

Univerzita Karlova  
Filozofická fakulta  
Ústav informačních studií a knihovnictví  
Obor: Informační studia a knihovnictví

## Bakalářská práce



Václav Jiroušek

### **Formáty pro dlouhodobou archivaci elektronických publikací v rámci systému e-deposit v České republice**

File formats for long-term preservation of electronic publications in the e-deposit system in the Czech Republic

Praha 2018

Vedoucí práce: Ing. Martin Souček, Ph.D.

**Poděkování:**

Děkuji Ing. Martinu Součkovi, Ph.D. za pomoc při vedení bakalářské práce. Mé poděkování patří též PhDr. Evě Bratkové, Ph.D. za pomoc s formulací zadání této práce a také všem domácím i zahraničním konzultantům, kteří mi při psaní práce vyšli vstříc a poskytli mi data nebo informace potřebné pro vypracování této práce.

**Prohlášení:**

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 30. července 2018

.....

podpis studenta

## **Identifikační záznam**

Jiroušek, Václav. *Formáty pro dlouhodobou archivaci elektronických publikací v rámci systému e-deposit v České republice*. Praha, 2018, 97 s. Bakalářská práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí bakalářské práce Ing. Martin Souček, Ph.D.

## **Abstrakt (česky)**

Bakalářská práce se zabývá výběrem archivačních formátů pro potřeby dlouhodobého uchování elektronických publikací v rámci českého systému pro e-deposit. Analyzuje současný stav v ČR a na základě vzorku sedmi zahraničních institucí jej porovnává se zahraničními přístupy a zkušenostmi. První část práce se zabývá elektronickými formáty elektronických knih po teoretické stránce. Na základě analýzy trhu jsou určeny formáty relevantní pro český trh s elektronickými publikacemi a jsou popsána rizika jednotlivých formátů s ohledem na možnosti jejich dlouhodobého uchování. Druhá část práce je věnována případovým studiím, ve kterých jsou zdokumentovány formátové politiky zahraničních knihoven, zabývajících se dlouhodobou archivací elektronických publikací. Je také provedena analýza dat elektronických publikací, dobrovolně odevzdaných vydavateli do NK ČR v rámci pilotního provozu systému pro e-deposit. Závěr práce shrnuje zjištěné poznatky formou doporučení archivačních formátů pro další rozvoj českého systému pro e-deposit.

## **Klíčová slova (česky)**

povinný výtisk, e-knihy, elektronické publikace, dlouhodobá ochrana digitálních dat, digitální informace, digitální kurátorství, standardizace, DRM, souborové formáty

**Abstract (in English):**

This bachelor's thesis focuses on selection of archival formats for the purposes of long term preservation of electronic publications within the Czech e-deposit system. It analyzes the current situation in the Czech Republic and, on the basis of a sample of seven foreign institutions, compares it with foreign approaches and experiences. The first part of the thesis deals with the electronic formats of electronic books on the theoretical side. Based on the market analysis, the formats relevant to the Czech market for electronic publications are identified and the risks of the individual formats are described with regard to the possibilities of their long-term preservation. The second part is devoted to case studies in which the format policies of foreign libraries dealing with long-term archiving of electronic publications are documented. An analysis of electronic publications, voluntarily deposited by the publisher to the NL CR during the pilot operation of the e-deposit system, is also performed. The conclusion of the thesis summarizes the findings in the form of recommendations for archival formats for the further development of the Czech e-deposit system.

**Klíčová slova (anglicky):**

legal deposit, e-books, electronic publications, long-term preservation, digital information, digital curation, standardization, DRM, file formats

## Obsah

Seznam zkratk	8
Předmluva	10
1 Úvod	13
1.1 Vymezení terminologie	14
2 Formáty e-publikací	16
2.1 Formáty e-publikací	16
2.2 Typologie formátů	16
2.3 Multimediální a interaktivní publikace	17
2.4 Formáty e-publikací používané na českém trhu	18
2.5 Prodejci elektronických knih v ČR	20
2.6 Dílčí závěry k 2. kapitole	21
3 Historie a aktuální stav řešení problematiky e-depositu v ČR	22
3.1 Projekt eDpilot	22
3.2 Projekt NAKI – Správa elektronických publikací v síti knihoven ČR	22
3.3 Stav legislativy v ČR	23
4 Výběr archivačních formátů	25
4.1 Dlouhodobá ochrana digitálních dokumentů	25
4.2 Kritéria výběru archivačních formátů	25
4.3 Formátové profily	27
4.4 Signifikantní vlastnosti	27
5 Správa digitálních práv (DRM)	29
5.1 Uzavřený ekosystém	29
5.2 Digitální vodoznak	29
5.3 Šifrování obsahu	30
5.4 DRM schémata	30
5.5 Nejdůležitější DRM schémata pro e-publikace	31
5.6 Zhodnocení dopadů užití DRM pro dlouhodobou archivaci	31
6 Zhodnocení vybraných formátů e-publikací	34
6.1 Open eBook (OEBPS)	34
6.2 Mobipocket (MOBI)	36
6.3 EPUB 2	39
6.4 EPUB 3	43
6.5 iBooks	46
6.6 Wooky	48

6.7	PDF .....	50
6.8	Dílčí závěry k 6. kapitole .....	56
7	Případové studie vybraných zahraničních knihoven .....	57
7.1	Cíle a metoda.....	57
7.2	Tematicky blízké výzkumy .....	57
7.3	Deutsche Nationalbibliothek .....	58
7.4	Bibliothèque nationale de France .....	61
7.5	The National Library of Australia.....	63
7.6	Narodna in univerzitetna knjižnica (Slovinsko).....	64
7.7	Koninklijke Bibliotheek (Nizozemí).....	65
7.8	The National Library of New Zealand .....	67
7.9	The British Library.....	68
7.10	Dílčí závěry k 7. kapitole .....	70
8	Analýza e-publikací, dobrovolně odevzdaných do NK ČR.....	72
8.1	Použité nástroje .....	72
8.2	Struktura formátů přijatých dokumentů .....	73
8.3	Soubory ve formátu EPUB.....	73
8.4	Soubory ve formátu PDF .....	74
8.5	Soubory ve formátu MOBI .....	75
8.6	Ostatní souborové formáty .....	75
8.7	Dílčí závěry k 8. kapitole .....	76
9	Závěr .....	77
	Bibliografie .....	79
	Seznam počítačových programů.....	94
	Použité počítačové programy.....	94
	Ostatní počítačové programy zmiňované v textu .....	94
	Seznam tabulek .....	95
	Seznam obrázků.....	96
	Seznam příloh .....	97

## Seznam zkratk

Níže jsou uvedeny v textu často používané zkratky. Zkratky, které jsou uvedené v textu pouze jednou a jsou v něm také rozepsány, nejsou v tomto seznamu uvedeny.

ADEPT	Adobe Digital Experience Protection Technology
AAC LC	Advanced Audio Coding – Low Complexity
AZW	Kindle Format 7
BnF	Bibliothèque nationale de France – Francouzská národní knihovna
CSS	Cascading Style Sheets
ČR	Česká republika
DNB	Deutsche Nationalbibliothek – Německá národní knihovna
DOC	DOCument file
DPC	Digital Preservation Coalition
DRM	Digital Right Management – Správa digitálních práv
DROID	Digital Record Object Identification
DTBook	DAISY Digital Talking Book
EPUB	Electronic PUBlication
EXE	Executable file
FIDO	Format Identification for Digital Objects
FITS	The File Information Tool Set
FTP	File Transfer Protocol
GIF	Graphics Interchange Format
HTM	viz HTML
HTML	HyperText Markup Language
IDPF	International Digital Publishing Forum
InSPECT	Investigating the Significant Properties of Electronic Content Over Time – projekt
ISBN	International Standard Book Number
ISMN	International Standard Music Number
ISO	International Organization of Standardization
ISSN	International Standard Serial Number
JHOVE	JSTOR/Harvard Object Validation Environment
JPEG	Joint Photographic Expert Group
JPG	viz JPEG
KB	Koninklijke Bibliotheek – Královská knihovna Nizozemí
KF8	Kindle Format 8
KFX	Kindle Format X
LIT	Microsoft LIT format
LTP	Long-term preservation
MK ČR	Ministerstvo kultury ČR
MIME	Multipurpose Internet Mail Extensions



MOBI	Mobipocket
MP3	Music file (MPEG Layer 3)
NAKI	Program aplikovaného výzkumu a vývoje národní a kulturní identity
NDK	Národní digitální knihovna – projekt
NK ČR	Národní knihovna České republiky
NLA	National Library of Australia – Národní knihovna Austrálie
NLNZ	The National Library of New Zealand – Národní knihovna Nového Zélandu
NUK	Narodna in univerzitetna knjižnica – Národní a univerzitní knihovna
ODLIS	Online Dictionary for Library and Information Science
ODT	OpenDocument Text
OEBPS	Open eBook Publication Structure
PDA	Personal Digital Assistant
PDB	Palm Database Format
PDF	Portable Document Format
PNG	Portable Network Graphics
PRC	Palm Resource Compiler
PSP ČR	Poslanecká sněmovna Parlamentu České republiky
PUID	PRONOM Persistent Unique Identifier
SCAPE	Scalable Preservation Environments – projekt
SČKN	Svaz českých knihkupců a nakladatelů
SNE	Syndicat National de l'Édition – sdružení francouzských nakladatelů
SPAR	Scalable Preservation and Archiving Repository
TDKIV	Terminologická databáze knihovnictví a informační vědy
TI/A	Tagged Image for Archival
TIF	Viz TIFF
TIFF	Tagged Image File Format
TPZ	Topaz format book
TXT	text file
W3C	World Wide Web Consortium
WYSIWYG	What you see is what you get
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language
ZIP	ZIP File Format

## Předmluva

Téma mé bakalářské práce je *Formáty pro dlouhodobou archivaci elektronických publikací v rámci systému e-deposit v České republice*. Cílem práce je zejména analyzovat současný stav řešení otázky výběru archivních formátů pro potřeby dlouhodobého uchovávání elektronických publikací v rámci systému e-deposit v ČR, a to také v porovnání se zahraničím. Téma práce jsem zvolil s ohledem na své několikaleté profesní působení v Odboru digitálních fondů Národní knihovny ČR, kde se věnuji správě dlouhodobého datového úložiště.

V roce 2012 jsem se úvodními formátovými analýzami podílel také na realizaci projektu NAKI *Správa elektronických publikací v síti knihoven ČR* (DF12P01OVV006). V roce 2015 jsem spolupracoval také na závěrečném testování systému pro automatický příjem elektronických síťových publikací do NK ČR – tzv. systém e-deposit. Přestože potřebné novely zákonů k zajištění povinného odevzdávání síťových publikací nebyly doposud schváleny, a systém e-deposit nadále zůstává ve fázi interního testování NK ČR, technologický vývoj ve světě elektronického publikování se nezastavil. Po pěti letech jsem se proto rozhodl k tématu vrátit a nově jej zpracovat na základě současných poznatků, aktuálně užívaných formátů a také čerstvých zkušeností zahraničních národních knihoven s právem povinného výtisku pro síťové publikace, které za uplynulé roky tyto instituce nasbíraly. Z výstupů projektu NAKI jsem vycházel pouze v omezené míře, rozsah zachované dokumentace se ukázal být pro potřeby dosažení cílů této práce nedostačující a v mnoha ohledech již neaktuální. Zhodnocení jednotlivých formátů jsem tedy provedl zejména na základě aktuálních formátových studií od zahraničních institucí.

Zadání práce mi předepisuje mimo jiné provést *Analýzu výběru archivních formátů ve vybraných zahraničních institucích zabývajících se dlouhodobou archivací elektronických publikací*. *Realizován bude malý průzkum 5-6 zahraničních pracovišť formou dotazníku. Další relevantní informace budou čerpány z vybrané literatury k tématu*. Vzhledem k tomu, že úvodní rešerše poukázala na poměrně velké rozdíly v přístupech jednotlivých institucí, daných například rozdíly v národních legislativách, rozhodl jsem se po domluvě s vedoucím práce od dotazníkového šetření upustit a za využití metody kvalitativního výzkumu zadaný úkol zpracovat formou případových studií. Zvolená metoda umožnila lépe zachytit specifika různých přístupů u zkoumaných subjektů. Při zpracování případových studií jsem potom pracoval v první fázi s dostupnou odbornou literaturou a pokyny nakladatelům zveřejněnými na webových stránkách institucí. V druhé fázi jsem daná pracoviště kontaktoval e-mailem a snažil se doplnit informace, které v dostupných zdrojích chyběly. Jako užitečné se ukázalo také ověřování publikovaných informací, které nebyly vždy aktuální či přesné.

Konzultantem za nizozemskou Královskou knihovnu byl formátový specialista Johan van der Knijff, za australskou Národní knihovnu mi konzultace poskytl Libor Coufal, který zde vede oddělení dlouhodobého digitálního uchovávání. Za slovinskou Národní a univerzitní knihovnu mi konzultace poskytl vedoucí oddělení správy digitální knihovny Janko Klasinc, za Německou národní knihovnu Cornelia Diebel z oddělení informační infrastruktury a uchovávání, za Britskou národní knihovnu Maureen Pennock, vedoucí oddělení digitální ochrany a za Národní knihovnu Nového Zélandu Amy Joseph, spravující agendu elektronického povinného výtisku.

Předkládaná práce je logicky členěna do dvou hlavních celků. Zatímco v kapitolách 2–5 jsem se zaměřil na vymezení teoretického rámce práce, v kapitolách 6–8 se zabývám zhodnocením konkrétních formátů, případovými studii a analýzou vzorku dat odevzdaných e-publikací.

V kapitole Formáty e-publikací se zabývám formáty e-publikací nejprve v obecné rovině, v druhé části potom zejména s cílem určit relevantní souborové formáty, důležité pro český trh. Na základě analýzy báze ohlášených e-knih agentury ISBN, doplněné o průzkum distribučních formátů u největších prodejců e-knih, zde definuji formáty, kterým se později podrobněji věnuji zejména v kapitolách 7 a 8. Následuje kapitola shrnující stav řešení problematiky uchovávání e-publikací v ČR z hlediska aktivit NK ČR a také aktuální stav legislativy. Další kapitola popisuje teoretické parametry výběru vhodných archivačních formátů. V kapitole páté se potom věnuji problematice DRM a jejich dopadů pro dlouhodobé uchovávání dokumentů. Doplnění kapitoly pokládám za důležité zejména z důvodu definice terminologie pro různé kategorie DRM a dopadů jednotlivých opatření na dlouhodobé uchování.

Druhou část práce uvozuje část věnovaná formátovým studiím. Zde se podrobně věnuji jednotlivým formátům, které byly identifikovány jako významné pro české prostředí. Formáty jsou na základě dostupné literatury popsány, včetně stručné historie, důležité pro pochopení významu a vztahu mezi jednotlivými formáty, a následně také zhodnoceny z hlediska své vhodnosti pro dlouhodobé uchování. U formátů s vyšším stupněm rizika je posuzována také možnost jejich převodu do jiných formátů. Zjištěná rizika jsou poté poměřována s postupy zahraničních institucí v následující kapitole věnované případovým studiím, zaměřeným na formátové politiky vybraných zahraničních knihoven. Cílem případových studií je zejména ověření konkrétních způsobů implementace řízení formátových rizik u institucí s praktickými zkušenostmi s provozem systémů pro e-deposit. V závěrečné kapitole se věnuji analýze dat dobrovolně odevzdaných nakladateli do NK ČR. V této části je provedena nejprve identifikace formátů odevzdaných publikací, a následně také jejich validace za využití nástrojů, určených v rámci kapitoly Zhodnocení vybraných formátů e-publikací. Podrobné výstupy z analýzy dat

odevzdaných e-publikací jsou součástí elektronické verze bakalářské práce jako Příloha 1. Příloha je ve formě tabulky ve formátu CSV (kódování UTF-8). Data v příloze jsou anonymizována tak, aby nebylo možné identifikovat konkrétního nakladatele.

Bakalářská práce má celkem 97 stran (154 000 znaků vlastního textu, tj. předmluva až závěr včetně, s mezerami a poznámkami pod čarou, bez příloh a seznamu zdrojů – po přepočtu tedy 85 normostran textu). Práce byla vytvořena v souladu s vnitřními pravidly pro evidenci, odevzdávání, obhajování a zveřejňování závěrečných prací Filozofické fakulty Univerzity Karlovy. Použité zdroje byly v práci citovány v souladu s normou ČSN ISO 690:2011, pro citace v textu je použita metoda číselného odkazu (údaj o stránce následuje vždy po čísle odkazujícím na položku v seznamu použitých zdrojů). Seznam použitých zdrojů na konci práce je řazen podle pořadí použití zdrojů v práci. Seznam literatury je doplněn seznamem použitých a v textu zmiňovaných počítačových programů.

## 1 Úvod

Počátek povinnosti vydavatelů odevzdávat exempláře publikací, vydaných na území daných států, nejprve do královských a později do národních knihoven, sahá v Evropě až do raného novověku. Ačkoliv na počátku povinného výtisku nebyly vždy ušlechtilé cíle (sloužil například pro potřeby cenzury), postupem času se jeho primárním účelem stala, vedle bibliografické funkce, zejména snaha zachytit a uchovat národní písemné kulturní dědictví pro další generace.

Uchovávání písemného dědictví vydaného ve formě tištěných dokumentů vyžaduje určité úsilí. Zejména dokumenty tištěné na degradaci ohroženém kyselém papíru jsou proto v posledních dvou desetiletích předmětem záchranného reformátování do digitální podoby. Dlouhodobá ochrana těchto digitálních dat klade na knihovny zcela nové nároky. Přinejmenším v omezené míře a v krátkodobé perspektivě lze nicméně alespoň teoreticky v případě ztráty čitelnosti jednotlivých reformátovaných kopií uvažovat o vytvoření digitálních kopií nových.

Masivní nárůst produkce původně digitálních (born-digital) dat ve formě e-knih a e-časopisů v posledním desetiletí tuto záchrannou síť trhá a staví tím paměťové instituce po celém světě před novou výzvu. Bez zachycení a uchování těchto původně digitálních dokumentů, často již publikovaných výhradně v elektronické podobě, nebude v budoucnu možné považovat zaznamenaný obraz současné literární produkce za reprezentativní. Pro zabezpečení sběru elektronických publikací je proto potřeba řešit zejména otázky legislativní a otázky technické.

Česká republika bohužel v tuto chvíli nemá legislativu, která by povinné odevzdávání elektronických publikací nakladatelům předepisovala. Absence potřebné právní úpravy nicméně neznamená, že nelze konat. Příkladem jsou projekty NK ČR pro zabezpečení alespoň dobrovolného odevzdávání elektronických publikací. Jedním z klíčových úkolů pro technické zabezpečení dlouhodobé ochrany digitálních dokumentů jsou dobře zvolené datové formáty. Chybně zvolená formátová politika může vést na jedné straně k tomu, že přijaté publikace nebude možné dříve či později zobrazit, na straně druhé může potom vést k tomu, že podmínky pro uložení zcela vyloučí příliš velkou část produkce, vydanou v nechtěných formátech.

Sledování a vstřebávání zahraničních zkušeností s provozem systémů pro e-deposit může být jedním ze způsobů, jak zmírnit dopady chybějící legislativy v ČR. Bude-li NK ČR stavět na zkušenostech zahraničních časných osvojitelů legislativy k povinnému výtisku e-publikací, může se snáze vyvarovat již odhalených chyb a čekat na pomyslný startovní výstřel, tedy schválení potřebné legislativy, lépe teoreticky připravena.

## 1.1 Vymezení terminologie

### 1.1.1 E-deposit

Termín *e-deposit* (česky též *e-depozit*) lze, v závislosti na kontextu užití, vykládat několika způsoby. V původním významu se jedná o anglickou zkratku označující *electronic legal deposit*, případně též *e-publication legal deposit*. Do češtiny tedy lze termín jednoduše překládat jako *elektronický povinný výtisk* nebo *povinný depozit elektronických publikací*<sup>1</sup>. Lze se nicméně setkat také s užším výkladem, kdy již samotný termín *e-deposit* určuje také způsob doručení e-publikací. V takovém případě se jedná o aktivní odevzdávání publikací nakladatelem do knihovny, zpravidla prostřednictvím dedikovaného portálu, provozovaného depozitní knihovnou. Tento způsob odevzdávání lze vymežit zejména proti sběru publikací z webových stránek metodou web-harvestingu, kdy je role nakladatele převážně pasivní [1, s. 6]. V posledním případě se termínu *e-deposit* přeneseně využívá také pro označení automatizovaných systémů nebo portálů pro příjem elektronických publikací. V této práci budu užívat termín ve všech těchto významech, konkrétní význam bude dán kontextem.

### 1.1.2 Elektronická publikace

Americký výkladový *Online slovník pro knihovníky a informační specialisty (Online Dictionary for Library and Information Science, ODLIS)* definuje termín *elektronická publikace (electronic publication)* takto: „*A work in digital form capable of being read or otherwise perceived, distributed to the general public electronically. The category includes electronic journals and e-prints, electronic magazines and newspapers, electronic books, Web sites, Weblogs, etc. Some electronic publications are online versions of print publications; others are born digital. Synonymous with e-publication*“ [2].

Podobně NK ČR v monografii *Elektronické publikace v Národní knihovně ČR* [3, s. 14-18] definuje pět základních typů elektronických publikací: elektronické knihy (e-knihy), tiskové předlohy, elektronická periodika (e-časopisy), obsah webu, elektronické publikace na hmotných médiích. Termín elektronická publikace je tedy pro potřeby této práce, zaměřené na problematiku sběru prostřednictvím systému e-deposit, příliš široký. V práci jej proto budu užívat pouze pro takové e-publikace, které jsou ze své povahy určeny k odevzdávání prostřednictvím e-depositu. Vyloučen bude zejména obsah webu, jehož sklizení a uchování zajišťuje Webarchiv NK ČR [4] a vyloučeny budou také elektronické publikace odevzdávané

---

<sup>1</sup> Tento překlad navrhuje Zdeněk Matušík jako přesnější [192].

na hmotných médiích, jež jsou v současné době doručovány a uchovávány společně s fyzickými exempláři [3, s. 16-18]. Vzhledem ke stavu legislativy a podobě současných legislativních návrhů (viz kapitola 3.3) se v práci nebudu specificky zaměřovat ani na tiskové předlohy a elektronická periodika. Tyto dvě kategorie dokumentů se nicméně ve velké míře, zejména pokud se výběru vhodných archivačních formátů týče, s e-knhami překrývají.

Pro účely této bakalářské práce budu tedy termín elektronická publikace (e-publikace) vykládat jako ekvivalent tištěné formy periodického (e-periodikum) nebo monografického dokumentu (e-kniha), která je digitální původem (born-digital) a je šířena v nehmotné podobě.

