

Výběr vhodných formátů pro dlouhodobé uchování elektronických publikací [1]

Autor:

- [JIROUŠEK, Václav](#) [2], [OSTRÁKOVÁ, Natalie](#) [3]

Číslo:

- [2017, ročník 26, číslo 3](#) [4]

Rubrika:

- [Trendy v knihovnách](#) [5]

Klíčová slova:

- [dlouhodobé uchovávání digitálních informací](#) [6], [e-knihy](#) [7]

Elektronickou knihu neboli e-book či e-knihu definuje terminologická databáze jako knihu v digitální podobě, kterou je možné číst na počítači nebo na dalších zařízeních (těmi mohou být například čtečka elektronických knih, tablet nebo mobilní telefon)¹. Analogicky je tomu u periodik. Elektronické knihy i periodika jsou na trhu dostupné v několika formátech. [Wikipedie](#) [8] jich například popisuje třicet, komunitní stránka [MobileRead Wiki](#) [9] jich eviduje dokonce přes šedesát. Zahrnuty jsou formáty, které vznikly speciálně pro elektronické publikování dokumentů (EPUB, MOBI apod.) i formáty, které vznikly za jiným účelem (PDF, TXT, HTML, DjVu apod.). Formáty elektronických knih vzniklé speciálně pro elektronické publikování jsou vytvořené tak, aby je bylo možné číst na různých zařízeních s různou velikostí obrazovky. Na rozdíl například od formátu PDF tak dokáží přizpůsobit text (velikost písma, zalomení řádků) velikosti obrazovky čtecího zařízení. Pro některé elektronické publikace je však vhodnější pevně daná podoba stránky a nikoliv zalamování a změna velikosti textu; bývá to u dokumentů, které obsahují větší množství obrázků nebo jsou strukturované – například komiksy, noviny, periodika.

Interaktivní a multimediální publikace

Vedle běžných e-knih, které pouze převádějí obsah klasických papírových knih do elektronického prostředí, tvoří samostatnou kapitolu e-knihy interaktivní a multimediální. Multimediální e-knihy mohou vedle běžného textu a obrázků obsahovat také zvuk, video nebo různé interaktivní prvky. Nejčastěji používanými formáty podporujícími multimediální obsah jsou v současné době EPUB 3, KF8, iBooks, interaktivní PDF, případně jsou jednotlivé e-knihy tvořeny samostatnou aplikací pro mobilní zařízení nebo webovou aplikací. Tyto formáty nejsou obvykle vhodné z hlediska dlouhodobého uchování. Společným rizikem pro tento typ dokumentů je obvyklá proprietárnost daných řešení, kdy vytvořené e-knihy jsou velmi závislé na konkrétním zařízení a obsah zpravidla nelze zobrazit na jiných zařízeních ani jej jednoduše převést do otevřených formátů.

Elektronické publikace a jejich uchování v současnosti

Některé zahraniční knihovny – [Britská knihovna](#) [10] (British Library), [Národní knihovna Austrálie](#) [11] (National Library of Australia), [Národní knihovna Finska](#) [12] (National Library of Finland), [Národní knihovna Skotska](#) [13] (National Library of Scotland) a další – již přijímají elektronické knihy

a časopisy k dlouhodobému uložení; elektronické publikace se tak stávají předmětem digitální archivace. Jedním z hlavních témat digitální archivace je výběr formátů vhodných k dlouhodobému uložení digitálního obsahu, protože ne každý je k tomu vhodný a přes veškerou snahu knihoven není možné u rizikových formátů zaručit dostupnost těchto knih do budoucna. V únoru 2017 zveřejnila Britská knihovna svou [analýzu formátů publikací](#) [14]2. V této studii tým Britské knihovny pro digitální archivaci posuzoval jednotlivé formáty z hlediska vhodnosti pro dlouhodobou archivaci. Formáty k posouzení byly zvoleny s ohledem na jejich vlastní sbírky elektronických dokumentů; blíže bylo analyzováno třicet formátů souborů. U jednotlivých formátů byl krátce popsán jejich původ, standardy na kterých jsou založeny, jejich status (dostupnost specifikace, podpora v softwaru apod.), rizika pro dlouhodobou archivaci apod.

