



národní
úložiště
šedé
literatury

Převod dokumentů do PDF pro potřeby NUŠL

Národní technická knihovna; Charvátová, Michaela
2016

Dostupný z <http://www.nusl.cz/ntk/nusl-203659>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte původ-Neužívejte komerčně-Zachovejte licenci 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 27.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

Převod dokumentů do PDF pro potřeby NUŠL

297 mm

Národní úložiště šedé literatury

Národní úložiště šedé literatury (NUŠL) slouží jako agregátor záznamů a zároveň i jako vlastní repozitář dokumentů. Dle [platné typologie](#) jsou do NUŠL přijímány dokumenty digitalizované i tzv. born digital, které už vznikly v elektronické podobě. Jak digitalizované, tak původní digitální dokumenty ovšem vznikají v řadě různých formátů, jejichž množství brání ve sledování vývoje všech jednotlivých formátů. Na základě vybraných parametrů byly vybrány čtyři základní formáty (pdf, mp3, mp4 a jpeg), ve kterých jsou přijímány dokumenty do NUŠL. Rozvoj, obecné užívání i technologický vývoj těchto vybraných formátů je sledován, aby byla v případě potřeby zajištěna včasná konverze formátů archivovaných dokumentů a nehrozila jejich ztráta.

Tento dokument slouží jako manuál pro převod textových dokumentů z dalších často používaných formátů do formátu PDF.

Podoba PDF v NUŠL a na co si dát pozor

V NUŠL není zavedeno omezení na verzi PDF formátu, ale samozřejmě čím novější verze, tím lépe. Zajímavou variantou, která rozhodně stojí za zvážení, je formát [PDF/A](#).

PDF/A

Jedná se o archivační verzi formátu PDF, která je vhodná právě k dlouhodobému ukládání a je definována v příslušných ISO normách. Tento formát zaručuje, že soubor bude samostatně použitelný, tedy nebude vázaný na další externí zdroje a materiály (fonty apod.). Proto při ukládání do PDF/A může dojít ke změnám vzhledu, které je třeba brát v potaz a případně je řešit. Problémem mohou být i třeba obrázky s průhlednou vrstvou (formáty png, gif), kde tato vrstva při převodu zčerná. Tuto situaci lze snadno vyřešit převodem obrázku do jiného formátu (např. jpeg), je ovšem třeba počítat s možnou ztrátou kvality.

Správný převod textu

Při převodu z jiných formátů do PDF je třeba dbát na správný převod textu. Systém Invenio, ale třeba i internetové vyhledávače jsou schopny texty dokumentů zaindexovat a umožňují v nich vyhledávat, a odkazovat tak na ně více případných zájemců.

U digitalizovaných dokumentů je tento problém řešen pomocí [OCR](#). Do NUŠL bude sice přijat i dokument, který bude obsahovat v podstatě jen obrázky stránek, ale takový dokument nebude možné prohledávat ani v Inveniu, ani jinými vyhledávači. Při dalším převodu do jiného formátu (např. PDF) je pak znovu třeba dbát na správný převod textu.

U dokumentů born digital převáděných do PDF může vinou převáděcího nástroje dojít k nesprávnému převodu textu. Může jít o problémy s českými znaky, spojování či multiplikaci slov a vět, nebo prostě o uložení textu jako obrazu. Při prvním použití nového nástroje je nutné zkontrolovat, zda je text převáděn korektně, a případně kontrolu opakovat při dalších změnách (změna výchozího formátu, zásadní změny podoby dokumentů, aktualizace nástroje). Pro kontrolu stačí jednoduše označit a zkopírovat text pdf stránky a vložit ho do poznámkového bloku nebo jiného textového editoru, kde hned uvidíte případné chyby – nezáleží na chybném zalamování řádků apod., pouze na tom, zda jsou korektně převedeny znaky.

