

Preporuke za dugoročno čuvanje elektroničkih dokumenata

- [Pohrana elektroničkih dokumenata u Dabru](#)
- Što je PDF/A?
 - [Kako izraditi PDF/A inačicu u Microsoft Word-u?](#)
 - [Kako izraditi PDF/A inačicu u Libre Office-u?](#)
 - [Kako izraditi PDF/A inačicu u LaTeX-u?](#)
- [Preporuke vezane uz metapodatke PDF datoteke](#)
- [Preporuke za digitalizaciju tekstualnih sadržaja](#)

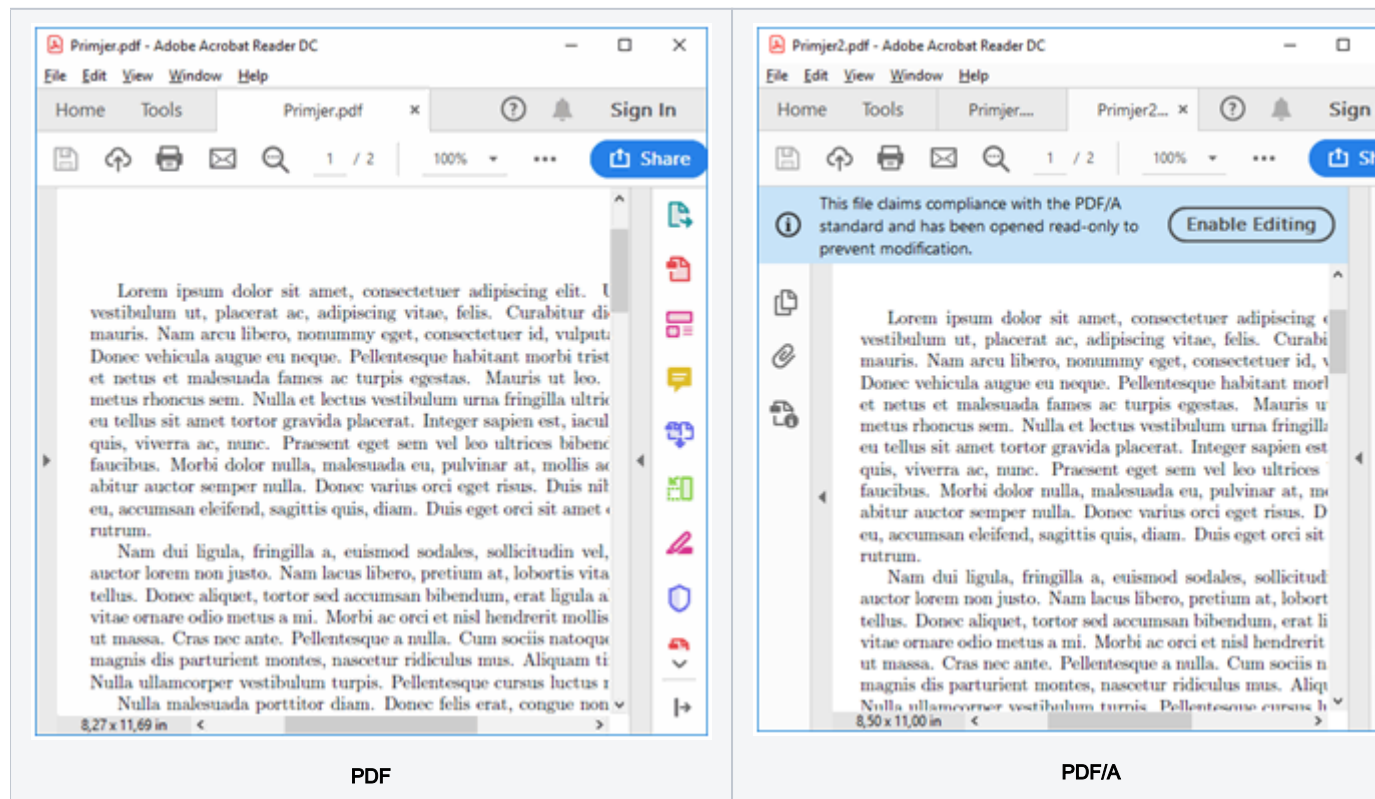
Pohrana elektroničkih dokumenata u Dabru

Za dugoročno čuvanje elektroničkih dokumenata u repozitorijima u Dabru preporuča se korištenje PDF/A inačice PDF formata.

