

# PDF/A-tiedoston tekeminen L<sup>A</sup>T<sub>E</sub>Xilla

Jarmo Niemelä (jarmo.niemela@tuni.fi)

27. syyskuuta 2023

## Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>PDF/A-tiedoston tekeminen L<sup>A</sup>T<sub>E</sub>Xilla</b>	<b>2</b>
<b>3</b>	<b>PDF/A-tiedoston tarkistus ja validointi</b>	<b>4</b>
<b>4</b>	<b>Suosituksia ja ongelmia</b>	<b>4</b>
<b>5</b>	<b>Muuntaminen PDF/A-muotoon muilla ohjelmilla</b>	<b>7</b>

## 1 Johdanto

PDF/A on pdf-tiedostoformaatin standardoitu versio, joka on tarkoitettu dokumenttien pitkäaikaista arkistointia varten. PDF/A-muotoisen tiedoston on sisällettävä kaikki sen lukemiseen ja tulostamiseen tarvittavat elementit kuten kuvat, väriprofiili ja fontit. Äänen, liikkuvan kuvan ja tiedoston salauksen käyttö on kielletty.

PDF/A-standardista on [neljä versiota](#): PDF/A-1, PDF/A-2, PDF/A-3 ja PDF/A-4. Versioissa 1–3 on vaatimustasot a ja b. Taso a (accessible) täyttää kaikki standardin vaatimukset. Tason b (basic) tarkoitus on varmistaa vain dokumentin visuaalisen ulkoasun säilyminen. Versioissa 2 ja 3 on lisäksi vaatimustaso u (Unicode), joka eroaa tasosta b siten, että dokumentin kaikki teksti on voitava esittää Unicode-muodossa. PDF/A-4 ei käytä vaatimustasoja a, b ja u, mutta määrittelee uuden tason f, joka vastaa versiota PDF/A-3. Yleisesti ottaen PDF/A-4 on vaatimuksiltaan sallivampi kuin standardin aiemmat versiot.

Tampereen yliopiston [opinnäytetöissä edellytetään PDF/A-muotoa](#). Mikä tahansa PDF/A:n versio käy.

## 2 PDF/A-tiedoston tekeminen L<sup>A</sup>T<sub>E</sub>Xilla

PDF/A-muotoinen tiedosto tehdään L<sup>A</sup>T<sub>E</sub>Xilla seuraavasti:

```
\pdfminorversion=7
\begin{filecontents*}[overwrite]{\jobname.xmpdata}
  \Title{Dokumentin otsikko}
  \Author{Tekijän nimi}
  \Language{fi}
  \Subject{Dokumentin lyhyt kuvaus tai tiivistelmä.}
  \Keywords{avainsana1\sep avainsana2\sep avainsana3}
\end{filecontents*}
\documentclass[a4paper,12pt]{article}
\usepackage[utf8]{inputenc}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
\usepackage[finnish]{babel}
\usepackage{colorprofiles}
\usepackage[a-2b,mathxmp]{pdfx}[2018/12/22]
\hypersetup{pdfstartview=}
\begin{document}
Dokumentin sisältö tulee tähän.
\end{document}
```

Tiedoston alussa olevalla komennolla `\pdfminorversion=7` asetetaan pdf-versioksi 1.7, johon PDF/A-2 ja PDF/A-3 perustuvat.

### Makropaketti pdfx

L<sup>A</sup>T<sub>E</sub>X-tiedosto saadaan tallennettua PDF/A-muotoon makropaketilla [pdfx](#):

```
\usepackage[a-2b,mathxmp]{pdfx}[2018/12/22]
```

Haluttu PDF/A:n versio ja taso annetaan valinnaisella argumentilla, jonka oletusarvo on `a-1b`. Valinnainen argumentti `mathxmp` mahdollistaa matemaattisten merkkien käytön [metatiedoissa](#) sekä korjaa eräitä virheitä metatietojen esittämisessä, joten sitä kannattaa käyttää aina. Komennon lopussa oleva valinnainen argumentti `[2018/12/22]` tarkoittaa, että makropaketista ei hyväksytä kyseistä päivämäärää vanhempia versioita. Makropaketin pdfx toiminta on vielä osin puutteellista, ja vanhemmissa versioissa noita puutteita ja ongelmia on enemmän.