## 2 Formáty e-publikací

### 2.1 Formáty e-publikací

Elektronické publikace mohou nabývat různých podob, disponovat různými vlastnostmi a vyskytují se v celé řadě různých formátů. Například Wikipedia uvádí ve svém často citovaném výčtu formátů přes 30 různých formátů e-knih [5]. Podobně rozsáhlý seznam zveřejnila Britská národní knihovna ve svém přehledu formátů e-knih z roku 2017 [6]. Jiný přehled na stránce MobileRead Wiki popisuje dokonce více než 60 různých formátů [7].

Britská národní knihovna uvádí jako hlavní formáty EPUB, MOBI, AZW, KF8, iBooks a PDF. MobileRead Wiki oproti tomu jako hlavní formáty zmiňuje AZW, AZW1, AZW3, AZW4, EPUB, KF8, KFX, MOBI, PDB, PDF, PRC a TPZ<sup>2</sup>. Mezi dalšími formáty uvádí například proprietární formáty výrobců hardwarových čteček (např. LIT od Microsoftu nebo BbeB od Sony), běžné rozšířené formáty primárně sloužící k jinému účelu (DOC, TXT, EXE, HTM, DJVU aj.), nebo zastaralé formáty, v současné době již nahrazené modernějšími (OEB, nyní nahrazen formátem EPUB).<sup>3</sup>

### 2.2 Typologie formátů

Jednotlivé souborové formáty mají různé vlastnosti, podle kterých je můžeme dělit. Distribuční formát e-publikací bývá zpravidla zvolen na základě kompromisu mezi vlastnostmi, které jsou od e-knihy požadované, a rozsahem hardwarových a softwarových platform, na kterých má být dokument zobrazitelný. Základní rozdělení formátů e-knih můžeme učinit dle schopnosti přizpůsobování textu velikosti displeje zařízení nebo preferencím konkrétního uživatele – např. zalamováním textu, možností přizpůsobování velikosti písma nebo výběru fontu<sup>4</sup>.

Dokumenty s pevně daným obsahem, ale přizpůsobitelnou formou zobrazení můžeme označit jako dokumenty nedefinitivní [8, s. 65]. Formáty, které toto chování podporují, jsou zpravidla založeny na technologiích XML a CSS a historicky vznikly pro čtení na mobilních zařízeních typu PDA, dnes na chytrých telefonech, čtečkách elektronických knih nebo tabletech. Nejvýznamnějším zástupcem této kategorie formátů je v současné době formát EPUB.

Druhou skupinou jsou formáty určené pro tzv. definitivní dokumenty, které zachovávají stejný vzhled nezávisle na zobrazovacím zařízení. Hegemony této skupiny formátů jsou dlouhodobě

---

<sup>2</sup> MobileRead Wiki velmi detailně rozlišuje formáty z rodiny Kindle – ve výčtu 12 nejdůležitějších formátů jich uvádí hned 9: AZW, AZW1, AZW3, AZW4, KF8, KFX, MOBI, PRC, TPZ.

<sup>3</sup> Srov. s [191].

<sup>4</sup> Toto rozdělení využívá ve svém doporučení archivačních formátů např. Kongresová knihovna [48, s. 6].



zejména formáty z rodiny PDF. PDF je využíváno pro svou univerzálnost napříč platformami různých typů a výrobců, a pro větší kontrolu nakladatele nad výslednou podobou díla napříč různými zařízeními. Časté je užití formátu PDF u děl se složitější grafickou úpravou nebo u odborných publikací. Je třeba nicméně poznamenat, že rozdíly mezi oběma skupinami formátů se v posledních letech stírají. Formáty původně určené zejména pro malá mobilní zařízení již v aktuálních verzích (např. EPUB 3) zpravidla podporují fixní sazbu textu a umožňují tak i velmi složité formátování obsahu děl. Naopak čtecí aplikace jsou již schopné poměrně úspěšně zalamovat text tagovaných PDF i pro čtení na malých displejích.

Dalšími vlastnostmi, podle kterých je možné formáty e-knih členit, jsou například podpora vložených obrázků, zvuku, videa, 3D objektů či podpora vložených interaktivních prvků. Pro odborné texty z oblasti přírodních věd může být velmi důležitá například podpora matematického značkovacího jazyka MathML [9]. Pro nakladatele a distributory může být významná podpora zabezpečení DRM, pro zákazníka naopak může být důležitá podpora formátů na různých zařízeních. Ta je zpravidla spojená s otevřeností dokumentace formátu.

### **2.3 Multimediální a interaktivní publikace**

Svébytnou kategorií e-publikací jsou multimediální a interaktivní publikace. Nejčastěji se s nimi lze setkat v podobě e-knih pro děti nebo učebních textů, potom se označují také jako i-učebnice [10]. Tyto dokumenty mohou obsahovat plně interaktivní prvky, vedle běžného multimediálního obsahu, jakým je audio, video či obrázková galerie, tak mohou obsahovat například 3D objekty nebo nejrůznější prvky určené k procvičování a testování učiva. [11] [10]

Za vhodné technologie pro vytváření interaktivních e-publikací se díky jejich univerzálnosti v současnosti považuje užití HTML 5 či EPUB 3<sup>5</sup> v kombinaci s JavaScriptem. Za zastaralé a spíše nevhodné se zejména z důvodu nekompatibility s některými zařízeními naopak již řadu let považují dříve rozšířené technologie Adobe Flash<sup>6</sup> nebo Java [12, s. 2-3]. Technologie Adobe Flash byla v minulosti klíčová zejména pro interaktivní prvky používané v rámci formátu PDF [13] [14]. Ve výčtu by neměly chybět ani samostatně vyvinuté aplikace pro iOS či Android, kdy jednotlivé dokumenty jsou tvořeny samostatnou aplikací [15, s. 2-4]. [16]

Příklady platform pro publikování interaktivních publikací z českého prostředí jsou například Publi.cz využívající technologii HTML 5 a JavaScriptu nebo Flexibooks.cz s vlastním

---

<sup>5</sup> Stejně tak se často využívají deriváty EPUB 3 – iBooks a KF8.

<sup>6</sup> Společnost Adobe oznámila v červenci 2017 ukončení podpory aplikace Flash Player. Uživatelům doporučuje převod obsahu na nové otevřené standardy HTML5, WebGL a WebAssembly. [201]

vyvinutým proprietárním formátem. Příkladem multimediálních e-knih je produkce nakladatelství Carpe Diem, které využívá formát EPUB 3 pro vydávání publikací obsahujících záznamy koncertů, divadelních představení nebo dokumentárních filmů [17].

## **2.4 Formáty e-publikací používané na českém trhu**

Zdaleka ne všechny formáty e-publikací, které jsou historicky na světě známy, jsou relevantní pro aktuální produkci e-knih na českém trhu. Prvním úkolem této práce tedy bylo vymezit množinu formátů, které jsou relevantní pro komerční produkci, vycházející v současné době v ČR. Jako hlavní strategii pro získání přehledu o výskytu a zastoupení jednotlivých formátů e-knih v ČR jsem zvolil analýzu přidělených čísel ISBN, doplněnou o základní průzkum distribuční sítě prodejců e-knih s ohledem na formáty, využívané největšími prodejci.

Identifikátor ISBN je online publikacím v ČR přidělován od roku 2010. Podmínkou přidělení je, že se jedná o ukončené dílo, které je určeno pro širokou veřejnost a je zpoplatněno [18]. Třetí podmínka nicméně není vždy dodržována a ISBN tak mohou dostat i publikace šířené zdarma<sup>7</sup>. Dle pravidel pro přidělování ISBN [19, s. 43] by mělo být přiděleno samostatné číslo ISBN každému jednotlivému formátu, ve kterém je online publikace vydána, a dokonce každé verzi – např. verze s DRM a bez DRM, verze s obrázky a bez obrázků apod. [19, s. 74]. V praxi se však různé způsoby DRM spíše nerozlišují a pomocí ISBN je možné rozlišit tedy jen jednotlivé datové formáty. Ani to se neděje příliš důsledně, lze se tak setkat s nakladateli, kteří přidělí online publikaci pouze jedno ISBN, pod kterým je následně publikace distribuována v několika různých formátech. Podobně je běžnou praxí, že nakladatel přidělí číslo ISBN např. produkčnímu formátu EPUB, a následně je e-kniha prostřednictvím distribučního kanálu nabízena ve formátu jiném, např. iBooks. Ohlášení e-knihy se děje prostřednictvím online formuláře<sup>8</sup>, kde nakladatel v případě online publikace ručně vyplňuje formát. Je tedy potřeba počítat také s tím, že nakladatel nemusí vždy označit formát zcela přesně<sup>9</sup>.

První analýza báze ohlášených online publikací proběhla v září 2012 v rámci *projektu NAKI*<sup>10</sup>. K uvedenému měsíci bylo v bázi evidováno 4711 čísel ISBN přidělených online publikacím od 205 nakladatelů. Strukturu nahlášených formátů ukazuje tabulka 1 [20, s. 7].

---

<sup>7</sup> Příkladem je Městská knihovna v Praze se svou edicí E-knihovna a více než 2000 přidělenými ISBN.

<sup>8</sup> Formulář je dostupný z adresy: [https://aleph.nkp.cz/F/?func=file&file\\_name=service-isbn](https://aleph.nkp.cz/F/?func=file&file_name=service-isbn)

<sup>9</sup> Zřejmé překlepy (Epub vs Epub apod.) či záměny formátů jsem při vyhodnocení dat korigoval.

<sup>10</sup> Viz kapitola 3.2 Projekt NAKI – Správa elektronických publikací v síti knihoven ČR.

formát	PDF	EPUB	MOBI	ostatní formáty (HTML, iBooks, Wooky aj.)
%	55 %	33 %	11 %	1 %

Tabulka 1. Souborové formáty e-knih dle báze ohlášených e-knih (2010 až září 2012)

Analýzu dat z databáze ISBN jsem pro potřeby této práce zopakoval k 30. červnu 2018. K tomuto datu již bylo v bázi evidováno 65 195 čísel ISBN přidělených online publikacím od 1370 nakladatelů. Struktura nahlášených formátů byla následující:

formát	EPUB	PDF	MOBI	HTML	PRC	Wooky	iBooks	Android <sup>11</sup>	Flexibooks	ostatní
počet	28943	21420	12504	1113	548	216	203	67	42	103
%	44,4 %	32,9 %	19,2 %	1,7 %	0,8 %	0,3 %	0,3 %	0,1 %	0,1 %	0,2 %

Tabulka 2. Souborové formáty e-knih dle báze ohlášených e-knih (2010 až červen 2018)

Tabulka 2 ukazuje strukturu ohlášených formátů e-knih od roku 2010 do 30. června 2018 a potvrzuje dominantní postavení formátů EPUB, PDF a MOBI. Pro zajištění skutečně aktuálních hodnot jsem samostatně zpracoval ještě poslední ukončený rok, tedy 2017. Pro lepší představu o významu jednotlivých formátů jsem mezi sledované ukazatele zahrnul vedle absolutního počtu ohlášených e-knih v daném formátu a jejich procentuálního zastoupení také počet nakladatelů, kteří daný formát využívají<sup>12</sup>. V roce 2017 bylo přiděleno 12 710 čísel ISBN e-publikacím od 377 nakladatelů. Počet titulů, dle počtu jedinečných hodnot v poli 245, odpovídá přibližně 8900. Je tedy zřejmé, že významná část titulů je ohlášena ve více formátech, netýká se to ale absolutní většiny titulů. Strukturu formátů za rok 2017 ukazuje tabulka 3.

formát	EPUB	PDF	MOBI	HTML	PRC	CoverPage <sup>13</sup>	Wooky	Flexibooks	iBooks
počet	7062	3418	1595	317	269	9	8	6	1
%	55,7 %	26,9 %	12,6 %	2,5 %	2,1 %	0,1 %	0,1 %	0,0 %	0,0 %
nakladatelů	180	322	122	8	1	1	1	3	1

Tabulka 3. Souborové formáty e-knih dle báze ohlášených e-knih (rok 2017)

Přehled, který ukazuje tabulka 3, poskytuje určitou představu o zastoupení jednotlivých formátů na českém trhu, ve skutečnosti ale nemá absolutní výpovědní hodnotu ohledně důležitosti těchto formátů pro dlouhodobé uchování. V případě, kdy by například každá e-publikace vydaná v MOBI měla ekvivalent v paralelně vydaném EPUB, význam formátu MOBI by byl výrazně menší, než by vyplývalo z jeho procentuálního zastoupení. Na vzorku dat za rok 2017 jsem proto tuto hypotézu za pomoci porovnání názvů v poli 245 ověřil.

<sup>11</sup> Uvedený formát „Android“ odkazuje na obchod Google Play, jedná se tedy buď o formát EPUB nebo PDF.

<sup>12</sup> Během zpracování dat se totiž ukázalo, že například u formátu PRC a HTML statistiku významně ovlivňuje produkce jediného nakladatele, konkrétně se jednalo o zdarma šířené e-knihy Městské knihovny v Praze.

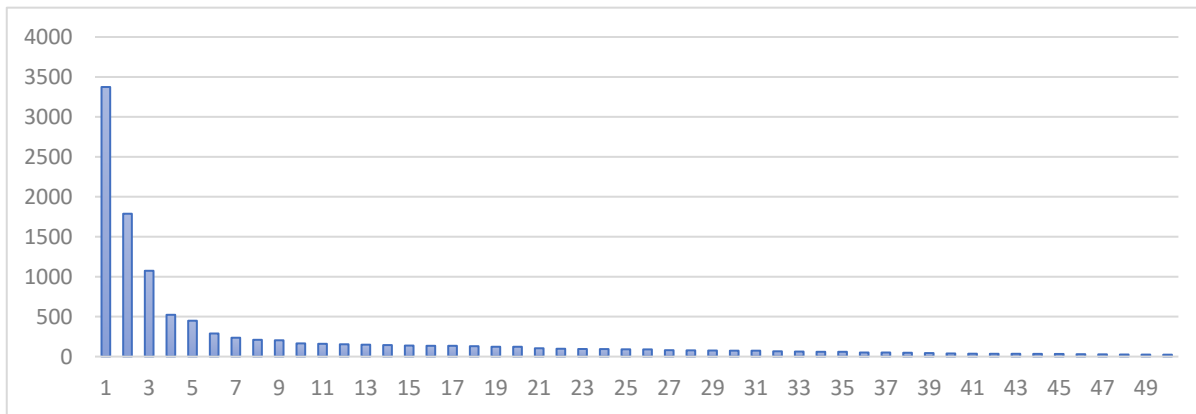
<sup>13</sup> Formát „CoverPage“ se odkazuje na stejnojmennou čtecí platformu, viz: <http://www.coverpage.cz/>

MOBI celkem	MOBI + EPUB (+ ?)	MOBI + PDF	pouze MOBI	MOBI + EPUB + PDF	MOBI + EPUB + Wooky
1595	1564	11	20	995	8

Tabulka 4. Výskyt formátů MOBI v kombinaci s dalšími formáty

Jak ukazuje tabulka 4, z 1595 e-publikací ve formátu MOBI rovnou ve 1564 případech (98 %) vyšla publikace také ve formátu EPUB a v dalších 11 případech ve formátu PDF. Pouhých 20 e-knih, pocházejících z produkce 4 nakladatelů<sup>14</sup>, vyšlo jen ve formátu MOBI. V 16 případech se přitom jednalo o publikace vydané v anglickém jazyce, určené pro zahraniční trh. Pouze 4 e-knihy od jediného nakladatele byly skutečně určeny pro český trh<sup>15</sup>.

Analýza dat za rok 2017 ukázala ještě jednu zajímavou skutečnost. Téměř polovinu přidělených čísel ISBN, konkrétně 49 %, ohlásila nejméně aktivní trojice z celkových 377 nakladatelů. Prvních deset nejméně aktivních nakladatelů ohlásilo 65 %, prvních dvacet 76 % a prvních padesát nejméně aktivních nakladatelů (13 %) reprezentovalo již 90 % přidělených ISBN. Naopak 225 (60 %) nejméně aktivních nakladatelů ohlásilo méně než 3 čísla ISBN, odpovídající například jediné publikaci, vydané ve třech nejběžnějších formátech. Tato čísla mohou teoreticky znamenat například to, že aktivní spolupráce NK ČR s pouhou dvacítkou nakladatelů by mohla zabezpečit dodání 75 % elektronických publikací v kvalitě potřebné pro dlouhodobé uložení.



Obrázek 1. Počet ohlášených ISBN u 50 nejméně aktivních producentů e-knih v roce 2017

## 2.5 Prodejci elektronických knih v ČR

Vzhledem k tomu, že statistiky přidělených ISBN nedokážou dokonale zachytit změny formátů u některých distribučních kanálů, zvolil jsem jako doplňkovou strategii při určování relevantních formátů pro český trh s e-knihami také zběžnou analýzu prodejní sítě. Jako podklad jsem využil *Zprávu o českém knižním trhu*, vydávanou každoročně Svazem českých

<sup>14</sup> Např. z Prahy působící mezinárodní nakladatelství Magic Dome Books, ve vzorku zastoupeno 13 publikacemi.

<sup>15</sup> Jedná se o nakl. Beletris (např. 978-80-7520-112-6), aktuálně již nejsou uvedné e-knihy na trhu k dispozici.

knihkupců a vydavatelů. Zatím poslední vydaná zpráva za rok 2016/2017 [21] uvádí mezi sedmi nejvýznamnějšími prodejci e-knih ČR za rok 2016 následující internetové obchody: ABZ knihy, Alza.cz, eBux.cz, eReading.cz, Kosmas.cz, Martinus.cz a Palmknihy.cz [21, s. 19]. S výjimkou eBux.cz všechny obchody nabízejí e-knihy ve formátech EPUB, MOBI a PDF. Ebox.cz nabízí všechny e-knihy ve formátu Wooky, pouze u části nabídky, u které není vyžadována přítomnost tvrdého DRM, umožňuje zaslat e-knihy také v jiném formátu<sup>16,17</sup>

Starší zprávy SČKV z předchozích let sledovaly také dostupnost česky psaných e-knih prostřednictvím zahraničních internetových obchodů Google Play Books a Apple iBooks [22, s. 21]. Zatímco Google staví na otevřených formátech EPUB a PDF, Apple využívá proprietární formát iBooks. U obou prodejců lze do budoucna očekávat spíše růst významu [23].

Navzdory velkému významu, který pro formování českého trhu s e-knihami hrála společnost Amazon prostřednictvím své hardwarové čtečky Kindle, internetový obchod Amazonu v současné době oficiálně e-knihy v českém jazyce neprodává. Nabídka českojazyčných publikací na Amazonu čítá pouze jednotky publikací, které navíc obcházejí omezení Amazonu označením angličtiny jako jazyka dokumentu. Význam Amazonu pro český trh je tedy pouze nepřímý. Přítomnost čtečky Kindle na českém trhu historicky výrazně formovala způsoby distribuce e-knih k čtenářům. Stála například za velkou oblibou formátu MOBI, který čtečky Kindle, na rozdíl od EPUB, umí zobrazit, a přispěla také k velkému rozšíření tzv. sociálního DRM, které pro čtečky Kindle oproti jiným DRM metodám nepředstavuje problém [24].

## **2.6 Dílčí závěry k 2. kapitole**

V průběhu analýzy bylo potvrzeno dominantní postavení formátů EPUB a PDF. Nepotvrdil se naopak zásadní význam formátu MOBI pro současnou českou produkci. Přestože procentuální zastoupení formátu v databázi ohlášených e-knih je značné, zpravidla se vyskytuje v kombinaci s formátem EPUB. Pohled na dlouhodobé statistiky formátů u přidělených ISBN, v kombinaci se zhodnocením podoby prodejní sítě e-knih v ČR, dále potvrzuje jako formáty hodné zřetele proprietární Wooky a iBooks. Podrobnějšímu zhodnocení těchto pěti formátů z hlediska možností jejich dlouhodobého uchování se proto budu věnovat v následujících kapitolách.

---

<sup>16</sup> Formát není specifikován, zaslání je možné výhradně na čtečka Kindle a PocketBook, s jistotou tedy půjde o formát MOBI a v případě PocketBook o formát EPUB či MOBI.

<sup>17</sup> Srov. s [23].

### **3 Historie a aktuální stav řešení problematiky e-depositu v ČR**

#### **3.1 Projekt eDpilot**

První pokus o zahájení systematického uchovávání elektronických publikací se v České republice datuje k dubnu 2011. V tomto měsíci Národní knihovna ČR spustila svůj pilotní projekt eDpilot [25], jehož cílem bylo „*mimo jiné získat výchozí podklady pro formulaci nového legislativního opatření, které ukládání elektronických dokumentů zákonem upraví.*“ [25] V rámci projektu mohli nakladatelé uzavírat s NK ČR rámcovou *Smlouvu o poskytování elektronických publikací* [26] a na jejím základě odevzdávat do knihovny své online vydávané publikace či elektronické tiskové předlohy tištěných publikací. Národní knihovna se na oplátku zavázala dodané publikace, při splnění předepsaných podmínek, dlouhodobě uchovat.

V rámci pilotního provozu NK ČR garantovala schopnost poskytnout plnou logickou dlouhodobou ochranu dokumentům odevzdaným v preferovaných formátech PDF/A-1, PDF (některé verze, např. PDF/X), HTML, XML, TXT, JPG, JPEG2000, TIFF, PNG. Mezi akceptovanými formáty byly dále tiskové podklady v jazyce PostScript, formáty DOC, RTF, ODT a další verze PDF, u nichž bylo možné provádět normalizaci do formátů PDF/A-1a nebo PDF/A-1b. U ostatních formátů, zejména MOBI nebo EPUB, byla předpokládána rovněž budoucí normalizace do archivačních formátů. V rámci tohoto projektu jim byla nicméně garantována pouze ochrana bit-streamu, nikoliv tedy ještě plná logická ochrana, jako by tomu bylo v případě odevzdání v některém z preferovaných či akceptovaných formátů. Odevzdávané dokumenty nesměly být chráněny heslem, šifrovány nebo obsahovat DRM ochranu. [27]

Projekt eDpilot byl formálně ukončen v průběhu roku 2012 v souvislosti se zahájením prací na nástupnickém čtyřletém projektu *Správa elektronických publikací v síti knihoven ČR*, financovaném z programu aplikovaného výzkumu a vývoje národní a kulturní identity – NAKI (dále jen jako *projekt NAKI*). Dobrovolný příjem elektronických publikací byl poté znovu v omezené míře obnoven v průběhu roku 2014.