Jaké formáty se tedy používají?

Za dnes nejrozšířenější formáty považují autoři studie formáty EPUB, MOBI, PDF a díky popularitě Amazon Kindle i formáty AZW, KF8 a KFX. Pro dlouhodobé uchování elektronických publikací ze zprávy jako nejvhodnější vychází formáty EPUB a PDF, protože jsou dostatečně rozšířené a široce podporované; formát PDF podle této studie podporují dnes nejrozšířenější čtečky (Amazon Kindle, Apple iPad, Barnes & Noble Nook, Kobo a čtečka od firmy Sony), formát EPUB podporují všechny výše zmíněné čtečky kromě čtečky Amazon Kindle.

Zpráva dále zmiňuje i méně běžné a zastaralé formáty, například formát Apabi, který je populární na čínském trhu elektronických knih, nebo proprietární formát BBeb společnosti Sony. Též zde hodnotí formáty, v kterých byly dostupné první elektronické knihy (například v rámci [projektu Gutenberg](#) [15]), tedy formáty TXT a HTML i například DjVu a Microsoft Word (DOC).

Situace ve světě

Předmětem zájmu autorů zprávy bylo také užití formátů v dalších paměťových institucích, zejména pak národních knihovnách s elektronickým „povinným výtiskem“, ať již legislativně ukotveným, nebo probíhajícím na dobrovolné bázi ve spolupráci s vydavateli. Vedle Britské knihovny zpráva zahrnuje Národní knihovnu Francie (Bibliothèque nationale de France), nizozemskou Královskou knihovnu (Koninklijke Bibliotheek) a americkou Kongresovou knihovnu (Library of Congress). Všechny uvedené knihovny využívají ve svých archivech formát PDF a pracují nebo v budoucnu plánují pracovat s formátem EPUB 2. Dvě z knihoven – Britská knihovna a nizozemská Královská knihovna – přijímají dokumenty ve formátu Microsoft Word. Kongresová knihovna spravuje jako součást svých fondů také dokumenty ve formátech založených na HTML/XHTML a XML/TEI. Samozřejmostí je absence ochrany prostřednictvím DRM (digital rights management, správa digitálních práv) u přijímaných dokumentů.

Situace v ČR

Poslanecká sněmovna dosud neschválila potřebnou [novelu legislativy povinného výtisku](#) [16], která by zabezpečila povinné odevzdávání elektronických publikací do Národní knihovny ČR. To však samozřejmě nesmí znamenat, že by Národní knihovna ČR zcela rezignovala na své poslání uchovávat písemné kulturní dědictví, byť by vyšlo pouze v elektronické podobě. Možným řešením tak i nadále zůstává dobrovolné odevzdávání elektronických publikací ve spolupráci s vydavateli. Národní knihovna ČR aktuálně počítá s archivací elektronických knih v otevřených formátech EPUB 2.0.1 a ve formátech PDF/A-1 a PDF/A-2.

Formáty vhodné k dlouhodobé archivaci

Formát PDF/A je archivní podobou známého formátu PDF a je určen přímo pro dlouhodobé uchovávání. Vychází ze specifikace formátů PDF 1.4 (PDF/A-1) a PDF 1.7 (PDF/A-2), ale byly vyřazeny vlastnosti, které jsou z hlediska dlouhodobé archivace problémové (např. vložení JavaScriptu, audia, videa), a jiné vlastnosti se naopak nově staly povinnými (např. vložení všech fontů použitých v dokumentu). Díky tomu v sobě dokument obsahuje všechny informace potřebné ke svému

správnému zobrazení.