Makropaketti pdfx kutsuu muun muassa makropaketteja hyperref ja xcolor, joten näitä ei tarvitse kutsua erikseen. Makropaketin hyperref optioita voi asettaa komennolla `\hypersetup`.

## Väriprofiili

PDF/A-tiedoston edellyttämä väriprofiili otetaan käyttöön makropaketilla `colorprofiles`, jota `pdfx` kutsuu automaattisesti. Makropakettia `colorprofiles` ei kuitenkaan kutsuta, jos sitä ei ole asennettu, joten se täytyy tarvittaessa asentaa ensin. Jos  $\LaTeX$ -ohjelmistona on `MiKTeX`, puuttuvat makropaketit asennetaan yleensä automaattisesti, jolloin riittää, että lisätään komento

```
\usepackage{colorprofiles}
```

ennen makropaketin `pdfx` kutsua. Kun makropaketti `colorprofiles` on asennettu, ei tätä komentoa enää tarvita.

## Metatiedot

PDF/A-standardi edellyttää, että dokumentin metatiedot, kuten dokumentin otsikko, tekijän nimi ja avainsanat, esitetään määrätyssä muodossa. Makropaketti `pdfx` lukee nämä metatiedot tekstitiedostosta `\jobname.xmpdata`, missä komento `\jobname` sisältää dokumentin päätiedoston nimen, esimerkiksi `pdfa-ohje.xmpdata`. Tiedosto `\jobname.xmpdata` kannattaa luoda päätiedoston alussa ympäristöllä `filecontents*`:

```
\begin{filecontents*}[overwrite]{\jobname.xmpdata}
  \Title{Dokumentin otsikko}
  \Author{Tekijän nimi}
  \Language{fi}
  \Subject{Dokumentin lyhyt kuvaus tai tiivistelmä.}
  \Keywords{avainsana1\sep avainsana2\sep avainsana3}
\end{filecontents*}
```

Ympäristön `filecontents*` optio `overwrite` ei toimi ennen [lokakuuta 2019](#) julkaistuilla  $\LaTeX$ in versioilla, joilla komento

```
\begin{filecontents*}[overwrite]{\jobname.xmpdata}
```

on korvattava komennoilla

```
\RequirePackage{filecontents}
\begin{filecontents*}{\jobname.xmpdata}
```

Käytettävissä olevat metatietokentät ja niiden merkitys on esitetty [makropaketin pdfx käyttöohjeissa](#). Kaikki metatietokentät ovat valinnaisia, mutta on suositeltavaa käyttää ainakin kenttiä `\Title`, `\Author` ja `\Language`.

Dokumentin eri tekijät, kielet ja avainsanat on parasta erottaa toisistaan komennolla `\sep`. Tällöin metatietoja käsittelevät ohjelmat osaavat tulkita ne selvästi erillisiksi.

$\LaTeX$ in erikoismerkit `$`, `&`, `#`, `_`, `~`, `^`, `{` ja `}` tulostuvat metatietokenttiin sellaisinaan. Merkit `%` ja `\` täytyy esittää komennoilla `\%` ja `\textbackslash`.

Tämän vuoksi muun muassa ylä- ja alaindeksejä ei voi esittää metatiedoissa muuten kuin niitä vastaavilla [Unicode-merkeillä](#).

Jos metatietokentät sisältävät  $\LaTeX$ in komennoilla tai [Unicode-muodossa](#) esitettyjä matemaattisia merkkejä, on käytettävä makropaketin pdfx optiota `mathxmp`. Matemaattisia merkkejä ei kuitenkaan pidä ympäröidä dollarin merkeillä, koska dollarin merkit tulostuvat metatietoihin sellaisinaan.