#### **3.2 Projekt NAKI – Správa elektronických publikací v síti knihoven ČR**

Projekt NAKI, řešený NK ČR v letech 2012–2015, byl realizován paralelně s řešením projektu *Vytvoření Národní digitální knihovny*, realizovaného NK ČR v letech 2010–2014 [28]. Projekt NAKI měl sdílet infrastrukturu vytvořenou v rámci projektu NDK, zejména LTP úložiště, kam měly být dokumenty z vyvíjeného systému pro e-deposit ukládány. Hlavní cíle projektu shrnul Martin Žížala v roce 2012 takto: „*navrhnout a otestovat infrastrukturu, která umožní výběr*

*elektronických publikací od vydavatelů a nakladatelů, zpracování těchto publikací, jejich sdílení a zpřístupnění v knihovní síti ČR. Dalším velmi důležitým úkolem (a zároveň cílem) je dlouhodobá archivace a ochrana elektronických dokumentů. Významnou součástí projektu bude též analýza a návrh legislativy v oblasti elektronického povinného výtisku“ [29].*

Součástí úvodní etapy projektu, realizované v průběhu roku 2012, byla analýza současného stavu zejména s ohledem na legislativní a technologické otázky sběru publikací. Cílem druhé etapy (2013–2014) byl vývoj infrastruktury a metodiky přípravy elektronických publikací k archivaci a konečně pro poslední etapu, probíhající v roce 2015, byl naplánován „*poloprovoz infrastruktury pro sběr, zpracování a sdílení elektronických publikací a návrh konceptu pro centrální bibliografické zpracování.*“ [29]

V rámci úvodní analytické fáze projektu byly provedeny technické analýzy formátů elektronických publikací [30]. Byly identifikovány a zhodnoceny v té době nejčastěji používané formáty pro e-publikace na českém trhu – PDF, EPUB, MOBI, Wooky, iBooks – a určeny vhodné archivní formáty EPUB 2 a PDF/A [31, s. 123]. V případě odevzdání e-publikací v jiném formátu měla být deklarována pouze bitová ochrana [31, s. 124].

Závěrečná zpráva projektu z roku 2015 zpřesnila užití konkrétních verzí PDF/A-1 a PDF/A-2 a deklarovala budoucí snahu normalizovat dokumenty odevzdané v nearchivních formátech do archivních. Podmínkou byla zobrazitelnost jejich obsahu pomocí běžně obvyklých prostředků v čase odevzdání dokumentu a absence tvrdého DRM [3, s. 19]. Jako vhodné validační nástroje byly určeny nástroje EPUBCheck pro EPUB a Apache PDFBox pro PDF/A [3, s. 84]. Součástí závěrečné zprávy projektu byla také doporučení pro přípravu elektronických publikací v požadovaném formátu PDF/A v aplikacích Adobe InDesign, Adobe Acrobat, OpenOffice, LibreOffice a Microsoft Office a pro produkci EPUB 2 v aplikaci Sigil [3, s. 23-34].

### **3.3 Stav legislativy v ČR**

Povinný výtisk je v ČR legislativně zabezpečován souborem tří zákonů. Povinnost odevzdávat vydané neperiodické publikace nakladatelům či vydavatelům předepisuje *Zákon č. 37/1995 Sb. O neperiodických publikacích* [32], povinnost odevzdávat periodické publikace potom *Zákon č. 48/2000 Sb. O právech a povinnostech při vydávání periodického tisku (tiskový zákon)* [33]. Třetím předpisem je *Zákon č. 257/2001 Sb. o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon)* [34], který ukládá Národní knihovně ČR povinnost uchovávat a zpřístupňovat získané dokumenty. Žádný z uvedených zákonů v současné podobě situaci síťových publikací neošetřuje. [35, s. 5]

O změnu a rozšíření povinného výtisku na elektronické, online šířené, dokumenty se NK ČR snaží již od roku 2008. V průběhu řešení projektu NAKI proto vznikl komplexní návrh na úpravu legislativy k e-depositu [36], na jehož základě připravilo Ministerstvo kultury ČR v roce 2016 návrh novelizace stávajících zákonů. Vzhledem ke snaze o zachování povinnosti nakladatelů odevzdávat tištěné exempláře, návrh zcela upustil od povinnosti odevzdávat také elektronické tiskové předlohy tištěných publikací. V lednu 2017 byl vládní návrh novely zákona o neperiodických publikacích a knihovního zákona předán do Poslanecké sněmovny Parlamentu ČR [37]. K návrhu novely třetího ze zákonů – tiskového zákona – se nenašla dostatečná politická vůle a povinné odevzdávání e-periodik tak z vládní novely, navzdory protestům NK ČR, vypadlo [38] [35, s. 8]. Do konce funkčního období PSP ČR, v říjnu 2017, nakonec nebyla projednána ani novelizace zbývajících dvou zákonů. [35]

Přestože tedy novela nakonec nebyla schválena, a budoucí úprava e-depositu může, ale nemusí z nynějšího návrhu vycházet, budu v rámci této práce vycházet z parametrů tohoto návrhu. Povinnost odevzdat Národní knihovně ČR kopii online vydané neperiodické publikace by podle navrhované úpravy měl mít nakladatel u těch neperiodických publikací, kterým bylo uděleno číslo ISBN či ISMN, nebo takových publikací, které jsou šířeny komerčně. Publikace by musely být odevzdány bez užití technických ochranných prostředků (DRM). Novela zákona o neperiodických publikacích dále počítala s vydáním vyhlášky Ministerstva kultury ČR, která by stanovila konkrétní technické požadavky, způsob odevzdání a akceptované formáty dat. [37]

V případě odevzdávaných formátů se předpokládaly běžně dostupné formáty přijatelné jak pro nakladatele, tak pro knihovny. Konkrétní preferované či akceptované formáty nebyly ještě součástí návrhu, měly být MK ČR teprve definovány ve spolupráci s knihovnami a nakladateli [39, s. 11]. Národní knihovna ČR v současné době předpokládá, v souladu s výstupy projektu NAKI a v souladu s nynějším nastavením vlastního systému pro e-deposit, využití formátů PDF/A-1, PDF/A-2 a EPUB 2 [35, s. 10] [40].



## 4 Výběr archivačních formátů

### 4.1 Dlouhodobá ochrana digitálních dokumentů

Termín dlouhodobá ochrana digitálních dokumentů můžeme podle Ladislava Cubra [41] definovat takto: „*Systematická opatření pro zajištění uchování, ochrany, integrity, authenticity a dostupnosti digitálních dokumentů v dlouhodobém horizontu.*“ Opatření vedoucí k tomuto cíli můžeme v zásadě rozdělit do dvou úrovní. První a nezbytný stupeň představuje tzv. bitová ochrana (bit-level preservation)<sup>18</sup>, druhý stupeň potom představuje tzv. ochrana logická. Zatímco bitová ochrana zabezpečuje neměnné uložení binárních streamů, reprezentujících uložená data, logická ochrana cílí zejména na zachování srozumitelnosti a dlouhodobé použitelnosti digitálních dokumentů [42, s. 4]. Jako základní strategie logické ochrany jsou přitom obecně uznávány migrace či emulace [42, s. 25]. Zatímco emulace spočívá v imitaci funkcionality původního technologického prostředí, migrace předpokládá konverzi digitálního objektu do nového formátu. Zvláštní podmnožinu migrace představuje tzv. formátová normalizace, kdy dochází při vstupu digitálního objektu určitého typu, před uložením do dlouhodobého úložiště, k jeho konverzi do preferovaného archivačního formátu [43] [42, s. 26].

Výběr vhodného archivačního formátu představuje zásadní krok pro dlouhodobé uchování digitálního obsahu. Cílem výběru vhodného archivačního formátu, tedy formátu, ve kterém je dokument odevzdán k dlouhodobému uchování, nebo do kterého je po odevzdání převeden během formátové normalizace, je určení formátu, který v současnosti ani v blízké budoucnosti nebude pravděpodobně představovat větší riziko pro uchovávání digitálního obsahu [44, s. 44]. V případě formátové normalizace je nutné, aby archivační formát umožnil zachovat všechny signifikantní vlastnosti (viz dále kapitola 4.4) původního formátu [44, s. 93-94].

Provádění preventivní normalizace formátů nicméně není všeobecně a bezvýhradně přijímanou strategií digitální ochrany, vzhledem k riziku ztráty obsahu či authenticity [42, s. 25-26] [43]. V případech, kdy je to možné, je tedy vhodnější raději podporovat vznik digitálních dokumentů již rovnou ve formátech, vhodných též k dlouhodobému uložení [42, s. 26].

### 4.2 Kritéria výběru archivačních formátů

Při výběru vhodných archivačních formátů je třeba formáty posuzovat podle celé řady kritérií. Mezi nejčastěji zmiňovanými kritérii, která můžeme nalézt v doporučených postupech dokumentovaných řadou významných institucí, jsou dle Cubra tyto: rozšířenost, otevřenost a

---

<sup>18</sup> Srov. s [45, s. 76].

nezatíženost patenty [44, s. 45] [45, s. 83-86].<sup>19</sup> Otevřeností formátu je myšlena dostupnost formátové specifikace. Nezatížeností patenty se potom chápe situace, kdy případný vlastník formátu neuplatňuje v současné době výkon práv k formátu a jeho použití tak není svázáno licenčními omezeními. Kritérium rozšířenosti formátu posuzuje velikost uživatelské komunity, která formát využívá, a tím zpravidla nepřímou také dostupnost nástrojů nutných pro zobrazení či manipulaci s digitálním obsahem v daném souborovém formátu. [44, s. 45]

Jako další kritéria pro výběr archivačních formátů uvádí literatura dle Cubra [45, s. 85] tato: „*standardizace formátu uznávanou normalizační organizací (např. ISO), míra zpětné kompatibility formátu, možnost exportu do jiných formátů, míra nezávislosti formátu na specifických hardwarových a softwarových platformách, rozumná rovnováha mezi podporou různých funkcionalit na straně jedné a přiměřenou komplexitou na straně druhé.*“

Archivní formát musí být v neposlední řadě schopen plnit roli tzv. archivní matrice (*archival master*) [44, s. 43]. V praxi to znamená, že pro různé skupiny formátů by vždy měl být určen vhodný archivní formát dle jejich vlastností tak, aby normalizací do archivního formátu zůstala zachována nejvyšší možná kvalita zdrojového digitálního dokumentu [46, s. 143]. Klíčové vlastnosti, jejichž uchování je nezbytné pro zachování autenticity dokumentu, potom označujeme jako signifikantní vlastnosti (podrobněji viz kapitola 4.4).

Hodnocení formátů není vždy právě jednoduchou záležitostí, často i proto, že ideální formát, který by splňoval všechna zvažovaná kritéria, nemusí vůbec existovat. Dobrou praxí tedy je, řídit se doporučeními uznávaných institucí [44, s. 44]. Mezi základní zdroje informací pro hodnocení jednotlivých souborových formátů patří dlouhodobě formátový registr Kongresové knihovny [47], doplňovaný každoročně aktualizovaným doporučením preferovaných a akceptovaných formátů pro jednotlivé typy dokumentů [48]. Cenný zdroj informací pro oblast e-publikací představuje také hodnocení formátů Britské národní knihovny [49] [46] nebo Harvardské knihovny [50]. Zejména pro formáty z rodiny PDF je relevantním zdrojem také formátový registr s hodnocením formátových rizik od švýcarské expertní skupiny KOST-CECO<sup>20</sup> [51].

---

<sup>19</sup> Srov. s [176], [194], [200], [198], [199], [195].

<sup>20</sup> KOST-CECO – Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen - Centre de coordination pour l'archivage à long terme de documents électroniques.

### 4.3 Formátové profily

Samotný výběr vhodného archivačního formátu nemusí být vždy dostačující. Pro zajištění dlouhodobého bezpečného uchování může být nutné definovat užší vymezení vlastností a nastavení v rámci specifikace konkrétního formátu – formátový profil. Příkladem profilu u obrazových dat může být povinnost použít bezeztrátovou kompresi [44, s. 51]. V případě e-publikací může jít například o zákaz šifrování obsahu za využití DRM, zákaz využívat JavaScript nebo zákaz vkládat jiné formáty než ty, které jsou přímo podporovány specifikací formátu<sup>21</sup>. Profil u archivačního formátu má za cíl snížit do budoucna rizika, plynoucí například ze závislosti na externích prvcích nebo z využití vlastností formátů, které nejsou v současné době ještě v dostatečné míře podporovány aplikacemi a je u nich tedy vyšší riziko, že je nebude v budoucnu možné zobrazit či migrovat do jiného formátu. Pro některé souborové formáty existují, nebo vznikají přímo standardizované profily určené specificky pro potřeby dlouhodobého uchování – příkladem je PDF/A, odvozený od standardu PDF nebo aktuálně vznikající standard TI/A [52], odvozený od specifikace TIFF. Konkrétní vymezení vlastností přijímaných formátů si instituce mohou určit jako součást publikované formátové politiky<sup>22</sup>.

### 4.4 Signifikantní vlastnosti

Projekt InSPECT<sup>23</sup>, probíhající v letech 2007–2009 [53] definoval pojem signifikantních vlastností ve své závěrečné zprávě [54] takto: „*The characteristics of an Information Object that must be maintained over time to ensure its continued access, use, and meaning, and its capacity to be accepted as evidence of what it purports to record*“ [54, s. 4].

Stanovení signifikantních vlastností je důležitou pomocnou strategií v případě provádění formátové migrace, která je mnohdy nevyhnutelně ireverzibilní [45, s. 89]. Před případnou normalizací dat do archivačního formátu je proto vhodné definovat soubor klíčových vlastností, které musí být zachovány i po převedení digitálního obsahu do formátu nového. Zachování těchto vlastností je základem k udržení autenticity digitálního dokumentu. Naopak vlastnosti, které nejsou repozitářem považovány za signifikantní, mohou být v průběhu formátové migrace změněny. Příkladem nesignifikantních vlastností může být koncovka souboru nebo bitová velikost. Příkladem signifikantních vlastností počet slov v dokumentu, počet odstavců nebo formátování textu. Slabinou této strategie je určitá míra subjektivity při výběru signifikantních

---

<sup>21</sup> Srov. s [99].

<sup>22</sup> Příkladem takto definovaného formátového profilu jsou profily pro vytváření archivních a uživatelských kopií obrazů v JPEG2000, vydané NK ČR jako součást tzv. NDK standardů [202].

<sup>23</sup> InSPECT – Investigating the Significant Properties of Electronic Content Over Time.

vlastností, a v některých případech rovněž při hodnocení těchto vlastností [44, s. 93]. Různá může být také míra detailu<sup>24</sup>, na jakém jsou signifikantní vlastnosti definovány.<sup>25</sup>

Určení signifikantních vlastností jednotlivých souborových formátů (v kontextu konkrétních digitálních sbírek) je důležité nejen pro kontrolu úspěšnosti provedené normalizace, ale zejména již pro samotný výběr přijímaných a archivačních formátů. V případě provádění normalizace dat do archivačních formátů musí být tento cílový formát schopen zachovat všechny zvolené vlastnosti formátů původních. Za signifikantní vlastnost nebude nejspíše považována taková vlastnost, která odporuje zvolenému formátovému profilu instituce. Pokud například formát podporuje JavaScript, ale repozitář tuto jeho vlastnost považuje za rizikovou, nebude považována ani za signifikantní vlastnost, kterou by repozitář garantoval uchovat (za předpokladu, že takový soubor k uložení vůbec přijme).

---

<sup>24</sup> Např. závěrečná zpráva projektu inSPECT [54, s. 19-20] definuje 37 konkrétních signifikantních vlastností jen pro popis strukturovaného textu. Jiný příklad definice 26 signifikantních vlastností pro textové dokumenty nabízí pro srovnání na svém webu americká Kongresová knihovna [203].

<sup>25</sup> Srov. s [190].

## 5 Správa digitálních práv (DRM)

Akronym DRM vznikl z anglického *Digital Right Management*, česky se zpravidla překládá jako *Správa digitálních práv* [55]. Česká terminologická databáze z oblasti knihovnictví a informační vědy (TDKIV), definuje termín *Správa digitálních práv* jako: „*Technologie, která kontroluje a omezuje používání digitálních děl (elektronické hudby, filmů, elektronických knih, počítačových her apod.) s cílem zabránit nelegálnímu užití těchto děl, např. zamezuje uživatelům kopírování digitálního díla, stanovuje určité časové období, po které může být toto dílo používáno, omezuje počet zařízení, na kterých může být dílo užíváno, zakazuje konverzi díla do jiných formátů apod. Technologii DRM používají poskytovatelé digitálních děl jako např. hudební vydavatelé, filmoví producenti, knižní nakladatelé, autoři apod.*“ [55].

Za účelem ochrany autorského obsahu publikací je používáno několik různých technických metod zabezpečení publikací, které je možno rozdělit do tří základních skupin [56, s. 42]:

- uzavřený ekosystém
- digitální vodoznak
- šifrování obsahu

V praxi se často setkáváme s kombinacemi těchto metod v podobě komplexních DRM schémat.

### 5.1 Uzavřený ekosystém

Jedná se o takový distribuční model, kdy distributor umožňuje stahovat a konzumovat obsah pouze prostřednictvím jím určené aplikace. Součástí uzavřenosti ekosystému bývá zpravidla užití proprietárních formátů, které výrazně znesnadňují užití obsahu na jiných zařízeních, neboť nejsou jiným aplikacím srozumitelné. Úspěšným zahraničním příkladem užití této metody je iBookstore společnosti Apple, využívající proprietární formát iBooks. Podobným příkladem z českého prostředí je platforma Wooky [57]. [56, s. 42-43]

### 5.2 Digitální vodoznak

Metodou digitálního vodoznaku nebo též watermarkingu rozumíme proces vkládání specifických informací, majících zpravidla za cíl identifikovat uživatele, do viditelných či neviditelných částí souboru tvořícího e-publikaci. Informace jsou vkládány takovým způsobem, aby je nebylo možné snadno odstranit, a to zpravidla ani změnou souborového formátu [58, s. 26]. Cílem watermarkingu je zejména jednoznačná zpětná identifikace uživatele, který je zodpovědný za neoprávněné šíření díla. Pokud je uživatel v dokumentu označen viditelným

způsobem a cílem digitálního vodoznaku je tedy zejména preventivně psychologicky působit proti kopírování, hovoříme o tzv. sociálním DRM [58, s. 26]. Naopak v případě, kdy označení uživatele není viditelné, a jeho účelem je tedy spíše zpětně vypátrat viníka případného úniku, lze mluvit o tzv. forezním DRM<sup>26</sup> nebo také forezních vodoznacích [59].

Specifickým způsobem užití watermarkingu je kombinace se zašifrováním souboru, kdy jsou do knihy vložena neviditelná metadata s informací o uživatelských právech k dokumentu, spolu s dešifrovacím klíčem. Čtecí aplikace potom zkontroluje uživatelská práva zobrazit dokument např. s ohledem na povolená zařízení a pokud potřebné vložené informace najde a podmínky zobrazení jsou splněny, publikace je dešifrována. [56, s. 43] [58, s. 26-27]

V českém prostředí je ochrana prostřednictvím tzv. sociálního DRM, kdy je uživatel viditelně označen jménem a například adresou, vůbec nejčastější formou ochrany. Jednou z příčin je značná obliba čteček elektronických knih Kindle mezi českými čtenáři. Protože čtečka Kindle je produktem firmy Amazon, prosazující svůj uzavřený ekosystém, čtečky Kindle nepodporují jiná DRM schémata než vlastní. Distributoři, kteří chtěli prodávat obsah uživatelům čtečky Kindle, tak v důsledku měli na výběr mezi žádnou ochranou nebo sociálním DRM. [57]

### **5.3 Šifrování obsahu**

Metoda šifrování obsahu byla částečně již zmíněna v souvislosti s watermarkingem. Jedná se o způsob zabezpečení, kdy je digitální obsah distribuován v zašifrované podobě a nemůže být zobrazen, pokud nedojde k jeho dešifrování za užití dešifrovacího klíče [60, s. 229] [56, s. 43-44] Prostou variantou užití metody šifrování je podmínění odemčení obsahu vložáním hesla. V takovém případě nevyžaduje odemčení obsahu připojení k žádnému vzdálenému licenčnímu serveru. Zabezpečení pomocí hesla je běžné zejména u PDF dokumentů. [60]

### **5.4 DRM schémata**

Složitější variantou DRM využívající šifrování obsahu je využití komplexních DRM schémat. Ta sestávají nejčastěji ze 3 závislých částí – čtecí aplikace, obsahového serveru a licenčního serveru. Zatímco obsahový server poskytuje soubor čtecí aplikaci a dešifrovací klíč licenčnímu serveru, licenční server řídí správu uživatelů a poskytuje čtecí aplikaci po úspěšné autentizaci dešifrovací klíč. [60, s. 229]

---

<sup>26</sup> V anglickém jazyce se používá termín „forensic DRM“ nebo též „forensic watermarking.“

Pro vyšší míru zabezpečení je zpravidla možné využít dešifrovací klíč jen v kombinaci s konkrétními registrovanými zařízeními uživatele [56, s. 44]. Pro velké distribuční platformy, jakými jsou například Apple či Amazon, je typická kombinace šifrování a užití uzavřených ekosystémů s proprietárními formáty. V českém prostředí je podobným příkladem již dříve zmiňovaná platforma Wooky.

### **5.5 Nejdůležitější DRM schémata pro e-publikace**

Na světě jsou v současné době dostupná čtyři velká DRM schémata využívaná pro e-publikace. Tři proprietární schémata Amazon DRM (Amazon), Apple FairPlay (Apple) a ADEPT<sup>27</sup> (Adobe) doplňuje open-source schéma Marlin (Marlin Developer Community<sup>28</sup>). [56, s. 44-47]

Schéma Marlin není na trhu e-publikací zatím příliš rozšířené, v ČR pak není dle dostupných informací používáno vůbec. Schéma od Amazonu je významné v globálním měřítku, vzhledem k jeho užití výhradně na platformě Amazonu a k absenci podpory Amazonu pro český obsah<sup>29</sup> se u českých e-publikací nicméně prakticky nevyskytuje. Na českém trhu se tak nejčastěji lze setkat se schématy ADEPT, které dlouhodobě využívá řada českých distributorů e-knih [61], nebo se schématem Apple FairPlay pro obsah distribuovaný prostřednictvím iBooks Store.

V kontextu českého trhu je třeba zmínit rovněž DRM řešení používané v ČR distribuční platformou Wooky. O používaném schématu nicméně nejsou, vyjma jeho proprietárnosti a nekompatibilitě s čtečkami ostatních výrobců [57], žádné bližší informace.

### **5.6 Zhodnocení dopadů užití DRM pro dlouhodobou archivaci**

Ve vztahu k dlouhodobému uložení e-publikací je nutné rozlišovat použité metody DRM. Metoda digitálního vodoznaku ve formě tzv. sociálního DRM zvláštní rizika pro dlouhodobou ochranu nepředstavuje. Fakticky se jedná o pouhou obdobu běžných provenienčních znaků, jaké známe z analogového knižního světa ve formě ex-libris nebo razítek. Případná přítomnost

---

<sup>27</sup> ADEPT – Adobe Digital Experience Protection Technology

<sup>28</sup> Zakládajícími členy v roce 2008 byly společnosti Intertrust, Panasonic, Philips, Samsung a Sony [56, s. 44-46].