Formát EPUB je, jak již bylo výše zmíněno, jeden z nejrozšířenějších formátů elektronických knih. Nejčastěji je zřejmě možné setkat se s verzí formátu 2.0.1 a 3.0.1. Verze 3.0.1 podporuje vkládání zvukových a audiovizuálních dokumentů a JavaScriptu, tento obsah je však mnohem složitější a dražší archivovat. Verze 3.0.1 též umožňuje, aby zvukový či audiovizuální dokument nebyl přímo vložen v dokumentu, ale jen odkazován v textu, přičemž odkaz vede mimo samotný dokument, který tedy neobsahuje vše pro své správné zobrazení³.

Závěr

Jak bylo výše zmíněno, některé země již mají uzákoněn povinný výtisk elektronických publikací, jinde je jejich odevzdávání pouze dobrovolné. Pevně věříme, že Česká republika se brzy přidá k první skupině zemí. Bez ohledu na chybějící legislativu se Národní knihovna ČR na dlouhodobou ochranu a povinný elektronický výtisk již nyní intenzivně připravuje. Cílem musí být uchování současné knižní produkce v co nejuplněnější podobě, tedy bez ohledu na formu nosiče obsahu. Nejlepší předpoklady pro archivaci elektronických publikací mají v dnešní době výše uvedené formáty EPUB a PDF/A. Volba formátů ale není nikdy konečný stav. Je třeba průběžně posuzovat rizika a vhodnost pro dlouhodobou archivaci (podpora v aplikacích, migračních nástrojích apod.), případně seznam vhodných archivačních formátů na základě nových poznatků rozšiřovat.

- ¹ CELBOVÁ, Ludmila a Jaroslava HAVLOVÁ. Elektronická kniha. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna ČR, 2003- [cit. 2017-09-08]. Dostupné z: http://aleph.nkp.cz/F/?func=direct&doc_number=000000866&local_base=KTD [17]
- ² WHIBLEY, Simon, Peter MAY a Paul WHEATLEY. *EBook Summary Assessment* [online]. Version 1.0. British Library, 2017 [cit. 2017-09-08]. Dostupné z: http://wiki.dpconline.org/images/0/04/EBook_Summary_Assessment_v1.0.pdf [14]
- ³ KNIJFF, Johann van der. EPUB for archival preservation. In: *Open Planets Foundation* [online]. 20 July 2012 [cit. 2017-09-08]. Dostupné z: <http://openpreservation.org/system/files/epubForArchivalPreservation20072012ExternalDistribution.pdf> [18]

URL zdroje: <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-3/vyber-vhodnych-formatu-pro-dlouhodobu-uchovani>

Odkazy

- [1] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-3/vyber-vhodnych-formatu-pro-dlouhodobu-uchovani>
- [2] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/autori/jirousek-vaclav>
- [3] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/autori/ostrakova-natalie>
- [4] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-3>
- [5] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/rubriky/trendy-v-knihovnach>
- [6] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/klicova-slova/dlouhodobu-uchovavani-digitalnich-informaci>
- [7] <https://bulletinskip.osvobozena-knihovna.cz/vsechna-cisla/klicova-slova/e-knihy>
- [8] https://en.wikipedia.org/wiki/Comparison_of_e-book_formats
- [9] https://wiki.mobileread.com/wiki/E-book_formats
- [10] <http://www.bl.uk/aboutus/legaldeposit/websites/elecpubs/>
- [11] <https://www.nla.gov.au/legal-deposit-faq>
- [12] <https://www.kansalliskirjasto.fi/en/legal-deposit-office>
- [13] <https://www.nls.uk/about-us/what-we-are/legal-deposit/e-legal-deposit-guide>
- [14] http://wiki.dpconline.org/images/0/04/EBook_Summary_Assessment_v1.0.pdf
- [15] <https://www.gutenberg.org/>

[16] <http://www.psp.cz/sqw/text/tiskt.sqw?O=7&CT=1011&CT1=0>

[17] http://aleph.nkp.cz/F/?func=direct&doc_number=000000866&local_base=KTD

[18] <http://openpreservation.org/system/files/epubForArchivalPreservation20072012ExternalDistribution.pdf>