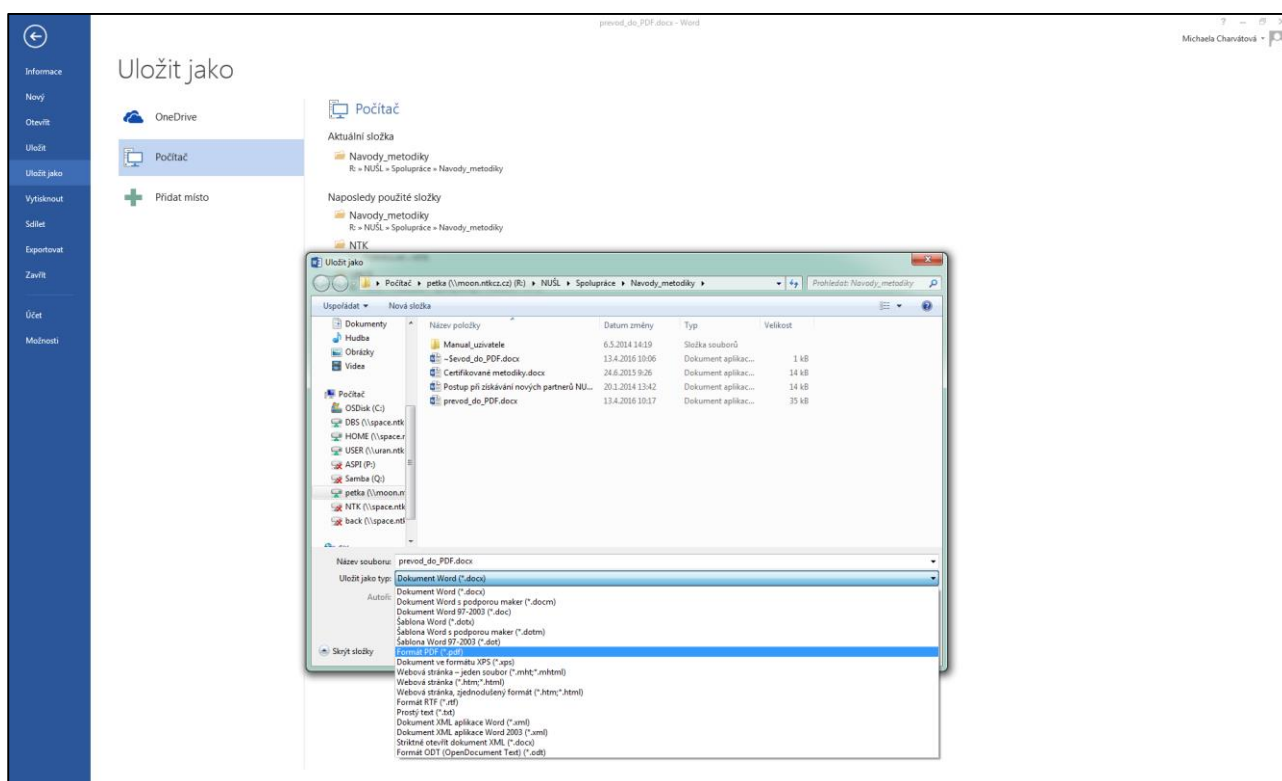
Převod dokumentů do PDF

Pro převod klasických textových dokumentů (mohou obsahovat ilustrace i fotografie) doporučujeme využít možnosti převodu, které poskytují běžně užívané kancelářské balíky. Tyto nástroje nebývají nijak omezeny (velikost souboru, počet stran) a nemívají problémy s prací s českými znaky.

Nástroje kancelářského balíku Microsoft Office

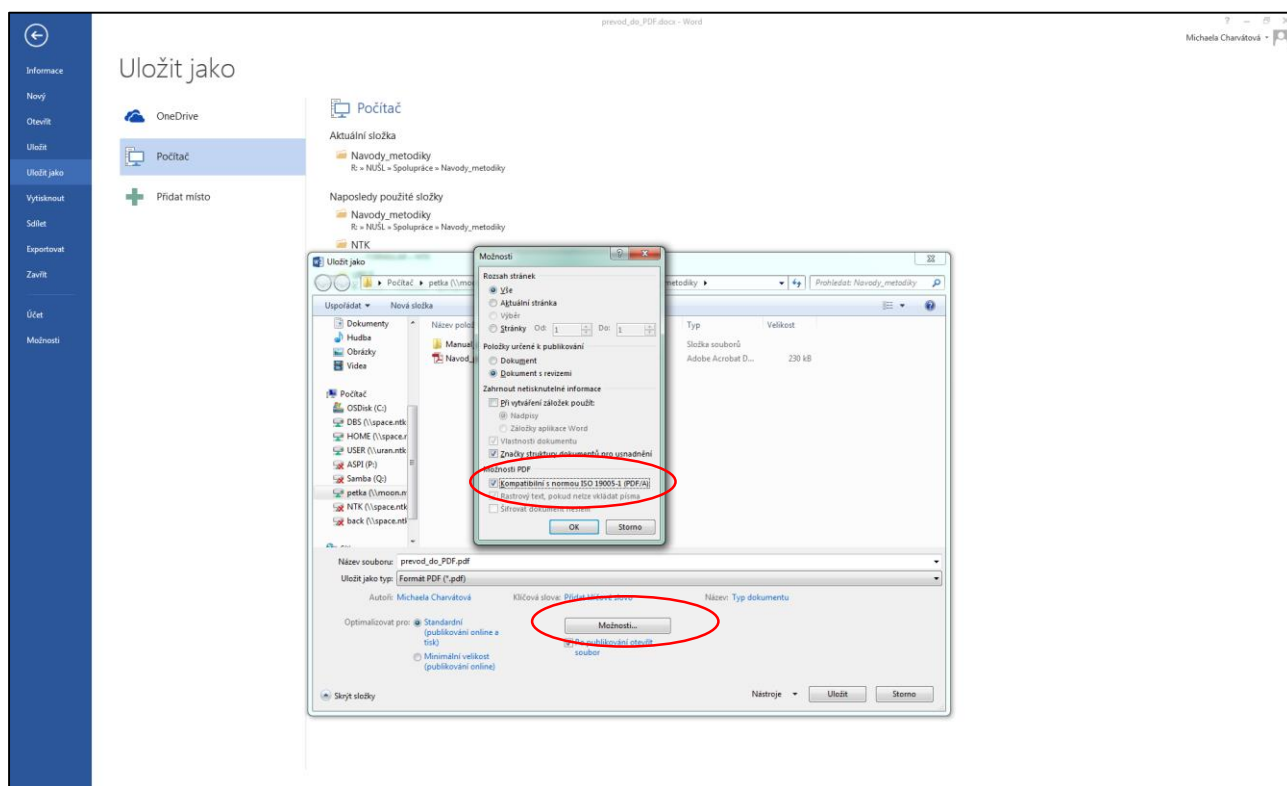
Následující návod pro uložení dokumentů ve formátu PDF byl vytvořen pro aplikace kancelářské sady Microsoft Office 2013. Ukázky jsou provedeny ve Wordu, ale obdobně se postupuje i v Excelu a PowerPointu. Starší verze tohoto balíku nemusí PDF uvádět jako variantu formátu pro uložení, pak je třeba využít jiných nástrojů, třeba tisku do PDF.