<sup>29</sup> Elektronický obchod Amazonu v současné době oficiálně nepodporuje prodej děl v českém jazyce, v dubnu 2018 tak bylo možné v nabídce nalézt pomocí hledání „czech edition“ pouze 11 česky psaných titulů, které obešly omezení Amazonu tím, že mají uveden jako jazyk díla anglický jazyk.

sociálního DRM v knize, pokud by se například nakladatel rozhodl vložit identifikační údaje instituce, které je e-publikace předána k uložení, uchování publikace neohrožuje.<sup>30,31</sup>

V případě tzv. tvrdého DRM za využití šifrování již rizika významně narůstají. Hrozí nejen ztráta přístupu k obsahu, ale případná technická omezení bránící kopírování nebo závislost na konkrétním softwarovém prostředí mohou zásadním způsobem komplikovat provádění základních ochranných akcí v rámci logické ochrany uložených digitálních objektů – formátové migrace a emulace. [60, s. 229] [62, s. 48-50]

Stefan Hein a Tobias Steinke z Německé národní knihovny dělí ve své studii [60] e-publikace chráněné pomocí šifrování na dvě skupiny podle toho, zda je dešifrování závislé na komunikaci s vzdáleným licenčním serverem, či zda se jedná pouze o lokálně zaheslovaný a zašifrovaný soubor. Zatímco v druhém případě považují (při znalosti hesla) riziko za středně vysoké, v případě závislosti na licenčním serveru, jsou již rizika považována jako vysoká.<sup>32</sup>

O tom, že obavy vyplývající z možné nedostupnosti licenčních serverů nejsou plané, svědčí řada příkladů z minulosti. Miles Carr [63, s. 164-167] například zdokumentoval více než dvacet příkladů z posledních 12 let, kdy původní licenční server přestal být dostupný, ať již z důvodu ukončení provozu distribučního kanálu, nebo z důvodu přechodu na novější technologii. V mnoha případech to znamenalo ztrátu přístupu uživatelů k legálně pořízeným souborům.

Jako modelový příklad poslouží ukončení provozu Adobe Content Server 3 v roce 2009. Společnost Adobe Systems poskytla uživatelům 9 měsíců k migraci dat na nový Adobe Content Server 4. Kdo tak v termínu neučinil, ztratil okamžikem reinstalace čtecí aplikace k zakoupenému obsahu zcela přístup [64]. Podobně společnost Microsoft v roce 2011 oznámila ukončení podpory svému formátu .lit a čtečky Microsoft Reader [65]. O rok později již nebylo možné aktivovat nové instalace aplikace Microsoft Reader a soubory v proprietárním formátu .lit, které byly chráněny DRM, se staly pro uživatele nedostupné. [63, s. 164-167]

Při posuzování dopadů technických opatření pro správu digitálních práv na možnost dlouhodobého uchování dat je potřeba zmínit také možnosti odstranění těchto technických opatření. Je obecně známou skutečností, že dlouhodobě je každá technologie bránící digitální obsah proti nelegálnímu kopírování odsouzena k prolomení. Některé způsoby ochrany jsou odstranitelné pouhou formátovou migrací, prolomitelná a prolomená jsou všechna velká

---

<sup>30</sup> Například portál e-deposit NK ČR možnost vložení sociálního DRM do e-publikace ve formě „Národní knihovna Praha“ ze strany vydavatele přímo zmiňoval [196].

<sup>31</sup> Srov. s [62, s. 71].

<sup>32</sup> Srov. s [62, s. 30-33].



globálně užívaná DRM schémata [66] [67], a rovněž prolomení šifrování za využití hesel je teoreticky podmíněné pouze dostupností dostatečné výpočetní kapacity v závislosti na složitosti použitého hesla [62, s. 49]. [68]

Při zvažování metod na odstranění DRM je nicméně nutné velmi důsledně posuzovat legální otázky, stejně jako otázky zachování autenticity a integrity dotčených děl. Nutné je zároveň zvažovat nezanedbatelné náklady na údržbu dodatečných nástrojů potřebných pro dané účely a celkové zvýšení nákladů pro správu dotčených dat. [62, s. 49, 72-73]

Na závěr kapitoly si vypůjčím výstižné hodnocení od Amy Kirchoff a Sheily Morrissey, které ve své studii *Preserving eBooks* [69] z roku 2014 shrnuly problematiku DRM takto: „*DRM constitutes a challenge in the preservation of eBooks. It is challenging from the business model perspective – if a book is ‘sold’ with limited use, what preservation rights come with that sale? And it is also challenging from a technical perspective, where DRM can impede the preservability of an eBook. Further, a change in digital rights technologies can be seen as a specialized case of format obsolescence. [...] If however DRM – and, more particularly, its encryption – cannot be removed by preservation institutions from the preservation copy of an object, that object will be opaque to future viewers. Preservation in such cases would be byte-level preservation of an unintelligible digital object*“ [69, s. 10].

## 6 Zhodnocení vybraných formátů e-publikací

### 6.1 Open eBook (OEBPS)

Název souborového formátu	Open eBook Publication Structure (OEBPS)
MIME type	application/oebps-package+xml
PUID	-
Přípona souboru	.opf
Tvůrce formátu	Open eBook Authoring Group
Datum publikování formátu	9.1999
Předcházející formát	-
Nástupnický formát	EPUB

Tabulka 5. Přehled vybraných údajů o formátu OEBPS

#### 6.1.1 Historie formátu

Formát Open eBook Publication Structure (OEBPS) byl zveřejněn v září 1999 sdružením Open eBook Authoring Group [70]. Toto sdružení vzniklo v říjnu 1998, při příležitosti konání první konference věnované e-knihám, svolané americkým Národním institutem standardů a technologií<sup>33</sup> (NIST). Mezi původními 25 členy sdružení byli vydavatelé, distributoři, autoři, akademici, ale také velké technologické firmy v čele s Microsoftem, Adobe nebo Nokií. Cílem bylo vytvořit otevřený standard na bázi XML, který by – podobně jako PDF – umožňoval uživatelům číst e-knihy napříč různými platformami a předešel tak vývoji vlastních, vzájemně nekompatibilních proprietárních řešení jednotlivými výrobci. [71] [56, s. 12]

Open eBook Authoring Group se v lednu 2000 přejmenovala nejprve na Open eBook Forum, v roce 2005 potom na International Digital Publishing Forum (IDPF). Formát OEBPS byl dvakrát aktualizován, nejprve v lednu 2001 na verzi 1.0.1 a poté v srpnu 2002 na verzi 1.2. [71]

Slabinou standardu bylo od počátku zaměření specifikace na definici obsahu dokumentu, nikoliv na podobu distribučního balíčku či ochranu duševního vlastnictví. To vedlo k různým proprietárním rozšířením pro kódování binárních souborů, jako byly formáty LIT od Microsoftu nebo úspěšnějšího MOBI od Mobipocket SA. Otevřené řešení nakonec přinesl až nový formát EPUB 2.0, kterým byl OEBPS v roce 2007 nahrazen. [56, s. 12]

<sup>33</sup> NIST – National Institute of Standards and Technology.

### 6.1.2 Popis formátu

OEBPS je formát založený na standardech XML 1.0, HTML 4.0, XHTML 1.0, CSS 2, meta-datovém standardu Dublin Core a znakové sadě Unicode. Může obsahovat vložené obrazové soubory ve formátech JPEG nebo PNG [72, s. 3-7]. Dokument může využívat také jiné vložené formáty než ty, které definuje specifikace standardu, v takovém případě ale musí obsahovat pro správné zobrazení na všech zařízeních také alternativní vyjádření v některém z podporovaných formátů [72, s. 18-19]. Každý dokument podle specifikace sestává z množiny souborů tvořících obsah a z povinného souboru *OEB package file* ve formátu XML s doporučenou příponou .opf. Tento *OEB package file* obsahuje výčet všech souborů, které tvoří výsledný dokument, určuje strukturu dokumentu a pořadí čtení jeho jednotlivých součástí. [72, s. 18-19].

V roce 2007 IDPF rozhodlo o nahrazení formátu OEBPS 1.2 a o jeho rozdělení na dva dílčí nástupnické standardy: OPS 2.0 (the Open Publication Structure), OPF 2.0 (the Open Packaging Format). Zjednodušeně řečeno lze říct, že zatímco OPS představuje standard pro prezentaci obsahu elektronické publikace, OPF je standardem definujícím formální strukturu dokumentu [73]. Již v září 2006 navíc vydala IDPF specifikaci standardu OCF 1.0 (OEBPS Container Format), který definuje způsob, jakým mohou být všechny součásti balíčku zabaleny do jednoho kontejneru ZIP, který pak tvoří výsledný dokument. Tyto tři dílčí specifikace společně vytvořily základy standardu EPUB, kterým byl OEBPS v září 2007 nahrazen. [74]

### 6.1.3 Zhodnocení formátu z hlediska dlouhodobého uložení

Standard OEBPS nepředstavuje z pohledu dlouhodobé archivace významná rizika. Jedná se o formát vystavěný na otevřených standardech, a je dobře zdokumentovaný. Přestože dokumentace není od roku 2012 na webu IDPF dostupná, je dosažitelná prostřednictvím služby Internet Archive's Wayback machine [75]. Vzhledem k zastaralosti formátu je v případě výskytu vhodná migrace do formátu EPUB 2.0.1. Pro konverzi lze využít nástroje Calibre nebo Sigil [76]. Význam standardu je však spíše historický, stal se základem pro odvození celé řady dalších formátů pro e-knihy. V současné komerční produkci se již nevyskytuje.

## 6.2 Mobipocket (MOBI)

Název souborového formátu	Mobipocket (MOBI)
MIME type	application/x-mobipocket-ebook
PUID	fmt/396
Přípona souboru	.mobi, .prc
Tvůrce formátu	Mobipocket SA (firma koupena 2005 Amazonem)
Datum publikování formátu	po roce 2000
Předcházející formát	OEBPS, PalmDOC
Nástupnický formát	AZW

Tabulka 6. Přehled vybraných údajů o formátu MOBI

### 6.2.1 Historie formátu

Formát byl vytvořen francouzskou firmou Mobipocket SA krátce po jejím vzniku, který se datuje k březnu 2000 [77, s. 2]. Cílem firmy bylo vytvořit formát, který by podporoval formátování textu a byl vhodný pro čtení na tehdy rozšířených PDA zařízeních<sup>34</sup>. V roce 2005 byla firma Mobipocket SA koupena americkou společností Amazon. Od té doby je tedy formát spojen s touto firmou a zejména potom s její čtečkou Kindle [56, s. 15]. V roce 2011 Amazon oznámil ukončení podpory formátu MOBI [78] a v roce 2016 došlo k ukončení provozu webu Mobipocket.com [79], obsahujícího též publikovanou část dokumentace formátu. Nástupcem MOBI se stal mírně modifikovaný formát AZW z roku 2007 [80]. V současné době je také AZW nahrazován novějším proprietárním formátem KF8<sup>35</sup>, odvozeným od EPUB 3 [81].

### 6.2.2 Popis formátu

Formát MOBI byl vytvořen na základě staršího standardu OEBPS, doplněného o řadu proprietárních rozšíření [82], a také formátu PalmDOC (.pdb<sup>36</sup> nebo .prc<sup>37</sup>) z roku 1996. Formát PalmDOC byl, podobně jako později MOBI, šířen v komprimované podobě ve formě binárního souboru, nepodporoval však ještě takřka žádné formátování textu [83]. Od formátu PalmDOC zdědil MOBI také příponu .prc, která se pro formát někdy používá namísto běžnější .mobi. Základem MOBI je XHTML a CSS. Formát umožňuje využití JavaScriptu a rámců (frames), podporuje ochranu pomocí DRM. MOBI umožňuje pouze zobrazení tzv. reflowable obsahu, tedy obsahu, který se dynamicky přizpůsobuje v závislosti na velikosti displeje zařízení a

<sup>34</sup> PDA – Personal Digital Assistant.

<sup>35</sup> Známý též pod označením AZW3.

<sup>36</sup> PDB – Palm Database Format.

<sup>37</sup> PRC – Palm Resource Compiler.

zvolené velikosti textu. V případě potřeby lze zobrazení obsahu s pevnou sazbou zajistit pouze vložením obrazu stránky v podobě bitmapového souboru. [69, s. 13-14]

Formát MOBI lze rozdělit na dvě části. První část tvoří definice obsahu, jeho formátování a připojená metadata. K této části, odvozené ve velké míře od OEPBS, byla do roku 2016 pro vývojáře dostupná dokumentace na webu Mobipocket.com<sup>38</sup> [82]. Druhou část tvoří binární formát sloužící k zabalení a distribuci publikace. Jedná se o kontejner, který vznikne kompilací ze zdrojových souborů, které dohromady tvoří obsah publikace. Specifikace binárního formátu byla odvozena od formátu PalmDOC [84] [85]. Zabalené soubory mohou být komprimované či nekomprimované a součástí může, ale nemusí být šifrování obsahu. K této části specifikace formátu nikdy nebyla oficiální dokumentace publikována [77, s. 2]. Částečnou dokumentaci se podařilo shromáždit na webu MobileRead Wiki pomocí metody reverzního inženýrství [84].

Formát defaultně podporuje tři typy šifrování. Prvním je jednoduché šifrování, které pouze limituje užití publikace na oficiální čtecí aplikaci Mobipocket. Druhým typem je šifrování v kombinaci s heslem a třetím je šifrování za využití externí DRM autentizace. [86]

### **6.2.3 Formát AZW**

Formát AZW je fakticky pouze specifickým proprietárním profilem formátu MOBI, který je výlučně používán pro distribuci e-publikací z elektronického obchodu Amazonu. AZW tedy odpovídá formátu MOBI, s tou podmínkou, že binární soubor je vždy kompilován za využití komprese. Formát AZW může, ale nemusí obsahovat DRM ochranu. V případě užití je vždy aplikováno proprietární DRM schéma od Amazonu. O shodě mezi formáty AZW a MOBI svědčí například zaměnitelnost koncovek .mobi, .prc a .azw. Pro vytvoření AZW souboru tak kupříkladu stačí přejmenovat koncovku souboru MOBI na .azw [77, s. 3]. [80]

### **6.2.4 Zhodnocení formátu z hlediska dlouhodobého uložení**

Při zhodnocení formátu jsem vycházel zejména z hodnocení formátu od Britské národní knihovny z roku 2018 [77]. Britská knihovna ve své zprávě shrnuje rizika formátu MOBI do těchto bodů: proprietárnost formátu, absence úplné oficiální dokumentace, morální zastarání formátu, klesající podpora čtecích aplikací, možná přítomnost DRM, absence validačních nástrojů a v menší míře také možnost vloženého obsahu v dalších formátech [77, s. 8-9].

---

<sup>38</sup> Provoz webového sídla Mobipocket.com byl společností Amazon ukončen v listopadu 2016, obsah webu je nicméně nadále dostupný prostřednictvím služby Internet Archive's Wayback machine.

Ukončením podpory formátu v roce 2011 a provozu webu Mobipocket.com v roce 2016 vyslala společnost Amazon světu zprávu, že formát MOBI považuje za zastaralý a jakoukoli další podporu tohoto formátu nepokládá za opodstatněnou. Spolu s ukončením provozu webu tak došlo k zneprístupnění nejen publikované části oficiální dokumentace k formátu, ale také například oficiální čtecí aplikace Mobipocket Reader. Dokumentace formátu je v současnosti dostupná pouze prostřednictvím služby Internet Archive's Wayback machine [82].

Z pohledu podpory softwarových aplikací se jeví za problematickou zejména absence validačních nástrojů. V současné době neexistuje validátor, který by dokázal zhodnotit validitu souboru MOBI alespoň v rozsahu publikované části formátové specifikace [77, s. 5]. Lepší situace je oproti tomu ohledně dostupnosti vhodných konverzních nástrojů třetích stran do otevřených formátů. Nejčastěji uváděným nástrojem je open-source nástroj Calibre, který umožňuje konverze MOBI souborů (nechráněných DRM) do jiných formátů, jako je například EPUB [77, s. 5]. Významnou pomocí při správě obsahu v MOBI do budoucna může být také open-source nástroj KindleUnpack, který dokáže z binárních MOBI souborů (nechráněných DRM) vytvořit zpět zdrojové soubory. [77, s. 5-6]

Rizikem v případě migrace z formátu MOBI může být nicméně důvěryhodnost migračních nástrojů. Vzhledem k nekompletní dokumentaci musely tyto nástroje nutně vzniknout přinejmenším z části metodou reverzního inženýrství, nikoliv podle podrobné specifikace formátu, nemusejí tedy ve všech případech garantovat zcela uspokojivý výsledek.

## 6.3 EPUB 2

Název souborového formátu	EPUB
MIME type	application/epub+zip
PUID	fmt/483
Přípona souboru	.epub
Tvůrce formátu	International Digital Publishing Forum (IDPF)
Datum publikování formátu	9.2007
Předcházející formát	OEBPS
Nástupnický formát	EPUB 3

Tabulka 7. Přehled vybraných údajů o formátu EPUB 2

### 6.3.1 Historie formátu

Formát EPUB vydala organizace International Digital Publishing Forum (IDPF) v září 2007 jako nástupnický formát OEBPS 1.2 [87, s. 3]. Jelikož evolučně navazoval na dřívější formát OEBPS a původně byl vyvíjen pod označením OEBPS 2.0, byl formát EPUB vydán pod novým označením rovnou jako verze 2.0. V květnu 2010 byla vydána aktualizovaná verze 2.0.1. V říjnu 2011 byl potom EPUB 2 nahrazen novou verzí formátu nazvanou EPUB 3. [88]

### 6.3.2 Popis formátu

Formát EPUB 2 je definován třemi otevřenými standardy: Open Publication Structure (OPS), Open Packaging Format (OPF) and Open Container Format (OCF) [89]. Tyto specifikace jsou založeny na standardech XML 1.0, XHTML 1.1<sup>39</sup>, DTBook, CSS 2, formátu ZIP, metadatovém standardu Dublin Core a znakové sadě Unicode [90] [91] [92]. EPUB 2 může dále obsahovat vložené bitmapové obrazy ve formátech JPG, PNG nebo GIF a vektorovou grafiku ve formátu SVG [92]. Mimo to může zahrnovat také jiné vložené formáty než v OPS definované tzv. Core Media Types, v takovém případě ale musí být vloženo pro správné zobrazení na všech zařízeních také záložní vyjádření v některém z podporovaných formátů [91].

Na rozdíl od OEBPS je EPUB šířen, stejně jako MOBI, ve formě binárního komprimovaného souboru, v tomto případě slouží jako kontejner formát ZIP, který je otevřeným formátem [93] [90, s. 17, 18]. Strukturu formátu EPUB popisuje Johan van der Knijff následovně: „Conceptually, an EPUB file is just an ordinary ZIP archive. Inside the archive, the contents of a publication are (in the simplest case) represented as one or more XHTML files, which may be organised in one or more directories. CSS files (Cascading Style Sheets) are used to define layout

<sup>39</sup> Textový obsah publikace může být alternativně ve formátu XHTML nebo ve formátu DTBook [94, s. 5, 8].

*and formatting. In addition, a number of XML files provide various kinds of (mostly structural) metadata. At a very basic level, an EPUB file could be described as "a webpage inside a ZIP file", although this is a bit of an oversimplification. Unsurprisingly, the EPUB format is largely based on existing web standards“ [94, s. 3].*

Formát EPUB 2 umožňuje pouze zobrazení tzv. reflowable obsahu, tedy obsahu, který se dynamicky přizpůsobuje v závislosti na velikosti displeje zařízení a zvolené velikosti textu. Neumožňuje pevnou sazbu dokumentu ani složitější typografické úpravy. Formát nepodporuje JavaScript ani multimediální obsah [94, s. 26, 28].

### **6.3.3 Zhodnocení formátu z hlediska dlouhodobého uložení**

Formát EPUB byl, vzhledem ke své rozšířenosti, předmětem celé řady formátových studií, které posuzovaly jeho vhodnost pro dlouhodobé uchování. Za nejvýznamnější považují studii Johana van der Knijffa pro nizozemskou Královskou knihovnu z roku 2012 [94], zhodnocení Britské národní knihovny z roku 2015 [87], záznam ve formátovém registru americké Kongresové knihovny [73], nebo formátovou studii Paula Wheatleyho vypracovanou pro Harvardskou knihovnu v roce 2014 [95]. Z těchto a dalších dílčích studií, odkazovaných dále v textu, vyplývá, že formát EPUB 2.0 nepředstavuje velká rizika pro dlouhodobé uchování. K některým částem formátové specifikace je nicméně přesto nutné přistupovat obezřetně.