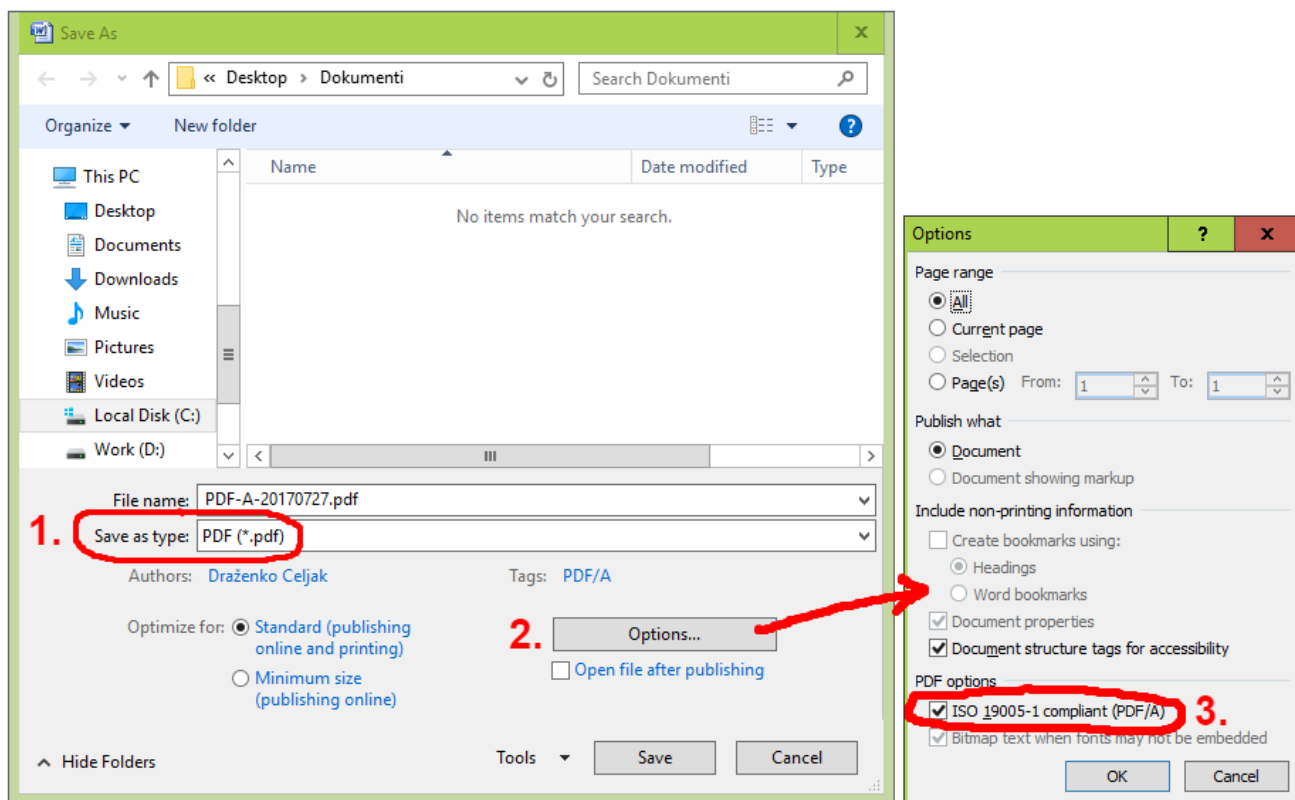
Što je PDF/A?

PDF/A inačica namijenjena je za dugoročno čuvanje elektroničkih dokumenata. Specificirana je standardom ISO 19005 i osigurava samodostatnost dokumenta na način da sve vanjske resurse potrebne za reprodukciju dokumenta (npr. fontove) uključi u PDF datoteku kako bi se dokument mogao reproducirati bez obzira na promjene tehnologije i neovisno o platformi. PDF/A ujedno isključuje mogućnost enkripcije te pokretanja programskih skripti i izvršnih datoteka.

Većina PDF preglednika prilikom otvaranja PDF/A dokumenta prikazuje informaciju da se radi o arhivskom formatu, te da se dokument iz tog razloga prikazuje u modu za čitanje:

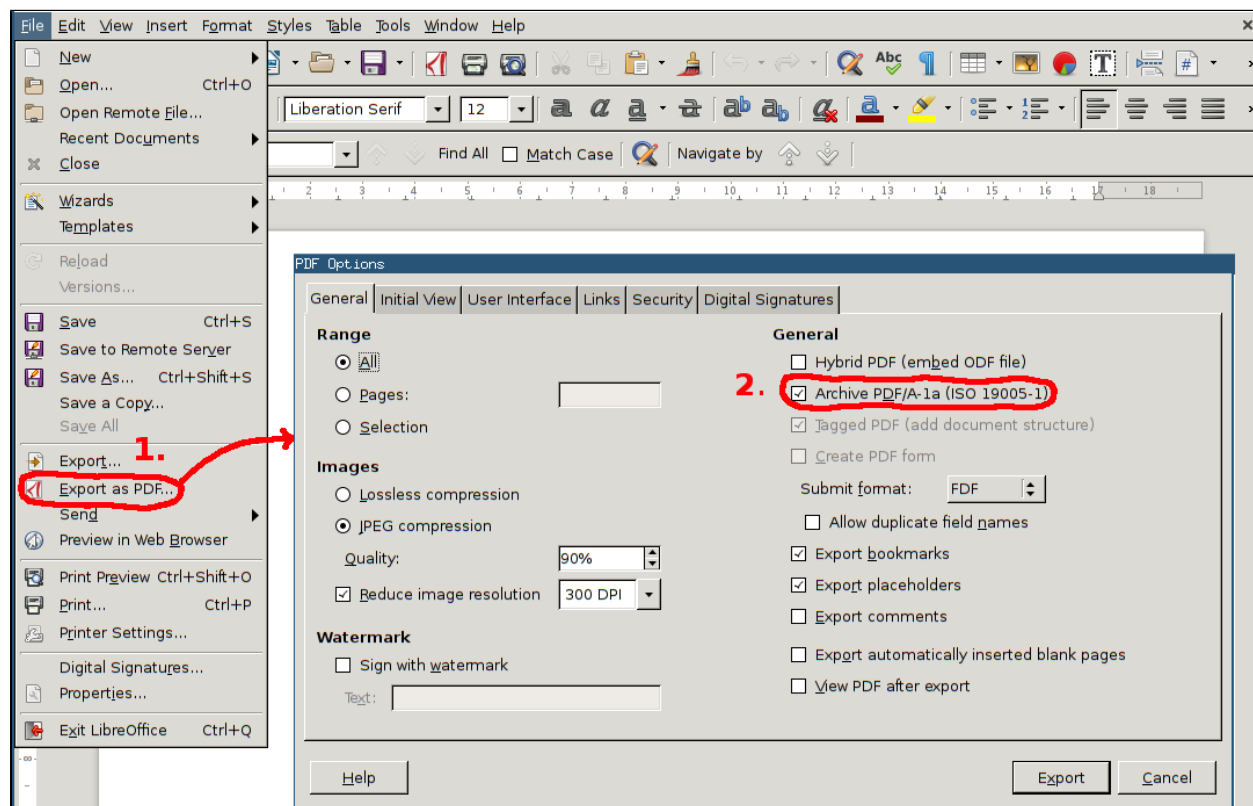


Kako izraditi PDF/A inačicu u Microsoft Word-u?



U većini slučajeva je za kreiranje PDF/A dovoljna verzija PDF/A-1 koja je podržana u programu za obradu teksta Microsoft Word, ali ovisno o verziji PDF/A postoje i neka ograničenja, npr. PDF/A-1 ne dozvoljavaju transparentne slike, pa je ponekad potrebno koristiti više verzije PDF/A (PDF/A-2 ili PDF/A-3) kako bi dokument bio ispravno prikazan.

Kako izraditi PDF/A inačicu u Libre Office-u?



Kako izraditi PDF/A inačicu u LaTeX-u?

Izrada PDF/A dokumenta u LaTeX-u vrši se uključivanjem paketa pdfx u preambuli dokumenta. Primjer:

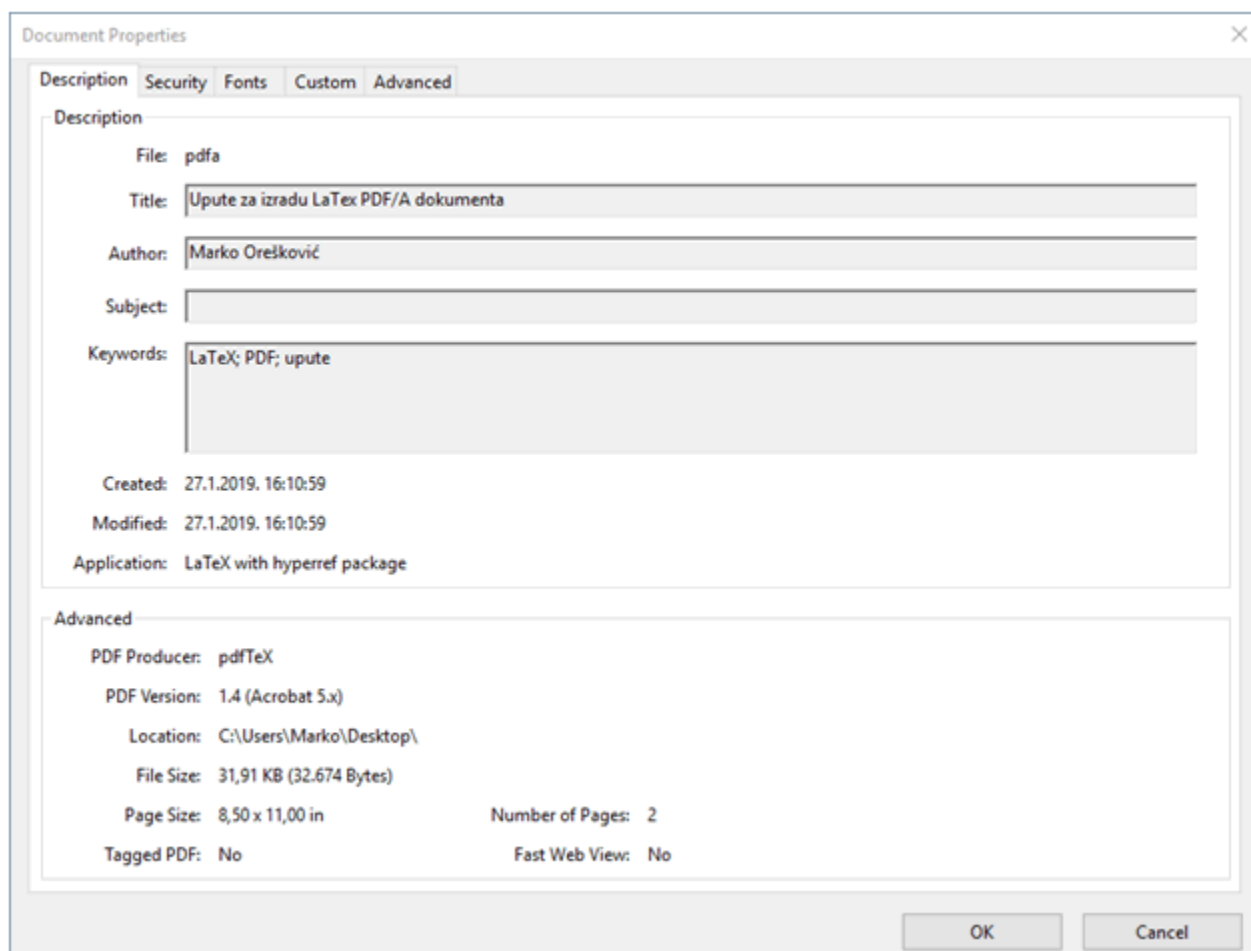
```
\documentclass{article}
\usepackage{lipsum}
\usepackage[a-1b]{pdfx}

\title{Upute za izradu LaTeX PDF/A dokumenta}
\author{Marko Orešković}

\begin{document}
\lipsum
\end{document}
```

Preporuke vezane uz metapodatke PDF datoteke

PDF datoteka treba imati upisane ispravne metapodake poput autora i naslova jer tražilice i baze podataka te podatke nerijetko koriste za indeksiranje. Metapodaci su u PDF čitaču Adobe Reader vidljivi kroz izbornik: *File > Properties*.



Slika: Prikaz metapodataka PDF dokumenta

U programu za obradu teksta Microsoft Word se metapodaci uređuju pomoću opcije Properties do koje se dolazi preko opcije Prepare u izborniku.

U LaTeX-u se metapodaci ugrađuju dodavanjem koda u preambulu:

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

\Title{Upute za izradu LaTeX PDF/A dokumenta}

\Author{Marko Oreškovi}

\Keywords{LaTeX\sep PDF\sep upute}

\Publisher{SRCE}

\end{filecontents*}
```

Preporuke za digitalizaciju tekstualnih sadržaja

Prilikom skeniranja tekstualnih sadržaja potrebno uključiti opciju OCR (Optical Character Recognition) ako ju alat za skeniranje ima ili je za OCR potrebno koristiti neki specijalizirani alat (npr. ABBY FineReader, Tesseract OCR Engine,...).

Proizvođači OCR softvera i literatura obično preporučaju da se pri digitalizaciji skenira s rezolucijom od 300 dpi za fontove veličine 10 točaka i više, a od 400-600 dpi s rezolucijom za fontove veličine 9 točaka i manje.

Zbog arhivske svrhe i pravne valjanosti dokumenta, tekst prepoznat u postupku OCR-a potrebno je spremiti u sloju ispod slike teksta kako bi se sačuvao izvorni izgled dokumenta i istovremeno dobila mogućnost označavanja teksta.