Otevřete dokument, který chcete převést do PDF. Vyberte možnost **Soubor** a **Uložit jako**. Po výběru cílového umístění rozbalte roletu nabídky „**Uložit jako typ:**“ a zvolte variantu **Formát PDF (*.pdf)**. Pak už stačí jen potvrdit nastavení tlačítkem **Uložit**.



Obrázek 1: Převod do PDF v MS Office - Uložit jako - rozbalena nabídka formátů

Pro uložení ve variantě formátu PDF/A je nutné před finálním stisknutím tlačítka **Uložit** změnit nastavení v nabídce **Možnosti** (označené tlačítko a rozbalená nabídka na obr. 2). V nastavení je nutné zaškrtnout pole **Kompatibilní s normou ISO 19005-1 (PDF/A)**. Toto nastavení si program zapamatuje a příště již není tento postup potřeba opakovat. Celá akce se dokončí potvrzením pomocí tlačítka **Uložit**.



Obrázek 2: Převod do PDF/A v MS Office - Uložit jako - Možnosti

297 mm

Tisk do PDF

U starších kancelářských balíčků nebyl převod do PDF implementován přímo do ukládací nabídky a používaly se tedy tzv. virtuální tiskárny. Po instalaci a případném nastavení programu takové „tiskárny“ stačilo v kancelářském balíku, ale i v jiných programech zvolit nabídku **tisk** a vybrat příslušnou virtuální tiskárnu. Programů pro virtuální tisk existuje [celá řada](#) a je na uživateli, jaký si vybere. Doporučujeme ovšem skutečně zkontrolovat výstup takového programu a ověřit, že si poradil s převodem textu.

Online konvertory

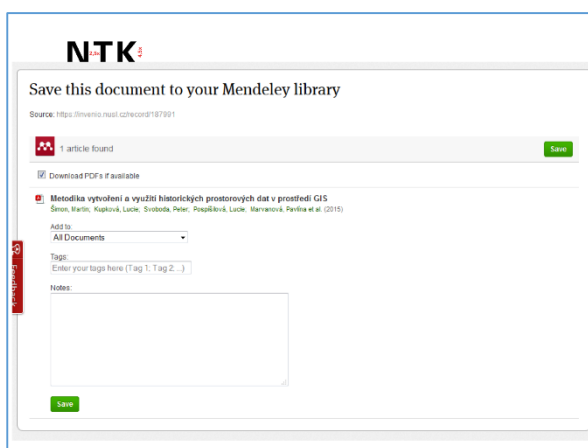
Na internetu lze nalézt nepřeberné množství konvertorů, které převedou do PDF téměř jakýkoli dokument. Vizuálně bývají jejich výsledky v pořádku, mívají ovšem často problém s převodem českého textu. Momentálně nevíme o žádném bezproblémovém, který bychom mohli doporučit pro převod dokumentů určených k uložení do NUŠL. Pokud bychom ale nějaký online konvertor doporučit museli, tak možná [Cloudconvert](#). Tento nástroj sice trpí častým neduhem, že text je po vykopírování vždy několikrát opakován, ale zato dobře pracuje s českými znaky, což je pro účely indexování a prohledávání dokumentů podstatnější.