Jedná se o otevřený formát s dostupnou dokumentací a bez licenčních omezení. V současné době je sice již nahrazen formátem EPUB 3, podpora softwarových aplikací pro zobrazení, validaci nebo též migraci z a do jiných formátů je dostačující. Formát je v současné době podporován většinou hardwarových i softwarových čteček určených pro čtení elektronických publikací. Jako příklad nástroje vhodného pro konverze mezi formátem EPUB 2 a jinými formáty se v literatuře nejčastěji uvádí open-source nástroj Calibre. K validaci formátu lze využívat nástroje EPUBCheck nebo Flightcrew. EPUBcheck je možné využít též k charakterizaci souborů nebo k detekci přítomnosti DRM [87, s. 5]. [87]

Nástroj EPUBCheck je také součástí modulárního validátoru Flint [96] vyvinutého v rámci evropského projektu SCAPE [97]. Flint využívá služeb řady validačních nástrojů a umožňuje jejich výstupy v XML automaticky porovnat oproti profilu formátu, definovanému institucí pomocí schématu v jazyce Schematron. Profil instituce může definovat například zákaz DRM šifrování nebo omezit souborové formáty vloženého obsahu. V době ukončení vývoje deklaroval Flint schopnost pracovat s formátem EPUB a PDF [96]. Vývoj nástroje Flint



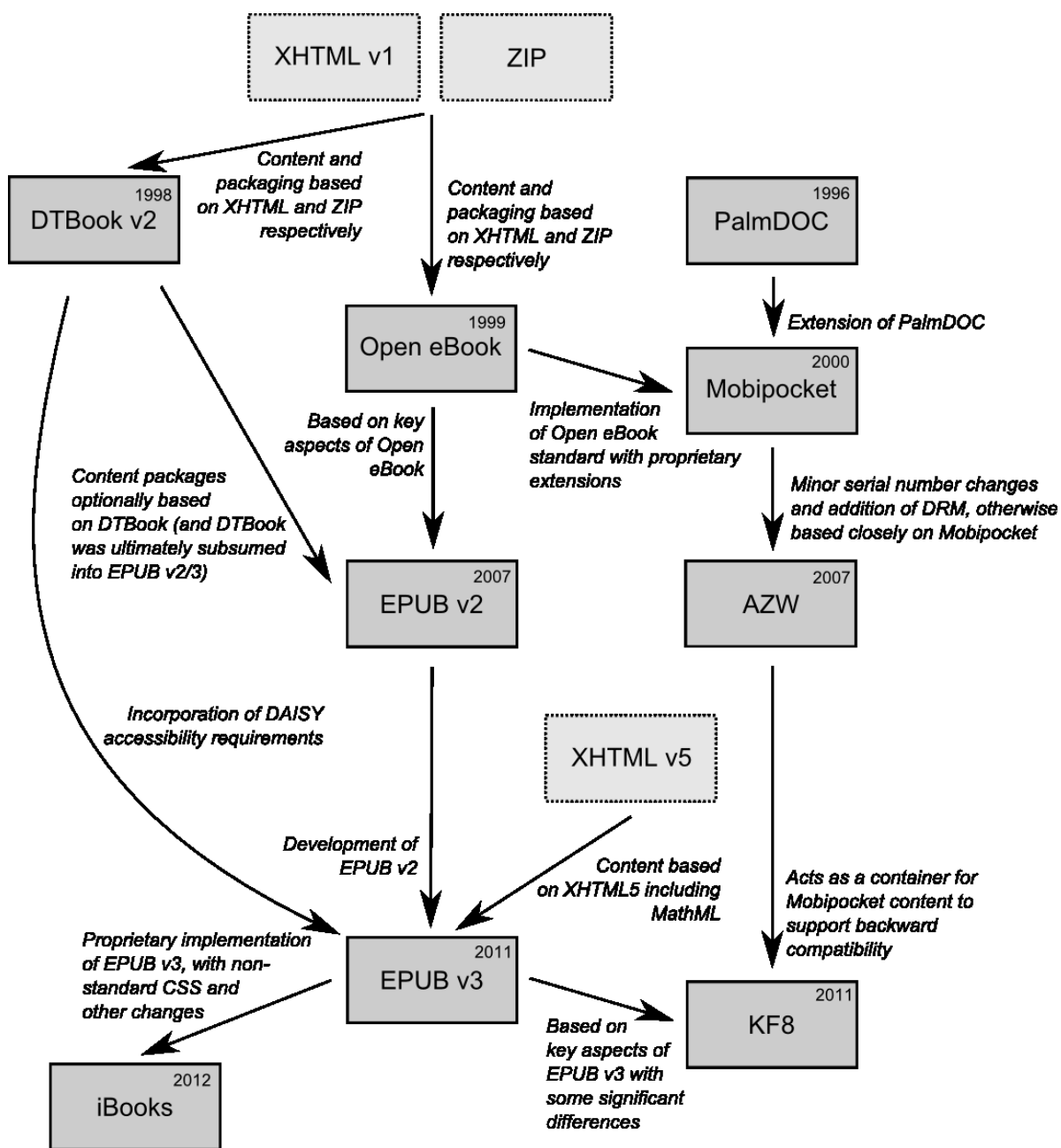
nicméně dle dostupných informací skončil spolu s ukončením projektu SCAPE v září 2014, a poslední verze nástroje je tedy předprodukční 0.8.0 [98].

Obsah vložený do souboru EPUB v podporovaných formátech JPG, PNG, GIF či SVG (tzv. Core Media Types) nepředstavuje v současné době zvláštní riziko. Pro potřeby dlouhodobé ochrany je nicméně vhodné veškerý obsah vložený do souboru identifikovat, zapsat při uložení do administrativních metadat v rámci archivního balíčku a zobrazitelnost veškerých vložených formátů pravidelně ověřovat.

Riziko naopak představuje možnost vložení obsahu ve formátech mimo specifikací EPUB explicitně podporované formáty. Obsah cizích formátů je ve formátu povolen za podmínky alternativního vyjádření jejich obsahu v některém z podporovaných formátů. EPUB tedy může teoreticky obsahovat například vložený soubor ve formátu JPEG2000, v takovém případě ale musí zároveň obsahovat záložní verzi například ve formátu JPEG, aby bylo zajištěno správné zobrazení na jakémkoliv čtecím zařízení. Výskyt nepodporovaných souborových formátů bez alternativního vyjádření validační nástroje našťastí dokážou detekovat. Existuje nicméně možnost ztráty informačního obsahu a také autenticity dokumentu též v případě, kdy by tvůrce vložil do dokumentu nestandardní obsah, například video soubor, jehož záložním vyjádřením může být pouze statický bitmapový soubor. [94, s. 28] [87, s. 6]

Textový obsah publikace v EPUB 2 může být ve formátu XHTML 1.1 nebo alternativně ve formátu DTBook (DAISY Digital Talking Book). Vzhledem k tomu, že v novější verzi formátu EPUB 3 není nadále formát DTBook podporován, a rovněž podpora v zobrazovacích nebo migračních nástrojích klesá, představuje přítomnost DTBook formátu v EPUB určité riziko [99]. Formát DTBook je nicméně založený na XML, vzhledem k dostupnosti dokumentace by tedy přítomnost formátu v balíčku neměla představovat zásadní problém [94, s. 8]. Přesto je potřeba zvážit, zda přítomnost DTBook v souboru pouze identifikovat a uvést do metadat archivního balíčku, nebo rovnou před vstupem do dlouhodobého úložiště provést normalizaci a předejít tak případným budoucím komplikacím.

Specifikace formátu nedefinuje užití konkrétního DRM schématu, podporuje všechna běžně užívaná DRM schémata. Nejčastěji je možné se setkat s DRM od Amazonu, FairPlay od Apple nebo ADEPT od Adobe [87, s. 6-7]. Rizika, která přítomnost DRM ochrany u e-publikací představuje pro dlouhodobé uchování, byla rozebrána v kapitole 5. Přítomnost DRM je nutné identifikovat a v souladu s formátovou politikou instituce řešit buď odmítnutím, normalizací na nechráněný formát, nebo podmíněným uložením s vědomím rizika pro uchování. Přítomnost DRM ochrany dokáže odhalit například kontrola validačním nástrojem EPUBCheck.



Obrázek 2. Grafické vyjádření vztahů mezi formáty e-knih, centrováno na EPUB [6, s. 3]

## 6.4 EPUB 3

Název souborového formátu	EPUB 3
MIME type	application/epub+zip
PUID	fmt/483
Přípona souboru	.epub
Tvůrce formátu	IDPF, ISO
Datum publikování formátu	11.10.2011
Předcházející formát	EPUB 2.0.1
Nástupnický formát	-

Tabulka 8. Přehled vybraných údajů o formátu EPUB 3

### 6.4.1 Historie formátu

V roce 2010 zahájila organizace International Digital Publishing Forum (IDPF) práce na nové verzi formátu EPUB, která byla jako verze 3.0 vydána v říjnu 2011. Formát byl poté dvakrát aktualizován, poprvé na verzi 3.0.1 v červnu 2014 a podruhé v lednu 2017 na současnou verzi 3.1 [100]. V roce 2014 byla specifikace formátu publikována jako mezinárodní ISO standard ISO/IEC TS 30135 (části 1-7) [101].

### 6.4.2 Popis formátu

Zatímco EPUB 2 byl definován 3 samostatnými specifikacemi, v případě EPUB 3 už je těchto samostatných specifikací 7 [102]:

- EPUB 3 Overview (také jako ISO/IEC TS 30135-1:2014)
- EPUB Publications 3.0 (také jako ISO/IEC TS 30135-2:2014)
- EPUB Content Documents 3.0 (také jako ISO/IEC TS 30135-3:2014)
- EPUB Open Container Format 3.0 (také jako ISO/IEC TS 30135-4:2014)
- EPUB Media Overlays 3.0 (také jako ISO/IEC TS 30135-5:2014)
- EPUB Canonical Fragment Identifier Specification (také jako ISO/IEC TS 30135-6:2014)
- EPUB 3 Fixed Layout Documents (také jako ISO/IEC TS 30135-7:2014)

Formát je založen zejména na standardech XML 1.0, XHTML 5.0, CSS 2.1 (částečně CSS 3.0), formátu ZIP, metadatovém standardu Dublin Core a znakové sadě Unicode. Nově přidává také podporu vkládání fontů ve formátech Open Type Fonts (OTF) a Web Open Font Format (WOFF), podporu JavaScriptu, technologii Text-to-Speech nebo podporu MathML. Do formátu je možné vkládat bitmapové obrazy ve formátech JPG, PNG nebo GIF, vektorovou grafiku ve formátu SVG a nově také zvukové soubory ve formátech MP3 a AAC LC (MP4).

Oproti EPUB 2 chybí podpora formátu DTBook [94, s. 8]. Formátová specifikace nedefinuje žádný konkrétní formát pro přehrávání videa, přehrávání videa je nicméně nativně podporováno již použitým formátem HTML 5 a s jeho přehráváním EPUB 3 počítá. [103]

Oproti EPUB 2 umožňuje EPUB 3 také fixní sazbu textu a obecně náročnější grafickou úpravu publikací. Podpora CSS modulu Media Queries umožňuje navíc definovat různou, ale předem danou, podobu vzhledu stránky v závislosti na velikosti displeje zobrazovacího zařízení [56, s. 17]. S pokročilými možnostmi formátování se tak formát EPUB 3 stává plně konkurenceschopný formátu PDF i v oblasti odborné či dětské literatury.

### **6.4.3 Odvozené formáty**

Již krátce poté, co byl formát publikován, došlo k jeho adopci ze strany velkých publikačních platforem, zejména Applu a Amazonu. Apple, který se na vývoji EPUB 3 ostatně aktivně podílel, přišel již v roce 2012 s odvozeným proprietárním formátem iBooks. iBooks prakticky kopíruje formát EPUB 3, přináší však několik nezdokumentovaných rozšíření a modifikací, které brání využití formátu na jiných zařízeních než v aplikaci iBooks. O formátu je blíže pojednáno v samostatné kapitole. Podobně Amazon přišel začátkem roku 2012 s formátem KF8<sup>40</sup>, který vychází z EPUB 3, rozšířeného o DRM schéma od Amazonu [81]. Formát KF8 z hlediska českého trhu významný není, nebude tedy v této práci podrobně popisován.

### **6.4.4 Zhodnocení formátu z hlediska dlouhodobého uložení**

Při zhodnocení verze formátu EPUB 3 jsem vycházel v první řadě z hodnocení formátu od Britské národní knihovny z roku 2015 [87], hodnocení od americké Kongresové knihovny [104], studie Paula Wheatleyho [95] pro Harvardskou knihovnu z roku 2014, studií Johana van der Knijffa pro nizozemskou Královskou knihovnu z let 2012 [94], 2013 [105] a 2015 [99] a dalších zdrojů, zabývajících se dílčími otázkami uchovávání EPUB 3, uvedených dále v textu. Hodnocení formátu se v obecné míře shodují na relativní bezpečnosti uchovávání dokumentů v EPUB 3, vzhledem k větší komplexnosti oproti verzi EPUB 2 nabádají nicméně k větší obezřetnosti. Týká se to zejména některých z nových funkcí, které formát oproti svému předchůdci přináší. Hlavními novinkami jsou podpora vloženého, nebo externě referencovaného multimediálního obsahu, JavaScriptu nebo např. obfuskace fontů.

---

<sup>40</sup> Známý též pod názvem AZW3.

Rozšíření podporovaných vložených formátů o audio formáty MP3 a AAC LC v kontejneru MP4 přidává formátu na komplexitě. Budoucí zobrazení obsahu dokumentu bude závislé na schopnosti spustit vložený multimediální obsah a udržovat nástroje nutné pro jeho úspěšné přehrání. Přítomnost vloženého multimediálního obsahu je tedy přinejmenším nutné identifikovat a status jednotlivých vložených formátů pravidelně ověřovat. [94, s. 27]

Jednou z novinek ve verzi 3 je také možnost referencování multimediálního obsahu, uloženého mimo vlastní soubor EPUB, zpravidla online. Tyto tzv. *Remote Resources* jsou poté uvnitř samotného dokumentu reprezentovány např. pouze obrázkem. V případě nedostupnosti odkazovaného tedy hrozí ztráta informačního obsahu a autenticity v budoucnu zobrazených souborů. [87, s. 6] [94, s. 27-28]

Za rizikovou lze považovat také přidanou podporu JavaScriptu. Vedle určitých bezpečnostních rizik, která jsou s JavaScriptem spojená [106], je problematická zejména různorodá podpora tohoto skriptovacího jazyka v prohlížečích. Není tedy garantována autentická podoba zobrazení dokumentu. [56, s. 32-33] [94, s. 28]

Formát i nadále podporuje všechna běžně užívaná DRM schémata, jejich přítomnost je tedy nutné detekovat a v souladu s politikou instituce na jejich přítomnost reagovat. Novinkou je v případě EPUB 3 možnost šifrování vložených fontů. Tzv. obfuskace fontů (*font obfuscation*) chrání vložené licencované fonty jejich částečným zašifrováním. Tato ochrana může nicméně vést ke ztrátě fontu v případě budoucí migrace na jiný formát [94, s. 26]. Obfuskaci fontů dokáže odhalit například validační nástroj EPUBCheck [99], měla by tedy být posuzována stejně jako přítomnost jiných DRM omezení.

Podobně jako v případě předchozí verze formátu EPUB 2 se jedná o otevřený formát s dostupnou dokumentací a bez licenčních omezení. Oproti EPUB 2 byla nová verze formátu navíc podrobena revizi technické komise Mezinárodní organizace pro normalizaci a vydána jako mezinárodní technická norma ISO/IEC TS 30135 [101].

Podpora softwarových aplikací pro zobrazení, validace nebo migrace mezi formátem EPUB 3 a jinými formáty je stále o něco nižší než u předchozího formátu, ale obecně ji lze hodnotit jako dostačující. K identifikaci souborového formátu lze využít nástroje DROID či FIDO, identifikace nicméně neumožňuje v současné době rozlišit mezi verzemi EPUB 2 a EPUB 3. Pro validaci a charakterizaci lze využít nástroj EPUBCheck. EPUBCheck lze využít také k identifikaci konkrétní verze formátu. Pro převody do formátu EPUB 3 je možné využít například open-source nástroj Calibre.

## 6.5 iBooks

Název souborového formátu	iBooks
MIME type	application/x-ibooks+zip
PUID	fmt/482
Přípona souboru	.ibooks
Tvůrce formátu	Apple Inc.
Datum publikování formátu	2012
Předcházející formát	EPUB 3
Nástupnický formát	-

Tabulka 9. Přehled vybraných údajů o formátu iBooks

### 6.5.1 Historie formátu

V lednu roku 2010 představila společnost Apple Inc. tablet iPad spolu s aplikací iBooks pro čtení e-knih, pracující s formátem EPUB a PDF [107]. V prosinci 2010 vyšla aplikace iBooks 1.2, která již pracovala s novým formátem nazvaným Apple Fixed Layout EPUB. Jednalo se o formát EPUB 2 rozšířený o několik proprietárních modifikací, zejména podporu pevné sazby textu nebo JavaScriptu [56, s. 18-20]. V lednu 2012 Apple představil aplikaci iBooks Author, určenou pro autory a nakladatele pro vytváření elektronického obsahu, která pracovala již se zcela novým formátem odvozeným od EPUB 3 a nazývaným iBooks [56, s. 20].

### 6.5.2 Popis formátu

Formát iBooks se stal poměrně oblíbeným nejen díky zajištěnému odbytu u držitelů zařízení Apple, ale také díky tomu, že aplikace iBooks Author jako první umožnila naplno využívat nových funkcí standardu EPUB 3. Autorům a nakladatelům například jako vůbec první poskytla uživatelsky přívětivý WYSIWYG editor pro interaktivní publikace [108].

Formát vychází z formátu EPUB 3, obsahuje však několik rozšíření a modifikací, které zamezují jeho užití mimo čtecí aplikaci iBooks. Předně, soubory samy sebe neidentifikují jako `application/epub+zip` ale jako `application/x-ibooks+zip`. Využívá nezdokumentované proprietární rozšíření standardu CSS spravovaného W3C [109] [107] [110] a proprietární Apple XML jmenné prostory [107] [110]. Celková dokumentace formátu není publikována<sup>41</sup>. Publikace v iBooks mohou kromě vložených formátů podporovaných specifikací EPUB 3 obsahovat prakticky jakýkoliv souborový formát, který je podporovaný multimediálním

---

<sup>41</sup> Srov. s [197].

přehrávačem QuickTime od Apple [111] [112]. Soubory iBooks jsou distribuovány v šifrované podobě a jsou chráněny DRM schématem FairPlay od Apple.

### **6.5.3 Zhodnocení formátu z hlediska dlouhodobého uložení**

Praktické dopady proprietárních rozšíření od Apple jsou ty, že formát není zpětně kompatibilní s formátem EPUB 3 a bez ztráty části informačního obsahu není možné provést zpětnou konverzi z iBooks do EPUB. Při převodu zůstane zachován text a struktura dokumentu, jsou ale narušeny ostatní vlastnosti, jako je formátování, obsah nebo titulní strana. [109] [107]

Další odchylky od EPUB 3 byly popsány v případě práce s interaktivními prvky. iBooks umožňuje do publikací vkládat řadu miniaplikací (tzv. widgets), které umožňují do knih umisťovat různé interaktivní doplňky, jako jsou galerie, multimédia, kvízy, 3D objekty apod. [113]. Podobné funkcionality umí též EPUB 3 za využití JavaScriptu, iBooks ale využívá vlastní proprietární řešení a maří tak možnou zpětnou kompatibilitu mezi iBooks a EPUB. [114]

Vytvoření konverzních nástrojů z iBooks do formátů nezávislých na platformě Apple by bylo možné pouze za předpokladu zpřístupnění dokumentace formátu, nebo velmi pracnou metodou reverzního inženýrství. Částečně pozitivní je alespoň skutečnost, že od verze 2.3 z června 2015 umožňuje iBooks Author export ve formátu EPUB [115] [116]. Tato funkce může pomoci autorům vyexportovat jejich knihy z aplikace iBooks Author v otevřeném formátu, který bude vhodný pro dlouhodobé uložení. Přetrvávající problém je nicméně zejména na straně interaktivních miniaplikací, u nichž není v aplikaci iBooks Author pro všechny typy těchto miniaplikací možnost exportu v profilu EPUB [116].

Formát iBooks je proprietární formát bez zveřejněné dokumentace, který lze číst pouze v rámci ekosystému firmy Apple. Soubory mohou obsahovat šifrování ve formě FairPlay DRM. Vzhledem k nedostupné dokumentaci a uzavřenému ekosystému Apple nejsou dostupné ani žádné validační nástroje. Soubory ve formátu iBooks nelze považovat za bezpečné z hlediska dlouhodobého uchování. Pokud je to možné, měly by být online publikace publikované výhradně prostřednictvím iBooks Store získávány přímo od nakladatele v jiných formátech.

## 6.6 Wooky

Název souborového formátu	Wooky
MIME type	-
PUID	-
Přípona souboru	.book, .magazine
Tvůrce formátu	Wooky, a.s.
Datum publikování formátu	2010
Předcházející formát	-
Nástupnický formát	-

Tabulka 10. Přehled vybraných údajů o formátu Wooky

### 6.6.1 Historie formátu

Společnost Wooky vznikla v roce 2010, kdy spustila svoji distribuční platformu provázanou s čtecími aplikacemi a později též hardwarovými čtečkami Wooky [117]. Na český trh tak přenesla model uzavřeného ekosystému, jaký v zahraničí úspěšně využívají například Amazon nebo Apple. Podobně jako v případě obou zahraničních prodejců se jádrem ekosystému stal proprietární formát obecně nazývaný prostě Wooky a stejně proprietární DRM schéma Wooky.

### 6.6.2 Popis formátu

Formát Wooky nemá dostupnou dokumentaci, veškeré poznatky k formátu tak lze získat převážně pouze metodou reverzního inženýrství. Dokumenty jsou šířeny ve formě šifrovaného binárního souboru s koncovkou .book nebo .magazine. V případě souborů s koncovkou .book se jedná o souborový formát odvozený od standardu EPUB [118]. V případě souborů .magazine, určeného zejména pro šíření periodických publikací s pevnou sazbou textu a náročnější grafickou úpravou, se pravděpodobně jedná o formát odvozený od PDF, jde ale spíše o spekulaci. Lze předpokládat, že podobně jako v případě Amazonu a Applu, jejichž model společnost Wooky převzala, bude hlavním rozšířením formátu přítomnost šifrování dle proprietárního schématu Wooky DRM. Samotná přítomnost proprietárního šifrování dostatečně připoutává čtenáře k dané distribuční platformě. Konkrétní změny mezi formátem Wooky a EPUB nelze bez možnosti ověřit nezašifrovanou podobu souboru posoudit.

### 6.6.3 Zhodnocení formátu z hlediska dlouhodobého uložení

Proprietárnost formátu, absence jakékoliv dokumentace a rovněž využití proprietárního Wooky DRM u všech publikací přinášejí zásadní rizika pro dlouhodobé uložení. Publikace v tomto



formátu by neměly být přijímány k dlouhodobému uložení. Z pohledu českého trhu s e-publikacemi je nicméně platforma Wooky významná a zejména díky obchodu Ebux.cz patří mezi vůbec největší distribuční platformy [23]. Produkci publikovanou v některých případech výhradně jen prostřednictvím Wooky nelze tedy jen tak jednoduše odepsat. Výhodou je odvození formátu od EPUB. Rovněž otevření se Wooky hardwarovým čtečkám Kindle [57] nebo PocketBook [119] naznačuje, že distributor publikace uchovává v některém z formátů, který lze bezpečně převádět do formátů srozumitelných těmto čtečkám. Možnou strategií k uchování digitálního obsahu publikovaného ve formátu Wooky by tedy mohlo být jednání s distributorem a zajištění dodávek e-publikací v alternativním, otevřeném formátu.

## 6.7 PDF

Název souborového formátu	PDF (Portable Document Format)
MIME type	application/pdf <sup>42</sup>
PUID	aktuálně registr PRONOM eviduje u PDF 30 různých PUID [120]
Přípona souboru	.pdf
Tvůrce formátu	Adobe Systems; ISO
Datum publikování formátu	1993 (PDF 1.0)
Předcházející formát	PostScript
Nástupnický formát	-

Tabulka 11. Přehled vybraných údajů o formátu PDF

### 6.7.1 Historie formátu

Formát PDF byl představen společností Adobe Systems v roce 1993. Specifikace PDF ve velké míře vycházela z programovacího jazyka PostScript. Nový formát měl zajistit zejména jednotné zobrazení dokumentů nezávisle na použitém software, hardware a operačním systému [121] [122, s. 221-222]. Do roku 2006 představila společnost Adobe Systems dalších 7 verzí formátu, poslední verze PDF 1.7 z roku 2006 pak byla v roce 2008 vydána také jako ISO standard ISO 32000-1:2008. Tím přešel další vývoj formátu do kompetence technické komise Mezinárodní organizace pro normalizaci (ISO), která v roce 2017 vydala také zatím poslední velkou aktualizaci formátu PDF 2.0 jako ISO 32000-2:2017.