### 3 PDF/A-tiedoston tarkistus ja validointi

Kun PDF/A-tiedosto on valmis, tarkista [Adobe Acrobat Readerissa](#) tai [PDF-XChange Editorissa](#) sen ulkoasu, [kirjanmerkit](#), linkkien toimivuus ja metatiedot (*Tiedosto*  $\rightarrow$  *Ominaisuudet*  $\rightarrow$  *Kuvaus*). Lopuksi on varmistettava, että tiedosto todella noudattaa PDF/A-standardia. Tämä validointi tulisi tehdä aina, kun tiedosto tallennetaan arkistointia varten. Opinnäytteiden kohdalla siis [ennen kuin opinnäyte jätetään arvioitavaksi](#).

PDF/A-tiedostoja voi validoida seuraavilla ohjelmilla:

**3-Heights PDF-validator online tool** Ilmainen verkkosovellus PDF/A-tiedostojen validointiin.

**veraPDF** Ilmainen avoimen lähdekoodin ohjelma PDF/A-tiedostojen validointiin. veraPDF on toteutettu Javalla, joten se edellyttää [Java Runtime Environmentin](#). Katso käyttöohjeet sivulta [docs.verapdf.org/gui](https://docs.verapdf.org/gui).

**Adobe Acrobat Pro** Maksullinen ohjelma, josta on saatavilla 7 päivän kokeiluversio. Acrobat Pro:n Preflight-työkalulla (*Työkalut*  $\rightarrow$  *PDF-standardit*  $\rightarrow$  *Preflight*) voi validoida PDF/A-tiedostoja sekä korjata tiedostoissa mahdollisesti olevat validointivirheet.

**Callas pdfaPilot** Maksullinen ohjelma, josta on saatavilla 14 päivän kokeiluversio. Callas pdfaPilotilla voi validoida ja korjata PDF/A-tiedostoja samaan tapaan kuin Adobe Acrobat Prolla.

Validointiohjelmien antamat tulokset poikkeavat joskus toisistaan. Yleensä riittää, että tiedosto validoituu yhdellä yllä luetelluista ohjelmista. Jos siis pdf-tiedosto läpäisee yhden ohjelman validointitestit, sitä ei tarvitse enää validoida muilla ohjelmilla. Jos taas tiedosto ei mene läpi yhden ohjelman testeistä, voi sen testata toisella ohjelmalla.

### 4 Suosituksia ja ongelmia

#### Kirjanmerkit

Tarkista dokumentin otsikot Adobe Readerin tai PDF-XChange Editorin kirjanmerkeissä. Matemaattisten merkkien käyttöä otsikoissa kannattaa vält-

tää. Mikäli otsikot sisältävät matemaattisia merkkejä tai muita erikoismerkkejä, ne eivät välttämättä näy oikein kirjanmerkeissä. Nämä voidaan usein korjata makropaketin `\texorpdfstring` komennolla `\texorpdfstring{ }{ }`, jonka jälkimmäinen argumentti tulostuu kirjanmerkkeihin. Esimerkiksi

```
\section{Integraali \texorpdfstring
{\displaystyle\int\sqrt{x^2 - a^2}\,dx}
{\int\text{surd}(x\text{ttwosuperior} \text{minus}
a\text{ttwosuperior})\unichar{"2009}dx}}
```

Komennon `\texorpdfstring` jälkimmäisessä argumentissa voi käyttää Unicode-merkkejä, joko kopioimalla ne jostain tai komennolla `\unichar{"XXXX}`, missä `XXXX` on Unicode-merkin heksadesimaalikoodi isoilla kirjaimilla kirjoitettuna.

