeigen.
- 3 `pdfauthor={Autor(en)}` wird bei den Metadaten eingetragen.
- 4 `pdfsubject={Untertitel}` wird bei den Metadaten eingetragen.
- 5 `pdfkeywords={Schlüsselwörter}` werden bei den Metadaten eingetragen.
- 6 `pdflang=de` Sprache des Dokuments ist Deutsch. Diese Information kann von einem Screenreader genutzt, um den Text des Dokument korrekt vorlesen zu können.
- 7 `bookmarks=true` Lesezeichen werden generiert.
- 8 `pdfdisplaydoctitle=true` Anstelle des Dateinamens wird der Titel des Dokuments in der Titelleiste ausgegeben und von einem Screenreader vorgelesen.
- 9 `colorlinks=true` Hypertext-Links werden farbig dargestellt.
- 11 `allcolors=black` Die farbige Darstellung der Hypertext-Links wird abgeschaltet. Die Option ist dann sinnvoll, wenn die PDF-Datei auf einem Schwarz-Weiß-Drucker ausgegeben werden soll.
- 15 `unicode=true` Metadaten können Unicode-Zeichen enthalten.

Archiv

Günter
Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

- 16 pdfcaptionwriter={*Meta-Autor*}: Autor der Metadaten-Einträge
- 17 pdfcontactaddress={*Adresse*}: Anschrift des Metadaten-Autors
- 18 pdfcontactcity={*Stadt*}: dazu Stadt
- 19 pdfcontactpostcode={*PLZ*}: dazu Postleitzahl
- 20 pdfcontactcountry={*Staat*}: dazu Staat
- 21 pdfcontactregion={*Land*}: dazu Bundesstaat / Bundesland
- 22 pdfcontactemail={*E-Mail*}: E-Mail-Adresse des Metadaten-Autors
- 23 pdfcontactphone={*Telefon*}: Telefonnummer des Metadaten-Autors
- 24 pdfcontacturl={*Kontakt-URL*}: Kontakt-Seite des Metadaten-Autors
- 25 pdfmetalang={*Meta-Sprache*}: Kürzel für die Sprache der Metadaten

Damit eine PDF-Datei auch nach einiger Zeit reproduzierbar ist, müssen die zugrunde liegenden Farben und Farb-Schemata eindeutig festgelegt sein. Im Folgenden wird das Procedere zur Vereinbarung eines Farbprofils kurz skizziert:

- ❶ Besorgen Sie sich in einem ersten Schritt ein passendes Farbprofil, beispielsweise von ColorManagement [[Color Management 2015](#)]. Wenn Sie sich unsicher sind, welches Profil Sie verwenden sollten, empfiehlt sich im Zweifelsfall die Verwendung von ISO Coated v2 300% (ECI), beispielsweise `ISOcoated_v2_300_bas.icc`
- ❷ Legen Sie diese Datei an einem Ort innerhalb Ihrer $\text{T}_{\text{E}}\text{X}/\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ -Installation ab, wo sie von $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ gefunden werden kann, z. B.
 - im aktuellen Projektordner oder
 - an einem geeigneten Ort im lokalen TeXMF-Baum oder
 - in dem Ordner, in dem [hyperxmp.sty](#) aufbewahrt wird (auf meinem PC beispielsweise `C:/Program Files (x86)/MiKTeX 2.9/tex/latex/hyperxmp`).
- ❸ Ergänzen Sie Ihre Präambel durch folgendes Code-Schnipsel:

Archiv

Günter Partosch

Abstract

Einleitung

PDFA

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links

Einbinden eines Farbprofils

```
\immediate\pdfobj stream attr{/N 3} file{ISOcoated_v2_300_bas.ICC}
\pdfcatalog{%
/OutputIntents [ <<
/Type /OutputIntent
/S/GTS_PDFa1
/DestOutputProfile \the\pdflastobj\space 0 R
/OutputConditionIdentifier (ISOcoated v2 300 basic)
/Info(ISOcoated v2 300 basic)
>> ]
}
```

[Archiv](#)[Günter Partosch](#)[Abstract](#)[Einleitung](#)[PDF/A](#)[anstehend](#)[Fonts](#)[Metadaten](#)[Profile](#)[Glyphen](#)[LZW](#)[Ausgabe](#)[Links](#)

Die Norm PDF/A-1b fordert eine eindeutige Zuordnung von Glyphen zu Unicode.