Do rodiny formátu PDF v současné době náleží několik dalších odvozených standardů s vlastními názvy, každý určený pro specifické účely. Jsou to PDF/A pro archivaci, PDF/E pro technické obory, PDF/H pro zdravotnictví, PDF/UA pro zrakově postižené uživatele, PDF/X pro tisk a PDF/VT [123] pro variabilní a transakční tisk [122, s. 223] [121]. Celkově již existují desítky verzí formátu PDF, například formátový registr PRONOM [120] registruje 30 různých verzí PDF. Z hlediska dlouhodobé ochrany je důležitý zejména standard PDF/A, specificky určený k dlouhodobému uchování elektronických dokumentů. PDF/A existuje ve 3 základních verzích: PDF/A-1, PDF/A-2 a PDF/A-3. První verze je založena na PDF 1.4, další verze již na PDF 1.7. Všechny verze PDF/A byly vydány jako standardy ISO.

<sup>42</sup> Vyskytují se také tyto MIME typy: application/x-pdf; application/x-bzpdf; application/x-gzpdf; application/acrobat; application/vnd.pdf; text/pdf nebo text/x-pdf.

### 6.7.2 Popis formátu

Formát PDF byl od svého vzniku zamýšlen jako stránkově orientovaný formát, do kterého bude možné vkládat rozmanitý typ obsahu, zpočátku zejména text, fonty, bitmapové obrázky a vektorovou grafiku. Postupně formát přidával podporu dalších typů vložených objektů, například hypertextových odkazů, interaktivních formulářů, JavaScriptu, 3D grafiky, videa, zvuku a dalších [124, s. 21]. Formát je uzpůsoben k tomu, aby fungoval jako kontejner, do kterého lze vložit prakticky jakýkoliv binární obsah [124, s. 31]. Soubory PDF mohou sice existovat také jako textové soubory ve formátu 7-bit ASCII [125], formátová specifikace nicméně předpokládá jejich distribuci výhradně ve formě binárních souborů [126, s. 15].

Formát podporuje různé způsoby zabezpečení obsahu. Od verze 1.1 umožňuje ochranu heslem, jímž lze kromě samotného zobrazení dokumentu zamezit u souboru samostatně také např. vykopírování textu, editaci nebo tisku. Od verze 1.4 umožňuje vkládání viditelného průhledného vodoznaku, využitelného například jako součást sociálního DRM. Od verze 1.3 je podporován také elektronický podpis, který je možné využít k autentizaci oprávněného uživatele. Samozřejmou součástí formátu je také podpora DRM schémat od Adobe. [124, s. 31-32]

### 6.7.3 Standard PDF/A

Přestože se o PDF/A zpravidla hovoří jako o samostatném formátu, není to zcela přesné. Spíše se jedná o formátový profil PDF ve verzi 1.4 nebo 1.7 s takovými parametry a omezeními, které jsou vhodné pro dlouhodobé uchování. Smyslem PDF/A je umožnit zobrazení obsahu dokumentu zcela nezávisle na použitém hardware, software a operačním systému. Soubor PDF/A tak musí vždy obsahovat veškeré informace a data potřebná k vykreslení obsahu, například fonty. Oproti PDF má profil PDF/A také určitá omezení ohledně možnosti vloženého obsahu, nepodporuje například vkládání audia, videa, 3D objektů, JavaScriptu nebo šifrování [122, s. 228]. Veškerá omezení PDF/A oproti PDF směřují k nezávislosti na obsahu, nacházejícímu se mimo samotný soubor. Současně jsou vyloučeny veškeré technologie a formáty, u nichž technická komise ISO očekávala v dlouhodobější perspektivě rizika z hlediska dlouhodobého uchování a zobrazitelnosti obsahu v původní autentické podobě.

Profil PDF/A má v současné době 3 základní verze. Verze PDF/A-1 byla vydána v roce 2005 jako ISO 19005-1:2005, jedná se archivní profil PDF 1.4. PDF/A-2 odvozená od PDF 1.7 byla jako ISO 19005-2:2011 publikována v roce 2011 a konečně zatím poslední verze PDF/A-3, odvozená rovněž od PDF 1.7 byla jako ISO 19005-3:2012 publikována v roce 2012. Verze PDF/A-3 se od verze PDF/A-2 liší zejména možnostmi vložení libovolného dalšího souborového

formátu, přičemž však garantuje dlouhodobé zachování pouze vizuální reprezentace takto vloženého obsahu. V procesu schvalování je také verze PDF/A-4 (ISO 19005-4), vycházející z PDF 2.0. Vydání se očekává v roce 2018. [127, s. 4]. [122, s. 229] [128, s. 7-8]

Jednotlivé verze PDF/A se v závislosti na míře naplnění požadavků standardu dále člení na verze A, B a U. Zatímco verze B (basic) zaručují pouze přesnou vizuální reprodukci<sup>43</sup>, verze A (accessible) garantuje také zachování logické struktury textu a namapování textu na standard Unicode. Verze U (unicode) představuje kompromis mezi požadavky úrovně A a B, nepožaduje tak sice zachování logické struktury textu, zachovává nicméně požadavek na namapování textu na standard Unicode [129]. V současné době takto existuje celkem 8 dílčích verzí PDF/A označovaných jako PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2b, PDF/A-2u, PDF/A-3a, PDF/A-3b, PDF/A-3u. [128, s. 10]

#### **6.7.4 Zhodnocení formátu z hlediska dlouhodobého uložení**

Při zhodnocení formátu jsem vycházel zejména z hodnocení formátu Britské národní knihovny z roku 2015 [130], hodnocení PDF [131] a PDF/A [132] od americké Kongresové knihovny z roku 2017, zprávy DPC k využití PDF/A v digitální archivaci z roku 2017 [128], hodnocení PDF od Johana van der Knijffa pro nizozemskou Královskou knihovnu z roku 2009 [133] a dalších článků, zabývajících se dílčími problémy PDF, odkazovaných dále v textu.

##### **6.7.4.1 Standardizace a rozšířenost formátu**

PDF, jako v prostředí internetu celosvětově vůbec nejrozšířenější publikační formát [134], vykazuje na první pohled velmi dobrou podporu prohlížečích aplikací. Oficiálním prohlížečem PDF od Adobe je bezplatný, avšak proprietární, prohlížeč Adobe Reader<sup>44</sup>. K dispozici je také řada open-source prohlížečů PDF. Dokumentace formátu je veřejně dostupná, od roku 2008 jsou všechny nové verze formátu publikovány jako standardy ISO. Formálně se tedy v současné době jedná o otevřený formát s dostupnou dokumentací. Zároveň nicméně PDF i nadále využívá některé dílčí proprietární technologie vlastněné Adobe [135, s. 10].

Dostupnost specifikace na první pohled řeší mnoho otázek spojených s budoucí srozumitelností formátu. V případě PDF nicméně riziko představuje značná komplexnost formátu [136, s. 40]. Složitost a rozsah formátové specifikace představuje problém nejen z pohledu programátorů při

---

<sup>43</sup> Zachování vizuální podoby dokumentu znamená, že není garantována neporušenost textu při vykopírování textu ze souboru PDF – neporušeností se myslí jak struktura, tak kódování [122, s. 230].

<sup>44</sup> Nejnovější verze je distribuována pod názvem Adobe Acrobat Reader DC.

implementaci standardu v průběhu vytváření aplikací pro produkci PDF souborů, ale také pro vývoj skutečně efektivních nástrojů pro validaci souborů PDF. [136]

Ještě větším problémem je potom vágnost formátové specifikace u nejstarších verzí formátu PDF. Dokumentace formátu totiž sice poskytovala soubor pravidel pro vytváření PDF, v mnoha aspektech však ponechávala vývojářům přílišnou volnost při výkladu těchto pravidel a vedla tím k velké rozdílnosti mezi jednotlivými implementacemi formátu PDF. [137]

Nejednoznačnost formátové dokumentace se Adobe rozhodlo řešit pragmaticky tím, že se při vývoji vlastní referenční implementace PDF, Adobe Readeru, rozhodlo zohlednit veškeré možné variantní podoby souborů. Adobe Reader díky tomu dovede zobrazit i velmi nestandardní či nevalidní soubory PDF, které vykazují značné odchylky od specifikace<sup>45</sup>. Aplikace Adobe Reader se tak stala de facto nadřazenou publikované specifikaci formátu, neboť při absenci validačních nástrojů se zobrazitelnost v Adobe Readeru stala pro vývojáře nástrojů pro práci s PDF hlavním cílem. [130, s. 5] [137] [138, s. 201-202]

Při snaze dlouhodobě ochránit obsah ve formátu PDF tak dochází k nejistotě, zda pro ochranu a budoucí zobrazitelnost postačuje znalost zveřejněné dokumentace formátu, nebo zda bude zapotřebí znát také veškeré odchylky od standardu, které nyní podporuje proprietární software Adobe Reader a které naopak zdokumentovány nejsou. [138, s. 201-202] [137]

#### **6.7.4.2 Dostupnost validačních nástrojů**

V současnosti je prakticky jediným validačním nástrojem, který je schopný validovat formát PDF JHOVE, respektive jeho modul PDF-hul [139]. JHOVE deklaruje schopnost validovat verze PDF 1.0 až 1.6 a také některé verze PDF/X a PDF/A, jeho spolehlivost ale vzhledem k výše uvedeným problémům s rozsahem a kvalitou definice standardu není vždy stoprocentní. Kvalitou validace pomocí JHOVE se ve své studii zabývali například v poslední době Michelle Lindlar, Yvonne Tunnat a Carl Wilson [140]. Silnou stránkou nástroje PDF-hul je zejména silná uživatelská komunita a aktivní vývoj pod dohledem Open Preservation Coalition.

Výrazně lepší je situace ohledně validačních nástrojů pro validaci profilu PDF/A. K dispozici je například validátor Apache Preflight, součást javovské knihovny Apache PDFBox®, který deklaruje schopnost validovat PDF/A-1 na úrovni PDF/A-1b. Ještě slibnějším nástrojem se potom zdá být javovská knihovna veraPDF, určená k validaci všech verzí PDF/A a také

---

<sup>45</sup> Tato tolerance vůči chybám naopak často chybí u prohlížečích aplikací třetích stran. Tyto aplikace potom nejen nemusí správně zobrazit dokumenty odchylovající se od formátové specifikace, často také neumí správně vyrenderovat ani soubor, který ve skutečnosti specifikaci odpovídá. [130, s. 3]

charakterizaci PDF. VeraPDF byla vyvinuta v rámci evropského projektu PREFORMA<sup>46</sup> [141], realizovaného v letech 2014–2017, a v současnosti je stejně jako JHOVE spravována organizací Open Preservation Foundation [142].

#### **6.7.4.3 Komplexnost formátu**

Významné riziko pro uchování formátu PDF představuje prakticky neomezená podpora vkládání obsahu v jiných souborových formátech. Vložení jakéhokoliv dalšího souborového formátu prakticky, z pohledu dlouhodobé ochrany, znamená nutnost zajistit zobrazitelnost takového vloženého formátu v budoucnu. Největší rizika jsou zpravidla spojována s vkládáním multimediálního audio, video nebo 3D obsahu, případně s využitím javascriptových prvků. Podpora tohoto typu obsahu je proto vyloučena z archivních profilů PDF/A a jejich přítomnost je důležité sledovat, nebo případně zcela omezit, také u ostatních verzí PDF. [130]

#### **6.7.4.4 Ostatní rizika**

Riziko pro dlouhodobé uchování může dále představovat přítomnost technických ochranných prostředků, jako je šifrování nebo DRM. Problém může v dlouhodobé perspektivě představovat též závislost na externím obsahu, který je ze souboru pouze linkován a bez něhož v budoucnu nebude možné obsah v úplnosti zobrazit [130, s. 7]. Opomenout nelze ani riziko přítomnosti malware [143], pokud tedy politika příjemce souboru PDF povoluje přítomnost JavaScriptu v souboru, je nutné se zaměřit také na zapojení relevantních antivirových nástrojů, které případnou hrozbu budou schopny zachytit. Prakticky všechna zmíněná rizika jsou ošetřena profilem PDF/A, který jejich vzniku předchází, nebo je alespoň výrazně minimalizuje.

#### **6.7.4.5 Konverze PDF do PDF/A jako řešení pro dlouhodobou archivaci?**

Mezi nejrozšířenější nástroje pro konverzi z formátu PDF do formátu PDF/A patří open-source nástroj Ghostscript a komerční Adobe Distiller nebo Adobe Acrobat Professional [144, s. 5]. Ke konverzím z PDF do PDF/A je nicméně nutné, stejně jako k jakékoliv jiné formátové migraci, přistupovat obezřetně. Jako optimální lze hodnotit situaci, kdy PDF/A je buď vytvořeno z původního produkčního formátu nakladatele, nebo je přímo vyexportováno z aplikace pro sazbu dokumentů [128, s. 12-13]. V případě provádění konverze do PDF/A z již dříve vytvořeného PDF souboru může dojít k několika problémům. Z již jednou vytvořeného souboru PDF například zpravidla není možné vytvořit PDF/A-a a může tedy dojít ke ztrátě

---

<sup>46</sup> PREFORMA – PREservation FORMAts for culture information.

logické struktury textu. Získané PDF/A-b pak sice zachová vizuální podobu původního dokumentu vhodnou pro čtení nebo tisk, indexace textu nebo jeho srozumitelnost pro čtenáře odkázaného např. na strojové čtení takového dokumentu může však utrpět. [144, s. 5-6]

Další rizika představuje možnost ztráty obsahu, nepodporovaného profilem PDF/A – zejména vloženého multimediálního obsahu, případně interaktivních funkcí JavaScriptu. PDF soubory obsahující takový obsah tedy není do PDF/A vůbec vhodné převádět. Vůbec nejčastěji uváděnou komplikací při převodu do PDF/A je ale nedostupnost fontů, které je nutné vložit přímo do souboru [145]. V případě převodu z PDF totiž v systému nemusí být potřebné fonty k dispozici. Licenční politika některých fontů navíc může jejich vložení do souboru znemožnit i v případě jejich dostupnosti. V takovém případě, bez nahrazení fontu, není možné PDF/A vůbec vytvořit. Proces dodatečného vložení nebo náhrady fontu jiným podobným fontem je potom nejen pracný, ale může vést také ke ztrátě autenticity, a v případě nevhodně zvoleného náhradního fontu dokonce i ztrátě informačního obsahu [146]. [128, s. 13-15]

Zcela samostatně je třeba posuzovat využití verze PDF/A-3. Možnost vložení jakéhokoliv obsahu, s garancí zachování statické vizuální reprezentace vloženého obsahu, činí z profilu PDF/A-3 z pohledu dlouhodobé archivace poměrně kontroverzní formát. V současné době proto není archivní komunitou PDF/A-3 obecně přijímán jako formát vhodný pro příjem dokumentů [127, s. 18-20]. Jeho doporučené využití se zpravidla omezuje na tzv. hybridní archivaci, kdy jsou jeho vlastnosti využívány při konverzích dokumentů, prováděných v průběhu ingestu již v režii repozitáře [127, s. 7]. Dokument v nepodporovaném textovém formátu (např. DOC) tak může být převeden do PDF/A-3, přičemž je zajištěno uchování jeho vizuální podoby, a zároveň je původní nepozměněný dokument vložen přímo do kontejneru PDF/A<sup>47</sup>. Případný příjem PDF/A-3 dokumentů by tedy měl být nejlépe omezen pouze na takové dokumenty, které jsou kompatibilní s požadavky profilu PDF/A-2. [127]

Problematika konverzí do PDF/A není zcela jednoznačná a klade před instituce přijímající elektronické dokumenty několik důležitých rozhodnutí. V první řadě se jedná o otázku, jakou roli (pokud nějakou) má PDF/A hrát v životním cyklu přijímaných dokumentů. Bude instituce archivovat pouze PDF/A, nebo i ostatní verze PDF? Pokud pouze PDF/A, bude povinností vkladatele provést konverzi do PDF/A, nebo bude konverzi do PDF/A provádět teprve depozitní instituce a vkladatel naopak bude mít povinnost dokument odevzdat v původní autentické podobě? Jako optimální varianta se jeví motivovat nakladatele k odevzdávání

---

<sup>47</sup> Srov. s [130, s. 7].

PDF/A pouze tehdy, kdy jsou schopni je přímo získat ze svých produkčních systémů. V případě, kdy by však byli nuceni odevzdávat výhradně PDF/A, narůstá riziko neobdobných konverzí do PDF/A, u nichž nebude možné ověřit autenticitu ani úplnost informačního obsahu.

### **6.8 Dílčí závěry k 6. kapitole**

Jako formáty převážně vhodné pro dlouhodobou archivaci se potvrdily otevřené formáty EPUB 2, EPUB 3 a PDF/A (ve verzích PDF/A-1 a PDF/A-2). Jako poměrně riziková se ukazuje nearchivní varianta PDF a jako nevhodný se potvrdil zastarávající formát MOBI s pouze částečně dostupnou dokumentací. Jako zcela nevhodné se ukazují plně proprietární formáty iBooks a Wooky. Ani formáty EPUB a PDF/A nicméně nevyšly z hodnocení plně bez rizik. Za slabinu formátu PDF/A lze považovat, že z pohledu podporovaného obsahu nepředstavuje zcela plnohodnotnou alternativu pro nearchivní PDF. Jen část obsahu, který je nyní produkován ve formátu PDF, by bylo možné produkovat ve formátu PDF/A, nebo jej do něj alespoň normalizovat. Nejviditelnějším příkladem jsou multimediální a interaktivní publikace. Dílčí rizika spojená s formátem EPUB, opět například možnost užití JavaScriptu nebo multimédií, je možné teoreticky řešit definicí povolených vlastností, v rámci formátové politiky repozitáře.



## 7 Případové studie vybraných zahraničních knihoven

### 7.1 Cíle a metoda

V případových studiích v této kapitole se zaměřuji především na otázky, které přinesly předchozí kapitoly této práce. Zejména se věnuji následujícím aspektům formátových politik:

- Preferované a akceptované formáty<sup>48</sup>
- Povolený obsah přijímaných dokumentů, zejména přítomnost interaktivního a multimediálního obsahu ve formátech EPUB 3 a PDF
- Přítomnost DRM u e-publikací (včetně hesla)
- Postoj k normalizacím na straně nakladatele / knihovny (otázka autenticity)
- Používané nástroje pro identifikaci / charakterizaci / validaci

Ke zpracování případových studií jsem využil údajů z literatury, informací uvedených na webových stránkách knihoven, e-mailové komunikace se zástupci zkoumaných institucí a případně také relevantních legislativních předpisů.

Výběr knihoven jsem učinil na základě předběžné rešerše literatury a informací na webových stránkách knihoven, která určila sedm institucí s dostatečným množstvím relevantních zdrojů v anglickém jazyce. Následně jsem tyto instituce oslovil prostřednictvím e-mailů s žádostí o poskytnutí doplňujících informací. Ze sedmi oslovených institucí se mi podařilo získat potřebné doplňující informace od šesti z nich. Jedinou knihovnu, od níž se mi nepodařilo dodatečné informace získat, francouzskou Národní knihovnu, jsem se nakonec rozhodl zpracovat na základě dostupné literatury také, neboť k problematice přináší nové přístupy a informace.

### 7.2 Tematicky blízké výzkumy

Průzkumem částečně volně navazuji na dotazníkové šetření Fredericka Zarndta, Dorothy Carner a Edwarda McCaina z léta 2017 *Digital e-Legal Deposit Survey (2017)* [147]. Týmu se podařilo shromáždit odpovědi od celkem 19 národních knihoven, z nichž 16 v dotazníku uvedlo, že elektronické publikace přijímají, z toho 2 instituce na dobrovolné bázi [147, s. 10-11]. S tématem bakalářské práce souvisela zejména otázka na přijímané formáty e-publikací<sup>49</sup>. Odpovědi respondentů ukazují tabulka 12 a tabulka 13.

---

<sup>48</sup> Jako preferované formáty jsou označovány formáty vhodné k dlouhodobému uchování. U akceptovaných formátů je očekávána dřívější potřeba formátové migrace, ani u nich však není zpravidla potřeba provádět při odevzdání preventivní normalizaci do preferovaných formátů. [193]

<sup>49</sup> Otázka zněla: „*In what format(s) does your library accept digital works?*“ [147, s. 17]

volba	EPUB	PDF	MOBI	TIFF	JPEG	Open Doc	Jiné
počet	12	14	3	9	8	2	9

Tabulka 12. Digital e-Legal Deposit Survey – akceptované formáty e-publikací [147, s. 17-18]

Příklady odpovědí od respondentů, kteří zaškrtnuli volbu *Jiné*:

„PDF 2000 goes for the e legal deposit newspapers“.
„We are reasonably format neutral, and will accept all formats“.
„Word; (as annex files we accept also other formats like video, sound and picture files)“
„XML (e.g. JATS/NLM - DTD), HTML“

Tabulka 13. Digital e-Legal Deposit Survey – další akceptované formáty e-publikací [147, s. 17-18]

Průzkum *Digital e-Legal Deposit Survey* nebyl specificky zaměřený na textové dokumenty, zároveň nerozlišoval například mezi verzemi PDF a PDF/A, přesto přinesl určitý základní obrázek přístupů národních institucí. Do současné doby se jedná zřejmě o jediný takto rozsáhlý průzkum, který se formátovými politikami v souvislosti s e-depositem vůbec zabýval.