## Kuva- ja liitetiedostot

L<sup>A</sup>T<sub>E</sub>X-dokumentissa käytettyjen pdf-muotoisten kuva- ja liitetiedostojen on sisällettävä niissä käytetyt fontit, koska muutoin dokumentista tuotettu pdf-tiedosto ei sisällä näitä fontteja eikä siten läpäise PDF/A-validointia. Fontit tulee sisällyttää pdf-muotoisiin kuva- ja liitetiedostoihin ensisijaisesti samoilla ohjelmilla, joilla nämä tiedostot on tehty. Jos tämä ei ole mahdollista, niin fontit saa sisällytettyä ilmaisella ghostscript-ohjelmalla, kuten L<sup>A</sup>T<sub>E</sub>X-kursin esimerkkitiedostossa [kuvat.pdf](#) on esitetty. Toinen vaihtoehto on tehdä L<sup>A</sup>T<sub>E</sub>Xilla tavallinen pdf-tiedosto ja muuntaa se sitten PDF/A-muotoon [jollain toisella ohjelmalla](#). Puuttuvat fontit sisällytetään PDF/A-muunnoksen yhteydessä.

## Makropaketin pdfx rajoitukset

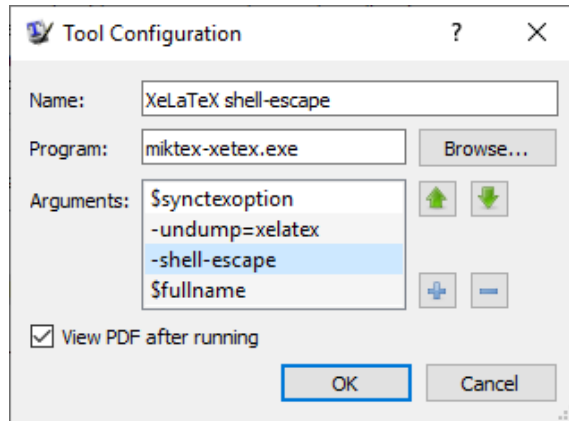
Makropaketilla pdfx ei toistaiseksi voi tehdä tasoa a (1a, 2a, 3a) olevia valideja PDF/A-tiedostoja.

Jos makropaketin pdfx kanssa käytetään makropakettia arev, kpfnts tai mathdesign, niin pdfLaTeXilla ei saa tehtyä validia PDF/A-tiedostoa. Muutkin fontteja määrittelevät makropaketit voivat joissain tapauksissa aiheuttaa validointiongelmia.

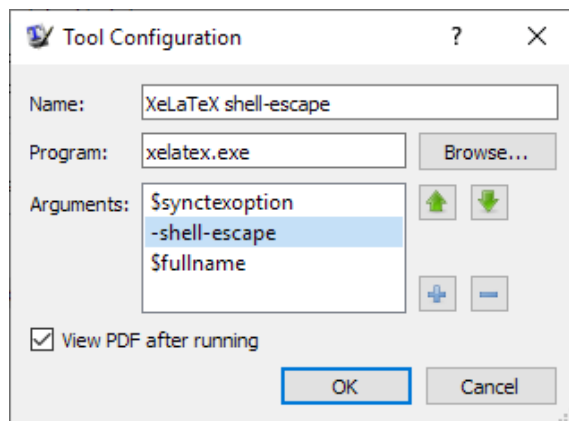
Makropaketti pdfx edellyttää, että XeLaTeXin kanssa käytetään optiota `-shell-escape`. Koska option `-shell-escape` käyttö on potentiaalinen tietoturvariski, kannattaa mieluummin tehdä uusi kääntäjävaihtoehto kuin muuttaa XeLaTeXin oletusarvoisia optioita. TeXworks-editorissa tämä tehdään seuraavasti:

1. Valitse *Edit* → *Preferences* → *Typesetting*.
2. Napsauta kohdan *Processing tools* vieressä olevaa pluspainiketta.

3. Luo uusi kääntäjävaihtoehto ”XeLaTeX shell-escape” kuvan 1 mukaisesti, jos käytät MiKTeX-ohjelmistoa, tai kuvan 2 mukaisesti, jos käytät TeX Live -ohjelmistoa.