Fehlermeldung beim Preflight: *Text cannot be mapped to Unicode.*

Ergänzen Sie die Präambel deshalb durch folgenden Code-Schnipsel:

Zuordnung Glyphen zu Unicode

```
\input{glyphtounicode.tex}  
\input{glyphtounicode-cmr.tex}  
\pdfgentounicode=1
```

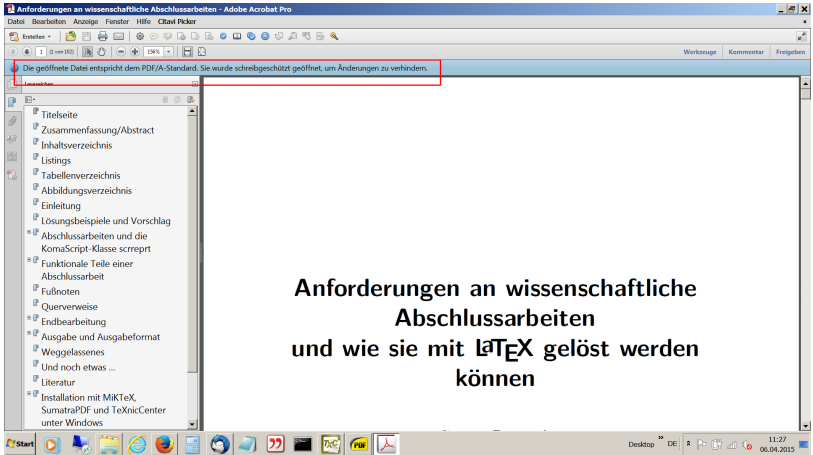
Die Norm PDF/A-1b fordert, dass die PDF-Datei keine LZW-Komprimierung enthält. Das können Sie erreichen, indem Sie die Präambel durch folgendes Code-Schnipsel ergänzen:

Ausschalten der LZW-Komprimierung

```
\pdfobjcompresslevel=0  
\pdfinclusioncopyfonts=1
```

[Archiv](#)[Günter Partosch](#)[Abstract](#)[Einleitung](#)[PDF/A](#)[anstehend](#)[Fonts](#)[Metadaten](#)[Profile](#)[Glyphen](#)[LZW](#)[Ausgabe](#)[Links](#)

„Die geöffnete Datei entspricht dem PDF/A-Standard. Sie wurde schreibgeschützt geöffnet, um Änderungen zu verhindern.“



**Anforderungen an wissenschaftliche
Abschlussarbeiten
und wie sie mit L^AT_EX gelöst werden
können**

Archiv

Günter Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links



Color Management: *Color Management*. 2009.

<http://www.colormangement.org/de/download.html> (besucht am 05.04.2015)



Wikipedia: *PDF/A*. Hrsg. von Wikipedia. 2014. url:

<http://de.wikipedia.org/w/index.php?oldid=130234016> (besucht am 05.04.2015)



River Valley: *Generating PDF/A compliant PDFs from pdfTEX*. Hrsg. von River Valley. 2010. url: http://support.river-valley.com/wiki/index.php?title=Generating_PDF/A_compliant_PDFs_from_pdftex (besucht am 05.04.2015)



Scott Pakin: *The hyperxmp package*. 2014. url:

<http://ctan.mirrorcatalogs.com/macros/latex/contrib/hyperxmp/hyperxmp.pdf> (besucht am 05.04.2015)



Sebastian P Q Rahtz und Heiko Oberdiek: *hyperref package options*. 2012. url: <http://ctan.tug.org/macros/latex/contrib/hyperref/doc/options.pdf>

(besucht am 05.04.2015)

Archiv

Günter Partosch

Abstract

Einleitung

PDF/A

anstehend

Fonts

Metadaten

Profile

Glyphen

LZW

Ausgabe

Links



Sebastian P Q Rahtz und Heiko Oberdiek: *Hypertext marks in L^AT_EX. a manual for hyperref.* 2012. url: <http://ctan.tug.org/macros/latex/contrib/hyperref/doc/manual.pdf>

(besucht am 05.04.2015)



xyMedia: *PDF Accessibility Checker (PAC) 2.0. PDF-Dokumente auf ISO Standard PDF/UA prüfen.* Hrsg. von xyMedia. 2013. url:

http://www.xymedia.ch/downloads/PAC_2_Uebersicht_de.pdf
(besucht am 05.04.2015)