Specifika převodu prezentací do PDF

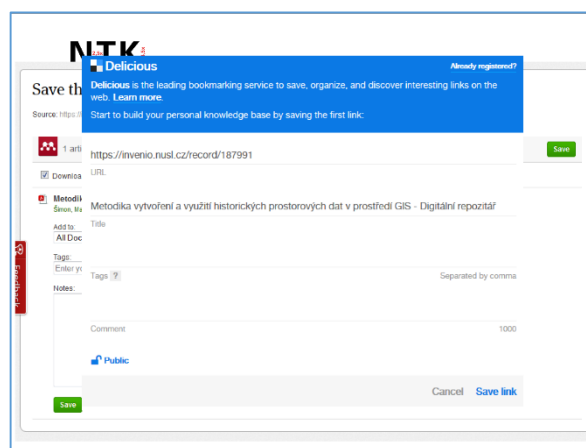
Samotné uložení .ppt prezentace do PDF se neliší od postupu při ukládání jiného typu dokumentu. Rozdíl zde ovšem může být v přípravě dokumentu před samotným uložením, jelikož powerpointové prezentace často obsahují různé efekty a animace, které za určitých okolností mohou způsobit ztrátu informací.

Postupně se překrývající obrázky

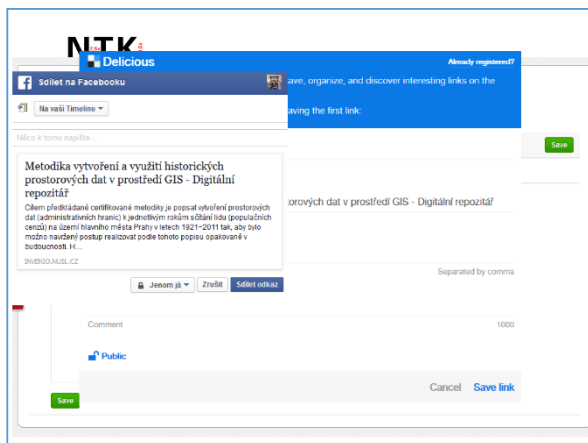
Poměrně často, když chce přednášející ukázat několik obrázků za sebou a zároveň takto nenavýšovat počet snímků prezentace, využije postupného objevování obrázků pomocí animace, kdy jeden obrázek vrší na druhý a ty se postupně překrývají. Pokud tato situace není ošetřena, bude po převodu do PDF na stránce vidět několik takto navršených obrázků a nebude možné je zobrazit jednotlivě. Pokud chceme předat do PDF i jednotlivé obrázky, je nutné tento proces „rozfázovat“ na více snímků prezentace (obr. 3-6). Jednoduše řečeno, nakopírujeme snímek podle počtu obrázků a na každém snímku necháme přesně o ten jeden obrázek více (viz obr. 6).



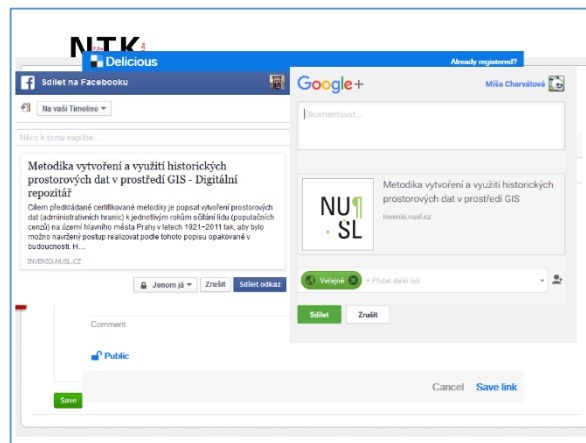
Obrázek 3: Vršení obrázků (1/4)



Obrázek 4: Vršení obrázků (2/4)



Obrázek 5: Vršení obrázků (3/4)



Obrázek 6: Vršení obrázků (4/4)

Údaje skryté za objekty

Dalším případem, kdy animace mohou způsobit ztrátu informace, je využití nějakého objektu ke skrytí např. údajů v prezentaci, kdy je následnou animací objekt odstraněn. To je praxe využívaná například ke skrytí odpovědí na otázky kladené v průběhu prezentace apod. Tuto situaci je třeba řešit obdobně jako v předchozím případě, protože jinak by se odpovědi do PDF nedostaly.