### 7.3 Deutsche Nationalbibliothek

Německá národní knihovna (Deutsche Nationalbibliothek, DNB) je příjemcem elektronického povinného výtisku od roku 2008, kdy bylo vydáno *Nariadení o povinném odevzdávání ze 17. 10. 2008* (BGBl. I S. 2013) [148]. Již v roce 2006 byl vydán také *Zákon o Německé národní knihovně (DNBG)*<sup>50</sup> z 22. 6. 2006 (BGBl. I S. 1338) [149], který knihovně předepisuje mimo jiné sbírat, skladovat, trvale zabezpečovat a připravovat pro zpřístupnění veřejnosti mediální díla bez rozlišení, zda vyšla na hmotném nosiči, nebo byla šířena nehmotně. Díla musí být dodána v bezvadné podobě a ve stavu vhodném pro jeho dlouhodobé uložení. [149] [150]

Nariadení o povinném odevzdávání umožňuje DNB odmítnout publikaci, pokud není její uchování ve veřejném zájmu. Od dodávky díla může knihovna upustit také v případech, kdy technické postupy neumožňují bez velkého úsilí, ať již sběr, nebo dlouhodobé uložení díla. Vydavatel je povinen odevzdat dílo přesně v takové podobě, v jaké je komerčně distribuováno. V případě vydání díla ve více formátech může knihovna upustit od povinnosti doručení díla v dalších kopiích. Vydavatel je dále povinen dodat spolu s dílem softwarové nástroje nezbytné pro užití díla nebo k vytvoření archivní kopie. [148]

Online publikace jsou tedy přijímány výhradně ve formátech, v jakých jsou publikovány. Akceptovány jsou v současné době formáty PDF a EPUB ve všech verzích, přijímány jsou také

<sup>50</sup> České překlady názvů zákonů převzaty z článku Zdeňka Matusíka [192].

multimediální a interaktivní publikace, pokud jsou v jednom z těchto formátů<sup>51</sup>. Z odevzdání jsou prozatím naopak vyloučeny e-knihy ve formátech rodiny Kindle. DNB formát MOBI nepovažuje za vhodný pro dlouhodobou archivaci, vzhledem k jeho uzavřenosti a využití proprietárního DRM schématu [151]. Od srpna 2011 DNB uplatňuje u formátu MOBI výjimku a umožňuje, aby v případech, kdy je autor sám vydavatelem a držitelem distribučních práv k dílu, mohl takové dílo dobrovolně odevzdat převedené do formátu EPUB [152] [151]. Z důvodu zachycení alespoň části produkce, která je publikována výhradně na platformě Amazonu tak v tomto případě DNB upouští od jinak důsledně vyžadovaného dodání díla v autentické podobě, jak je dílo běžně dostupné na trhu. Odevzdávané publikace nesmí obsahovat žádné ochranné prostředky proti čtení nebo kopírování. V případě vydání ve více formátech jsou preferovanými formáty PDF/A a ostatní verze PDF. [152]

Strategií a povinností<sup>52</sup> DNB je zachovat, pokud možno, nepozměněné originální dokumenty bez použití normalizace v průběhu ingestu do dlouhodobého úložiště [60, s. 230]. V průběhu ingestu jsou dokumenty hodnoceny a rozřazeny do 5 úrovní (0 až 4), podle stupně rizika pro dlouhodobé uchování. Hodnocení probíhá na základě posouzení 5 kritérií [153]:

Integrita dat je neporušena (kontrola hash, např. MD5)	ingest level 0
Souborový formát identifikován, patří mezi akceptované formáty (identifikace)	ingest level 1
Nejsou přítomná žádná technická omezení (např. DRM)	ingest level 2
Podářilo se extrahovat specifická technická metadata (charakterizace)	ingest level 3
Soubor je validní vůči formátové specifikaci (validace)	ingest level 4

Tabulka 14. Hodnocení rizik e-publikací dle DNB [153]

V případě, že je zjištěna porušenost dat po přenosu, je soubor ohodnocen 0. Proběhne-li úspěšně identifikace souborového formátu, soubor je označen 1 atd. Úspěšně identifikovaný souborový formát bez technických omezení, u něhož došlo k extrakci požadovaných technických metadat i validaci, dosahuje nejvyšší, tedy 4. úrovně. Formátová politika DNB [154] určuje pro jednotlivé formáty minimální úroveň podmiňující úspěšný ingest do úložiště. V případě formátů EPUB a PDF je podmínkou pro uložení minimálně 2. úroveň [154, s. 7]. Tato formátová politika je nastavena zejména na základě přesvědčení, že je-li, byť formálně nevalidní dokument prostý ochranných opatření, může být teoreticky transformován a

<sup>51</sup> Zjištěno z e-mailové korespondence s Cornelií Diebel z oddělení informační infrastruktury a uchování Deutsche Nationalbibliothek.

<sup>52</sup> Tamtéž.

vytvořena z něj bezchybná kopie, která bude dlouhodobě uchovatelná. V případě přítomnosti DRM jsou tyto transformace prakticky nemožné [155, s. 11].

V případě, že soubor dosáhne úrovně 0 nebo 1, je odmítnut a původce souboru je vyzván k dodání bezchybné verze. V případě, kdy soubor dosáhne úrovně 2–4, může být dokument přijat k dlouhodobé archivaci a DNB přejímá zodpovědnost za jeho dlouhodobou ochranu. Také v případě zjištění nevalidních souborů se DNB pokouší o vyjednání bezchybné kopie, pokud ale validní soubor není dodán, uchovává DNB soubor, který splnil požadovanou minimální úroveň 2, odpovídající dodávce souboru bez DRM<sup>53</sup>. Požadavky na dodání bezchybných kopií ze strany DNB v některých případech již motivovaly nakladatele k tomu, aby sami integrovali validační nástroje do svých publikačních systémů. [153, s. 2-4]

Pro identifikaci, charakterizaci a validaci používá DNB nástroj FITS (The File Information Tool Set), vyvinutý Harvardskou knihovnou, který pro svou práci využívá řadu zapojených nástrojů – např. Apache Tika, DROID nebo JHOVE. Pro charakterizaci a validaci souborů ve formátu EPUB využívá DNB vlastní vyvinutou aplikaci<sup>54</sup>. [153, s. 3-4]

nástroj	formát	typ služby
FITS	PDF, EPUB	Identifikace
FITS	PDF	Charakterizace / validace
Vlastní nástroj	EPUB	Charakterizace / validace

Tabulka 15. Seznam nástrojů využívaných DNB pro práci s přijímanými e-publikacemi

Zajímavé údaje přináší publikované informace o kvalitě přijatých dat za období 12/2012–4/2013 (tabulka 16) a 1/2013–12/2013 (tabulka 17). Statistika ukazuje zcela minimální zastoupení dokumentů s DRM (0,01 %) a velký rozdíl mezi validitou odevzdávaných souborů ve formátech PDF (94,90 % – 97,75 %) a EPUB (64,68 % – 76,44 %). Jedním z možných vysvětlení je nicméně kvalita dostupných validačních nástrojů pro PDF. Patrné je také zlepšení poměru validních a nevalidních souborů EPUB v obou vzorcích. [153, s. 4] [155, s. 10-11]

formát	Level 0	Level 1	Level 2	Level 3	Level 4
PDF	0	12 / 0,01 %	2 890 / 2,49 %	3 014 / 2,59 %	110 222 / 94,90 %
EPUB	0	0	0	4 654 / 34,31 %	8 908 / 65,68 %

Tabulka 16. Struktura kvality EPUB a PDF odevzdaných do DNB mezi 12/2012 a 4/2013 [153, s. 4]

<sup>53</sup> Zjištěno z e-mailové korespondence s Cornelií Diebel z oddělení informační infrastruktury a uchovávání Deutsche Nationalbibliothek.

<sup>54</sup> Tamtéž, srov. také s [153, s. 4].

formát	Level 0	Level 1	Level 2	Level 3	Level 4
PDF	0	128 / 0,01 %	23 648 / 1,73 %	6 807 / 0,50 %	1 332 604 / 97,75 %
EPUB	0	0	0	33 648 / 23,56 %	109 184 / 76,44 %

Tabulka 17. Struktura kvality EPUB a PDF odevzdaných do DNB v roce 2013 [155, s. 10-11]

#### 7.4 Bibliothèque nationale de France

Historie elektronického povinného výtisku ve Francii se datuje do roku 2006. Až do roku 2011 nicméně francouzská Národní knihovna (Bibliothèque nationale de France, BnF) mohla provádět pouze webové sklizně volně vystaveného webového obsahu (včetně e-publikací), práce na systému pro příjem elektronických publikací tak započaly až v roce 2012 [156, s. 4].

Předchozí zkušenosti s web-harvestingem ukázaly na možné problémy se zachycením části produkce e-publikací, které nejsou přímo uloženy na webu prodejce, ale zasílány zákazníkovi po uskutečnění nákupu jinými prostředky (např. e-mailem). Poskytnutí přístupových údajů k webu, na němž požadovaný obsah není přítomný, by tento problém neřešil. Pracovní skupina za účasti BnF a Syndicat National de l'Édition (SNE, sdružení francouzských nakladatelů) proto nakonec od metody web-harvestingu ustoupila a přiklonila se k povinnému odevzdávání e-publikací ze strany nakladatelů prostřednictvím vyhrazeného FTP serveru. [156, s. 4-5]

Společná pracovní skupina BnF a SNE zároveň dosáhla shody na tom, že odevzdávání musí probíhat v otevřených distribučních formátech a bez přítomnosti DRM. Vyloučeny byly také produkční formáty, které neodpovídají charakteristikám tištěné knihy (např. TXT nebo DOC), a proprietární uzavřené formáty (např. iBooks nebo KF8). [156, s. 5-6]

V řetězci vydavatel – distributor – prodejce potom BnF zvolila jako nejvýhodnější strategii spolupráci s distributory: „*In the French ebook market, the publisher is in charge of creating the book (both the intellectual content and the digital document); the digital bookseller is in charge of promoting and selling the ebook to end-user; the role of the distributor is to make a bridge between these two stakeholders: it receives the ebook, checks its format, verifies and enhances the quality of its metadata, and sends it to the bookseller*“ [157, s. 3].

Toto rozhodnutí je tedy podloženo několika argumenty. Předně se jedná o počet distributorů na francouzském trhu, kterých je řádově méně než nakladatelů. Dalším důvodem je skutečnost, že BnF takto fakticky může těžit z kontroly kvality prováděné distributorem. V neposlední řadě je důležitá skutečnost, že distributor, na rozdíl často od koncového obchodníka, disponuje ještě e-publikací bez DRM. Praktická implementace tohoto modelu zachovává povinnost za odevzdání publikace nakladateli, ten ji ale může delegovat na distributora. [156, s. 4-5] [157, s. 3].

Dodání dat probíhá prostřednictvím FTP, kam jsou soubory s e-knihami vystavovány producenty a na denní bázi sklíženy národní knihovnou [158, s. 7]. Některé detaily přenosu dat mezi distributorem a knihovnou jsou předmětem vyjednávání. Přijímány jsou formáty EPUB 2, EPUB 3 a PDF [157] [159]. Elektronické soubory s dokumenty nejsou při příjmu převáděny do jiných formátů a jsou tedy archivovány v originální podobě. [156, s. 5-6]

BnF ve svém systému pro dlouhodobou ochranu SPAR (Scalable Preservation and Archiving Repository) kategorizuje datové soubory, podle míry rizika spojeného s uchováním dat v těchto formátech, do čtyř úrovní [160] [161, s. 25-28]:

1. Stored	Bliže neidentifikovaná data, jimž instituce poskytne pouze nejnižší ochranu na úrovni zachování bit streamu.
2. Identified	Identifikované formáty s přiřazeným mimetypeem. Instituce nemá k formátu přiřazený žádný plán ochrany formátu.
3. Known	Identifikované formáty s dostupnou dokumentací a nástroji pro jejich správu. Instituce má plán ochrany, který garantuje možnost emulace nebo migrace do formátů s nejvyšším stupněm ochrany.
4. Managed	Identifikované formáty s dostupnou dokumentací, nástroji pro jejich správu, plánem ochrany a vlastním aplikačním profilem pro producenty, který specifikuje konkrétní potřebné vlastnosti archivovaných souborů.

Tabulka 18. Kategorizace formátů podle míry rizika pro uchování formátu dle BnF [160] [161, s. 25-28]

Dle publikované formátové politiky BnF z konce roku 2015 jsou v současné době všechny přijímané formáty, tedy varianty EPUB i PDF, řazeny do druhé až třetí kategorie úrovně ochrany [159]. Do nižší, druhé, úrovně ochrany jsou řazeny například soubory EPUB, obsahující prvky Flash nebo JavaScriptu [156, s. 8]. U těchto dokumentů BnF nemůže garantovat dlouhodobé uchování, zároveň je ale nemůže odmítnout přijmout [159, s. 2]. Příčinou, proč nejsou v současné době řazeny dokumenty ve formátech PDF ani EPUB do nejvyšší úrovně, je zejména chybějící metadatový aplikační profil pro extrakci technických metadat [157, s. 5].

Pro charakterizaci a validaci souborů EPUB využívá BnF aplikaci EPUBCheck [157, s. 5], pro charakterizaci a validaci PDF je využíván Apache Tika [162]. V minulosti byl zvažován také JHOVE pro validaci PDF, jeho využívání nicméně není ze strany BnF potvrzeno. [157, s. 5-6]

nástroj	formát	typ služby
Apache Tika	PDF	Charakterizace / validace
EPUBCheck	EPUB	Charakterizace / validace
JHOVE	PDF	Validace (zvažováno, aktuálně nepotvrzeno)

Tabulka 19. Seznam nástrojů využívaných BnF pro práci s přijímanými e-publikacemi [157, s. 5-6] [162]

## 7.5 *The National Library of Australia*

Povinný výtisk byl pro elektronické publikace rozšířen v Austrálii v roce 2016 [163, s. 30]. Nakladatelé nejsou povinni odevzdávat všechny online vydané publikace, povinnost jim vzniká teprve na základě požadavku ze strany Národní knihovny Austrálie (National Library of Australia, NLA). V případě vydání díla v tištěné i elektronické podobě je vydavatel povinen odevzdat pouze tištěnou verzi díla, pokud se s NLA nedomluví jinak. [164]

Odevzdávání publikací probíhá přes webový portál NLA [165]. Odevzdání publikací je možné výhradně ve formátech EPUB, PDF a MOBI [163, s. 42]. Maximální velikost souborů je nastavena na 500 MB. Obálku publikace je možné odevzdat ve formátech JPG nebo TIFF. V případě publikování ve více formátech je nakladatel povinen odevzdat pouze jednu verzi, v takovém případě NLA upřednostňuje odevzdání ve formátu EPUB. Když publikace není dostupná ve formátu EPUB, je upřednostňován formát PDF před MOBI<sup>55</sup>. Všechny soubory musí být prosty veškerých technických prostředků pro DRM, včetně vodotisku. [164]

NLA neukládá povinnost odevzdat publikace ve formátech, ve kterých byly určeny pro komerční trh. Naopak, v případě vydání publikací v neakceptovaných formátech je vydavatel povinen, za možné podpory NLA, převést vyžádané publikace do některého z akceptovaných formátů.<sup>56</sup> NLA nespecifikuje verze formátů EPUB či PDF, je tedy možné odevzdávat soubory ve všech verzích PDF i EPUB. Pokud jsou publikace v těchto formátech, je možné bez omezení odevzdat také publikace obsahující multimediální nebo interaktivní prvky.<sup>57</sup>

Při ingestu souborů do dlouhodobého úložiště není ze strany NLA prováděna normalizace formátů, všechny akceptované formáty jsou uchovávány v původní podobě. K migraci do jiného formátu bude přistoupeno teprve tehdy, pokud by formát přestal být dostatečně podporován.<sup>58</sup> Podpora MOBI je tedy v současné době považována ještě za dostačující.

Uchování dat probíhá v dlouhodobém úložišti Preservica. Obsah úložiště v březnu 2017 byl dle počtu souborů tvořen z 25 % soubory ve formátu PDF, ze 33 % EPUB, ze 32 % JPG a z 10 % ostatními souborovými formáty [166, s. 22]. Po dvou letech od zahájení příjmu e-publikací, v květnu 2018, se poměr uložených souborových formátů výrazně přiklonil k PDF, například

---

<sup>55</sup> Zjištěno z e-mailové korespondence s Liborem Coufalem, vedoucím oddělení dlouhodobého digitálního uchovávání v Národní knihovně Austrálie.

<sup>56</sup> Libor Coufal v emailové korespondenci zmiňuje případ publikací, vydaných původně ve formátu iBooks, které nakladatel převedl kvůli odevzdání do PDF, publikace tím ale přišly o multimediální prvky.

<sup>57</sup> Tamtéž.

<sup>58</sup> Tamtéž.

jen nejčastěji zastoupené verze PDF 1.3 bylo uloženo téměř 117 000 souborů, oproti cca 6 000 souborům EPUB a pouhým 222 souborů MOBI<sup>59</sup>.

nástroj	formát	typ služby
Apache Tika	všechny formáty	Identifikace (v rámci portálu eDeposit)
Droid	všechny formáty	Identifikace (v rámci Preservicy)
Flint <sup>60</sup>	PDF, EPUB	detekce DRM
JHOVE	PDF	charakterizace / validace

Tabulka 20. Seznam nástrojů využívaných NLA pro práci s přijímanými e-publikacemi<sup>61</sup>

## 7.6 Narodna in univerzitetna knjižnica (Slovensko)

Slovenská Národní a univerzitní knihovna (Narodna in univerzitetna knjižnica, NUK) má právo povinného výtisku pro elektronické publikace od roku 2006 [167], resp. 2007, kdy byly vydány též potřebné prováděcí předpisy [168] [150, s. 38-42]. Odevzdávání publikací probíhá sklízením ze strany knihovny metodou web-harvestingu, nebo, v případech, kdy to není možné například z důvodu zabezpečení u elektronického obchodu, jsou publikace deponovány do knihovny nakladatelem přímo prostřednictvím portálu Svarog [169] [170].

Vydavatel je povinen odevzdat e-publikace ve všech formátech, kterým přidělil číslo ISBN, nebo ve kterých bylo dílo publikováno. Veškeré e-publikace musí být odevzdány bez DRM zabezpečení. NUK doporučuje pro textové materiály odevzdání ve formátech PDF, EPUB nebo MOBI. Doprovodné obrazové soubory je potom doporučeno odevzdávat ve formátech JPEG2000, JPEG nebo PDF [170]. Jedná se nicméně ze strany NUK pouze o doporučení, knihovna je povinna přijmout e-publikaci ve všech formátech, ve kterých je vydána<sup>62</sup>. V případě, kdy je publikace vydána a odevzdána ve více formátech, je knihovna povinna zvolit k uchování takový formát, který umožní dílo uchovat v originální podobě a takto zajistit jeho autentičnost a integritu na straně jedné a dlouhodobou dostupnost na straně druhé [168].

NUK v současné době neprovádí validaci odevzdaných e-publikací, neuplatňuje tedy ani žádné jiné požadavky na vlastnosti dat, vyjma již zmíněného zákazu DRM<sup>63</sup>. Bez omezení tak mohou být odevzdávány interaktivní či multimediální publikace nejen ve formátech PDF a EPUB 3.

<sup>59</sup> Zjištěno z e-mailové korespondence s Liborem Coufalem, vedoucím oddělení dlouhodobého digitálního uchování v Národní knihovně Austrálie.

<sup>60</sup> Obsahuje validační nástroje Apache PDFBox, Apache Tika, Calibre, EPUBCheck a iText.

<sup>61</sup> Tamtéž.

<sup>62</sup> Zjištěno z e-mailové korespondence s Jankem Klasincem, vedoucím oddělení správy digitální knihovny slovenské Národní a univerzitní knihovny. Srov také s [168].

<sup>63</sup> Zjištěno z e-mailové korespondence s Jankem Klasincem. Nevalidnost souborového formátu není dle Janka Klasince ani důvodem pro odmítnutí souboru a vyžádání náhradní dodávky.



Při příjmu dat není prováděna normalizace, odevzdané publikace jsou tedy vždy uchovávány v původní podobě. V minulosti byla zvažována např. možnost předepsat vydavatelům povinnost odevzdávat všechny textové dokumenty, bez ohledu na původní formát (včetně např. EPUB či MOBI), rovnou ve formátu PDF/A, pro očekávaný nedostatek expertních znalostí na straně nakladatelů, potřebných pro korektně provedené konverze, nakonec knihovna od požadavku zcela upustila. Podobně z důvodu rizika ztráty některých vlastností odevzdaných publikací v současné době NUK neuvažuje ani o konverzi do PDF/A ve vlastní režii.<sup>64</sup>

Pokud to bude nezbytné pro zachování dostupnosti publikace, umožňuje zákon o povinném výtisku knihovně provést potřebné technické změny formou převodu do jiného formátu. V takovém případě je však NUK kladeno za povinnost zajistit zachování nejen informačního obsahu, ale také vizuální a případně též zvukové složky uchovávané publikace. [167]

### **7.7 Koninklijke Bibliotheek (Nizozemí)**

Nizozemí nemá zákon o povinném výtisku. Jak tištěné, tak i elektronické publikace jsou tedy nakladateli odevzdávány čistě na dobrovolné bázi. Právní rámec je dán dvoustrannou dohodou mezi nizozemskou Královskou knihovnou (Koninklijke Bibliotheek, KB) a asociací nizozemských nakladatelů (Nederlands Uitgeversverbond), uzavřenou v roce 1999 [171] a aktualizovanou v letech 2005 a 2014. [172] [173]

KB se v dohodě zavazuje uchovávat publikace po neomezenou dobu, zachovávat jejich integritu a obsah. V nezbytných případech je KB povinna provádět ochranná opatření vedoucí k zajištění dlouhodobé dostupnosti odevzdaných publikací, včetně konverzí na nové formáty. V případě formátové migrace by měla být vždy zajištěna autentičnost kopie, odvozené od původně odevzdaného díla. Pokud nelze zajistit uchování díla bez ztráty některých jeho vlastností, je KB povinna konzultovat provedení ochranných opatření s nakladatelem. KB má povinnost označit kopii v novém formátu tak, aby bylo zřejmé, že odchylky od originálního dokumentu vznikly jako důsledek ochranných opatření KB, a nikoliv na straně nakladatele. [171]

V případě vydání publikace ve více formátech je nakladatel v souladu s dohodou povinen odevzdat jednu kopii publikace. Vhodný formát odevzdání v takovém případě určuje KB [171]. U odevzdaných publikací je zakázána přítomnost DRM. Přítomnost DRM je potom kontrolována při příjmu publikací nástroji Apache Tika a EPUBCheck [69, s. 18].

---

<sup>64</sup> Zjištěno z e-mailové korespondence s Jankem Klasincem, vedoucím oddělení správy digitální knihovny slovinské Národní a univerzitní knihovny.

KB se v minulosti soustředila zejména na sběr vědeckých a vládních publikací, teprve v poslední době se zaměřuje také na pokrytí komerčního trhu e-knih včetně beletrie [69, s. 18]. Portál Webloket, určený malým nakladatelům a autorům<sup>65</sup>, v současné době omezuje přijímané formáty na PDF a formáty Microsoft Word [174]. V průběhu roku 2018 byl nově spuštěn také příjem publikací ve formátech EPUB 2 a EPUB 3<sup>66</sup>.

Přijímány jsou všechny verze formátu PDF. Jedinou uplatňovanou podmínkou je absence DRM ochrany. Přijímané soubory PDF jsou uchovávány v původní podobě, KB v současné době neprovádí normalizaci z formátu PDF na PDF/A<sup>67</sup>. Johan van der Knijff ve svém příspěvku z roku 2014 navrhuje hypotetickou formátovou politiku pro PDF [175], v současné době se ale nejedná o uplatňovanou politiku KB. Podmínky pro PDF, bezpečné z hlediska digitální archivace, navrhuje van der Knijff následující:

- *soubor nesmí být šifrován nebo chráněn heslem*
- *kompletní fonty jsou vloženy*
- *soubor nesmí obsahovat JavaScript*
- *soubor nesmí obsahovat vložené soubory (např. přílohy)*
- *soubor nesmí obsahovat multimediální obsah (audio, video, 3D objekty)*
- *soubor je validní PDF [175]<sup>68</sup>*

Podobně definoval Johan van der Knijff podmínky pro bezpečný příjem formátu EPUB 2 a EPUB 3 [99], jejichž parametrům pro dlouhodobé uchování se ve svých formátových studiích také dlouhodobě věnuje.<sup>69</sup> Tato pravidla jsou ze strany KB při příjmu uplatňována<sup>70</sup>:

- *musí se jednat o validní EPUB*
- *soubor nesmí obsahovat DRM nebo šifrování*
- *soubor nesmí obsahovat cizí formáty (pouze tzv. Core Media Types)*
- *soubor nesmí obsahovat obsah ve formě DTBook [99]*

Formáty Microsoft Word jsou po příjmu normalizovány do archivní podoby PDF/A, která je následně uložena společně s originálně dodaným zdrojovým dokumentem [176, s. 5]. Pro převody do PDF využívá KB aplikaci Adobe Acrobat Professional<sup>71</sup>.