Kuva 1: Option `-shell-escape` lisääminen MiKTeXissä.



Kuva 2: Option `-shell-escape` lisääminen TeX Livessä.

Makropaketin hyperref optiolla `pdfstartview` valitaan pdf-tiedoston aloitusnäky katseluohjelmassa. Katso tarkemmin [makropaketin hyperref käyttöohjeista](#). Option arvoja `FitH`, `FitV`, `FitR`, `FitBH` ja `FitBV` täytyy seurata aloituspisteen etäisyys sivun vasemmasta alakulmasta  $1/72$  tuuman yksiköissä. Muutoin seurauksena on validointivirhe [3-Heights PDF Validatorissa](#). Option voi antaa myös muodossa `pdfstartview=`, jolloin pdf-tiedoston aloitusnäkyinä käytetään katseluohjelman oletusarvoa.

## 5 Muuntaminen PDF/A-muotoon muilla ohjelmilla

Jos validin PDF/A-tiedoston tekeminen ei onnistu LaTeXilla, voi muunnoksen tehdä jollain toisella ohjelmalla. Tiedoston metatiedot kannattaa silti tallentaa L<sup>A</sup>T<sub>E</sub>Xilla. Tähän ei välttämättä tarvita makropakettia pdfx, koska metatiedot voidaan tallentaa myös makropaketeilla hyperxmp ja hyperref seuraavasti:

```
\pdfminorversion=7
\documentclass[a4paper,12pt]{article}
\usepackage[utf8]{inputenc}
\usepackage[T1]{fontenc}
\usepackage{lmodern}
\usepackage[finnish]{babel}
\usepackage{hyperref}
\usepackage{hyperxmp}
\hypersetup{%
  pdftitle={Dokumentin otsikko},
  pdfauthor={Tekijän nimi},
  pdflang={fi},
  pdfsubject={Dokumentin lyhyt kuvaus tai tiivistelmä.},
  pdfkeywords={avainsana1, avainsana2, avainsana3},
  pdfstartview=}
\begin{document}
Dokumentin sisältö tulee tähän.
\end{document}
```

Jos metatiedoissa tarvitaan L<sup>A</sup>T<sub>E</sub>Xin erikoismerkkejä, niin makropakettia hyperxmp käytettäessä ne on kirjoitettava niitä vastaavilla komennoilla `\%`, `\$`, `\&` jne. Matemaattiset merkit on kirjoitettava Unicode-merkkeinä.

### Muuntaja-palvelu

Tampereen yliopiston opiskelijat ja henkilökunta voivat muuntaa pdf-tiedostoja PDF/A-muotoon [Muuntaja-palvelulla](#).

### Adobe Acrobat Pro

[Adobe Acrobat Pro](#)lla, voi muuntaa pdf-tiedostoja PDF/A-muotoon: *Työkalut* → *PDF-standardit* → *Tallenna PDF/A-muodossa*. Preflight-työkalulla (*Työkalut* → *PDF-standardit* → *Preflight*) voi korjata PDF/A-tiedoston validointivirheet. Valitse haluamasi muunnosprofiili Preflightin *Profiilit*-välilehdeltä kohdasta *PDF/A* ja sen jälkeen ikkunan oikeasta alakulmasta *Analysoi ja korjaa*.

## Callas pdfaPilot

Callas pdfaPilotilla voi tehdä ja validoida PDF/A-tiedostoja samaan tapaan kuin Adobe Acrobat Prolla. Tiedostojen [metatietojen muokkaaminen](#) on hankalaa, joten metatiedot kannattaa tallentaa L<sup>A</sup>T<sub>E</sub>Xilla.

## Muita ohjelmia

Verkosta löytyy useita sivustoja, joilla voi muuntaa pdf-tiedostoja PDF/A-muotoon. Näistä ainakin [PDFen](#) on toimiva.