---

<sup>65</sup> Zjištěno z e-mailové korespondence s Johanem van der Knijffem, formátovým specialistou z nizozemské Královské knihovny.

<sup>66</sup> Tamtéž.

<sup>67</sup> Tamtéž.

<sup>68</sup> Srov. s [205], [206], [207].

<sup>69</sup> Viz např. [94], [105].

<sup>70</sup> Zjištěno z e-mailové korespondence s Johanem van der Knijffem.

<sup>71</sup> Tamtéž.

KB neuplatňuje při příjmu žádné požadavky na autenticitu dokumentů v původních formátech, naopak předpokládá, že mnoho odevzdaných PDF bylo vydáno v jiných formátech. Pracovníci KB rovněž nezaznamenali větší výskyt interaktivních nebo multimediálních publikací, v současné době tedy nejsou zvláštními pravidly řešeny. Z dalších formátů nyní KB zkoumá možnosti příjmů elektronických periodik vydávaných ve formátech na bázi XML.<sup>72</sup>

nástroj	formát	typ služby
Apache Tika	všechny formáty	identifikace
Apache Tika	PDF	charakterizace / validace / detekce DRM
EPUBCheck	EPUB	charakterizace / validace / detekce DRM

Tabulka 21. Seznam nástrojů využívaných KB pro práci s přijímanými e-publikacemi<sup>73</sup>

## 7.8 The National Library of New Zealand

Národní knihovna Nového Zélandu (The National Library of New Zealand, NLNZ) disponuje právem povinného výtisku pro elektronické publikace od roku 2003. Systém pro příjem elektronických publikací NLNZ spustila počátkem roku 2007. [177]

Příjem e-publikací je možný buď sklizením metodou web-harvestingu (po zaslání odkazu do NLNZ), nasdílením prostřednictvím DropBoxu nebo nahráním na FTP server knihovny. Publikace musí být odevzdány bez DRM omezení. V případě vydání díla ve více různých formátech je nakladatel povinen odevzdat pouze jeden z formátů, který určuje NLNZ<sup>74</sup>. Národní knihovna neurčuje konkrétní formáty, doporučuje nicméně otevřené formáty určené k distribuci, například EPUB, před produkčními formáty, jako je například DOC. [178]

Preferovány jsou originální distribuční formáty také v případech, kdy jsou publikace vydány pouze v proprietárních uzavřených formátech. V takových případech nicméně NLNZ usiluje také o dodání kopie pro zpřístupnění dokumentu v lépe přístupném formátu. Publikace v produkčních (např. DOC nebo HTML) formátech jsou přijímány zejména u dokumentů publikovaných prostřednictvím platformy Amazonu. Důvodem je, že nakladatelé v těchto případech k publikaci v distribučním formátu nemají vůbec přístup.<sup>75</sup>

Pokud je to možné, preferuje NLNZ odevzdání ve formátu EPUB, formátem druhé volby je potom PDF ve všech verzích. Z dalších formátů jsou nejčastěji přijímány publikace ve

<sup>72</sup> Zjištěno z e-mailové korespondence s Johanem van der Knijffem, formátovým specialistou z nizozemské Královské knihovny.

<sup>73</sup> Tamtéž.

<sup>74</sup> Zjištěno z e-mailové korespondence s Amy Joseph, legal deposit librarian z NLNZ.

<sup>75</sup> Tamtéž

formátech MOBI, DOC, DOCX a HTML (viz tabulka 22). Pokles zaznamenala NLNZ v poslední době u formátu MOBI, který již není podporovaný Amazonem.<sup>76</sup>

NLNZ využívá k uchování odevzdaných e-publikací dlouhodobé úložiště ExLibris Rosetta [179]. Pro identifikaci souborových formátů je používán nástroj DROID. V případech, kdy je publikace vydána jak v tištěné, tak v elektronické verzi, je zpravidla preferována již pouze elektronická verze dokumentu [180]. V případě odevzdání nevalidních souborů jsou od nakladatelů vyžádány bezchybné kopie, v některých případech jsou ale rovnou provedeny opravy digitálními kurátory – podmínkou je neporušení obsahu a podoby díla<sup>77</sup>. Z důvodu možné ztráty obsahu či vlastností dokumentů NLNZ neprovádí normalizace dat po příjmu a nepodporuje změny formátů ani na straně původce. Formátové migrace jsou prováděny pouze za účelem zpřístupnění, kdy dochází např. ke konverzím z MOBI do EPUB.<sup>78</sup>

V dubnu 2018 čítala sbírka e-knih odevzdaných prostřednictvím povinného výtisku více než 74 tisíc položek. Zastoupení jednotlivých formátů ukazuje tabulka 22 [181].

formát	PDF	RTF	EPUB	MOBI	DOC	DOCX	HTML	iBooks	AZW3	ODT	PRC	AZW	DOCM
počet	64 269	5 478	3 401	539	267	182	18	5	2	2	2	1	1
%	86,25 %	7,35 %	4,56 %	0,72 %	0,36 %	0,24 %	0,03 %	0,01 %	0,00 %	0,00 %	0,00 %	0,00 %	0,00 %

Tabulka 22. Struktura formátů e-publikací odevzdaných do NLNZ<sup>79</sup>

## 7.9 The British Library

Britská národní knihovna (The British Library, BL) zahájila pilotní provoz systému pro příjem elektronických publikací v roce 2000 [182, s. 33]. Příjemcem povinného elektronického výtisku se stala v roce 2013 poté, co vstoupilo v platnost *Nářízení o knihovnách pro povinné ukládání (netištěná díla) č. 777* [183]. Právní rámec pro povinné odevzdávání elektronických publikací stanovil již o deset let dříve *Zákon č. 28 o knihovnách pro povinné ukládání z roku 2003*<sup>80</sup> [184]. Nařízení předepisuje odevzdání e-publikací pouze v případě, že nemá tištěný ekvivalent. Nakladatelé jsou povinni odevzdat publikace výhradně v původních distribučních formátech. V případě vydání e-publikace ve více formátech je na dohodě vydavatele s příslušnou depozitní knihovnou, který z formátů je odevzdán [183, s. 5-6]. Preferovanými formáty jsou v takovém

<sup>76</sup> Zjištěno z e-mailové korespondence s Amy Joseph, legal deposit librarian z NLNZ.

<sup>77</sup> Tamtéž.

<sup>78</sup> Tamtéž.

<sup>79</sup> V tabulce byly ponechány pouze textové formáty, jejichž obsah lze považovat za textové e-publikace. Vynechané: jpg (223), xlsx (52), xlx (36), mp3 (7), png (7), jpeg (6), csv (2), ppsx, ppt, pptx, pub, rar, tiff, zip (1).

<sup>80</sup> České překlady názvů zákonů převzaty z článku Zdeňka Matušika [192].

případě PDF nebo EPUB<sup>81</sup>. Běžně akceptovány jsou nicméně také další formáty, jako jsou zejména XML, HTML, SGML, Microsoft Word nebo RTF [185]. Bez omezení jsou přijímány také interaktivní a multimediální publikace v těchto formátech.

Publikace jsou buď sklíženy metodou web-harvestingu, odevzdávány přes FTP server, nebo prostřednictvím portálu pro nakladatele [186]. Všechny dokumenty musí být odevzdány bez DRM [185]. Pokud odevzdané publikace neprojdou úspěšně validací souborového formátu, není to ze strany BL nahlíženo jako nesplnění povinnosti nakladatele odevzdat kopie v kvalitě vhodné pro dlouhodobé uchování. Náhradní kopie je zpravidla požadována teprve v případě poškození souboru, nebo pokud není dodanou publikaci možné zobrazit<sup>82</sup>. Přístup BL pramení mimo jiné ze zkušeností s kvalitou využívaných validačních nástrojů.

V rámci testování aplikace JHOVE v nové verzi v1.16 provedla BL v roce 2017 validaci vzorku složeného z 2759 e-knih a e-časopisů odevzdaných ve formátu PDF. Výsledky testu (tabulka 23) podávají jednak určitý obrázek ohledně kvality dat, dodávaných do BL, zejména ale poukazují na zásadní kvalitativní rozdíl mezi různými verzemi validačních nástrojů. Z 1079 (39,1 %) původně nevalidních souborů bylo podle nové verze validátoru nevalidních pouze 388 (14 %). Přinejmenším 691 souborů tedy bylo starší verzí chybně označeno jako nevalidní. [187]

BL neprovádí preventivní normalizace do jiných souborových formátů, strategií knihovny je uchování publikací v originálních formátech, pokud je to možné. Převod publikace do jiného formátu přichází v úvahu zejména pro potřeby zpřístupnění<sup>83</sup>.

V roce 2017 provedla BL vizuální kontrolu na vzorku 1,111 souborů PDF, aby ověřila efektivnost nastavených vstupních kontrol [188]. Jedním ze závěrů bylo, že stávající praxe, kdy jsou z přijatých nevalidních souborů vytvářeny modifikované validní kopie pro zpřístupnění, je jednak zbytečná, a v některých případech dokonce přímo způsobuje nové chyby. Na základě těchto zjištění bylo následně od této praxe upuštěno. [189, s. 12]

Výsledek	JHOVE v1.15	JHOVE v1.16
Well-Formed and valid	1680 / 60,9 %	2371 / 85,9 %
Well-Formed, but not valid	220 / 8,0 %	150 / 5,4 %
Not well-Formed	859 / 31,1 %	238 / 8,6 %

Tabulka 23. Ověření validity 2759 souborů PDF z fondu BL – porovnání JHOVE v1.15 a JHOVE v1.16 [187]

<sup>81</sup> Zjištěno z e-mailové korespondence s Maureen Pennock, vedoucí odd. digitální ochrany Britské knihovny.

<sup>82</sup> Tamtéž.

<sup>83</sup> Tamtéž.

Mezi lety 2015 až 2018 publikovala BL řadu formátových studií, zabývajících se zhodnocením formátů elektronických publikací z hlediska dlouhodobého uložení. V roce 2015 to byly studie věnované formátům PDF [130] a EPUB [87], v roce 2017 souhrnná zpráva k formátům e-publikací [6] a v roce 2018 studie k formátu MOBI [77]. Pro charakterizaci a validaci využívá BL zejména nástroje EPUBCheck a JHOVE.

nástroj	formát	typ služby
JHOVE	PDF	charakterizace / validace
EPUBCheck	EPUB	charakterizace / validace

Tabulka 24. Seznam nástrojů využívaných BL pro práci s přijímanými e-publikacemi<sup>84</sup>

### 7.10 Dílčí závěry k 7. kapitole

Všechny sledované knihovny přijímají bez omezení všechny verze PDF a formát EPUB 2 i EPUB 3. Pouze KB uplatňuje dodatečné podmínky pro podobu souborů ve formátu EPUB (zákaz cizích formátů a DTBook). Ostatní knihovny žádné jiné podmínky, než zákaz DRM, u žádného z formátů neuplatňují. Všechny knihovny tedy přijímají také interaktivní a multimediální e-publikace, pokud jsou odevzdány ve formátu EPUB 3 nebo PDF<sup>85</sup>. 3 z 7 knihoven přijímají také publikace v proprietárním formátu MOBI, dvě z knihoven (NUK, NLNZ) neuplatňují žádná omezení na akceptované formáty. Při bližším pohledu se jeví, že ani uvolněnější formátová politika nemusí způsobit dramatické zahlcení knihoven proprietárními formáty, pokud si mohou pro dlouhodobé uchování zvolit pouze jednu z vydaných verzí<sup>86</sup>.

knihovna	EPUB 2	EPUB 3	PDF	MOBI	další formáty	interaktivní a multimediální
DNB	✓	✓	✓	X	X	✓ (neomezují)
BnF	✓	✓	✓	X	X	✓ (neomezují)
NLA	✓	✓	✓	✓	X	✓ (neomezují)
NUK	✓	✓	✓	✓	(neomezují)	✓ (neomezují)
KB	✓	✓	✓	X	MS Word	✓ (neomezují)
NLNZ	✓	✓	✓	✓	(neomezují)	✓ (neomezují)
BL	✓	✓	✓	X	ano <sup>87</sup>	✓ (neomezují)

Tabulka 25. Přehled formátů, přijímaných sledovanými knihovnami

<sup>84</sup> Zjištěno z e-mailové korespondence s Maureen Pennock, vedoucí odd. digitální ochrany Britské knihovny.

<sup>85</sup> Nelze to však vykládat vždy jako aktivní podporu příjmu tohoto typu publikací, ve většině případů zkrátka knihovny tento typ obsahu nejsou při stávající úrovni vstupních kontrol sto odhalit, případně se nevyskytují tak často, aby je knihovny začaly považovat za větší riziko.

<sup>86</sup> Příkladem je velmi malé zastoupení formátu MOBI ve fondu NLA nebo NLNZ.

<sup>87</sup> Z dalších formátů jsou přijímány: XML, HTML, SGML, Microsoft Word, RTF.

S výjimkou normalizace dat z formátu MS Word do PDF v případě KB žádná z institucí nepotvrdila provádění normalizace do jiných formátů (např. MOBI do EPUB nebo PDF do PDF/A). Všechny tedy zatím uchovávají formáty v odevzdaných formátech. Čtyři z knihoven (DNB, NUK, NLNZ, BL) naopak důsledně požadují po nakladatelích odevzdání e-publikací v autentické podobě a formátu, v jakých byly určeny pro komerční trh. Dvě knihovny (BL, NLNZ) připustily, v omezené míře, provádění formátových konverzí pro potřeby zpřístupnění. Zjištění ohledně využívaných softwarových nástrojů ukázalo na převažující užívání aplikace EPUBCheck pro validaci EPUB a užití JHOVE nebo Apache Tika pro validaci PDF. Pro identifikaci souborových formátů je nejčastěji využíván DROID.

Srovnání zjištěných přístupů s formátovou politikou NK ČR ukazuje celou řadu rozdílů. Prvním a nejzásadnějším rozdílem je přístup k PDF/A. Přestože některé z knihoven uvádějí PDF/A jako preferovaný formát, žádná ze sedmi institucí neomezuje příjem nearchivních PDF a žádná z institucí také neprovádí normalizaci do formátu PDF/A v průběhu ingestu.

## 8 Analýza e-publikací, dobrovolně odevzdaných do NK ČR

V rámci projektu eDpilot a pozdějšího projektu NAKI zahájila NK ČR v roce 2011 v ověřovacím režimu přijímání e-publikací a elektronických tiskových předloh tištěných publikací. Odevzdávání e-publikací probíhá na dobrovolné bázi. Národní knihovna se zavazuje k dlouhodobé ochraně těch publikací, které jsou odevzdány ve formátech PDF/A-1, PDF/A-2 nebo EPUB 2. Ostatní jsou přijímány pouze k bitové ochraně, jejich dlouhodobá dostupnost tedy není garantována. Odevzdávání e-publikací do NK ČR probíhá nejčastěji prostřednictvím e-mailu, publikace vydané v režimu open access bývají stahovány přímo z webových stránek producenta. Méně časté je předávání na fyzických nosičích nebo prostřednictvím různých webových služeb pro výměnu dat<sup>88</sup>. Od roku 2011 přijala NK ČR cca 1300 jednotlivých souborů e-publikací, pocházejících od 72 různých nakladatelů. Analýze těchto dat se budu věnovat v této kapitole. Vzorek dat není svou velikostí zcela reprezentativní z pohledu celkové produkce e-publikací v ČR. Vzhledem k velkému zastoupení menších nakladatelů a samovydavatelů je nicméně shromážděná kolekce zajímavá z hlediska ověření kvality datových souborů právě u segmentu menších producentů e-publikací, u nichž lze očekávat největší rozmanitost v používaných produkčních nástrojích a tím také v kvalitě souborů.

### 8.1 Použité nástroje

Při práci s daty jsem se v první řadě zaměřil na identifikaci souborových formátů. zastoupených ve zkoumaném vzorku dat. K identifikaci jsem využil nástroj DROID ve verzi 6.4, s využitím aktuálního Signature File V93. U souborů PDF jsem zároveň využil také validátor veraPDF ve verzi 1.12.1, který dokáže na rozdíl od DROIDu identifikovat spolehlivě také soubory PDF/A. Následně jsem provedl validaci nástroji, relevantními pro zjištěné souborové formáty. Kromě ověření validity souborů jsem se zaměřil také na přítomnost DRM ochrany v kontrolovaných souborech. Pro formát EPUB byl využit validační nástroj EPUBCheck ve verzi 4.0.2, pro formáty PDF a JPG byl využit nástroj JHOVE ve verzi 1.20, resp. jeho moduly PDF-hul ve verzi 1.11 a JPG-hul ve verzi 1.4. Pro formáty MOBI, iBooks a MS Word jsem nezjistil žádné vhodné validační nástroje, místo toho jsem tedy ověřil zobrazitelnost dokumentů a možnost jejich konverze do otevřených formátů. V případě MOBI a iBooks byl využit nástroj Calibre ve verzi 3.21, u dokumentů DOC potom textový editor Microsoft Word 2016.

---

<sup>88</sup> Zjištěno z e-mailové korespondence s Martinem Žížalou, zástupcem vedoucí v Oddělení doplňování domácích dokumentů Národní knihovny ČR, který od roku 2011 příjem elektronických dokumentů zajišťuje.



## 8.2 Struktura formátů přijatých dokumentů

koncovka	název formátu	PUID	mime type	verze formátu	počet souborů	nakladatelů
epub	ePub format	fmt/483	application/epub+zip	2.0.1	1012	7
pdf	Acrobat PDF 1.4 - Portable Document Format	fmt/18	application/pdf	1.4	59	24
prc	PocketMobi (Palm Resource) File	fmt/396			53	1
pdf	Acrobat PDF 1.5 - Portable Document Format	fmt/19	application/pdf	1.5	50	19
pdf	Acrobat PDF 1.6 - Portable Document Format	fmt/20	application/pdf	1.6	34	15
pdf	Acrobat PDF 1.7 - Portable Document Format	fmt/276	application/pdf	1.7	23	16
jpg	JPEG File Interchange Format	fmt/43	image/jpeg	1.01	22	4
pdf	Acrobat PDF 1.3 - Portable Document Format	fmt/17	application/pdf	1.3	21	8
mobi	PocketMobi (Palm Resource) File	fmt/396			13	6
pdf	Acrobat PDF 1.2 - Portable Document Format	fmt/16	application/pdf	1.2	4	1
epub	ePub format	fmt/483	application/epub+zip	3.0.1	3	2
doc	Microsoft Word Document	fmt/40	application/msword	97-2003	2	2
ibooks	Apple iBook format	fmt/482	application/x-ibooks+zip		1	1
jpg	JPEG File Interchange Format	fmt/42	image/jpeg	1.00	1	1

Tabulka 26. Výsledek identifikace souborových formátů nástrojem DROID

## 8.3 Soubory ve formátu EPUB

Ve zkoumaném vzorku bylo zjištěno celkem 1015 souborů EPUB od 8 nakladatelů. Z tohoto počtu bylo ve verzi 2.0.1 celkem 1012 souborů a 3 soubory ve verzi 3.0.1. Validace nástrojem EPUBCheck určila celkem 1010 souborů jako Well-formed, z tohoto počtu se u 6 souborů objevilo varování na nedodržení doporučených částí formátové specifikace. U 5 souborů byly zjištěny vážnější chyby, a tyto byly tedy vyhodnoceny jako Not well-formed. Všechny nevalidní soubory byly zobrazitelné aplikací Calibre bez viditelných chyb. Zjištěné chyby ukazují tabulka 27 a tabulka 28. U žádného z kontrolovaných souborů nebyla zjištěna přítomnost DRM.

výsledek kontroly EPUBCheck	zjištěné chyby <sup>89</sup>	počet souborů
Well-formed		1010
Well-formed, s varováním	HTM-014; CSS-017; OPF-003	6
Not well-formed	RSC-005; RSC-012; OPF-015; CSS-017; PKG-010, OPF-055	5

Tabulka 27. Výsledek validace souborů EPUB validátorem EPUBCheck

<sup>89</sup> Vysvětlení chyb je k dispozici v Github repozitáři aplikace EpubCheck: [210].

kód chyby	kategorie	chybové hlášení	počet souborů
CSS-017	WARN	CSS selector specifies absolute position.	4
HTM-014	WARN	Invalid file extension for HTML file, expecting (html, htm or xhtml).	5
OPF-003	WARN	Item '%1\$s' exists in the EPUB, but is not declared in the OPF manifest.	1
OPF-055	WARN	%1\$s tag is empty.	1
PKG-010	WARN	Filename contains spaces, therefore URI escaping is necessary. Consider removing spaces from filename.	1
OPF-015	ERROR	The property '%1\$s' should not be declared in the OPF file.	1
RSC-005	ERROR	A fragment identifier is required for svg use tag references.	4
RSC-012	ERROR	Fragment identifier is not defined.	1

Tabulka 28. Chyby zjištěné validátorem EPUBCheck u souborů EPUB

## 8.4 Soubory ve formátu PDF

Mezi odevzdanými daty bylo přítomno celkem 191 souborů PDF od 67 různých nakladatelů. Zastoupeny jsou verze formátu od 1.2 po 1.7. Aplikace JHOVE označila jako *Well-Formed and valid* 153 z ověřovaných souborů. Další 5 souborů bylo označeno jako *Well-Formed, but not valid* a 33 jako *Not well formed*. U 24 souborů JHOVE detekoval soulad s profilem PDF/A-1b, následné ověření validátorem veraPDF však soulad s PDF/A vyloučil. Podle kontroly nástrojem VeraPDF není žádný ze 191 souborů PDF validní vůči některému z profilů PDF/A. Prostřednictvím veraPDF<sup>90</sup> byla také ověřena přítomnost DRM v souborech. Šifrování bylo zjištěno u 18 souborů od 4 vydavatelů, ve všech případech byly PDF soubory zabezpečeny heslem proti úpravám, v jednom případě také proti tisku a vykopírování textu.

Výsledek kontroly PDF-hul	Jméno chyby	ID chyby <sup>91</sup>	počet
Well-Formed and valid			153
Well-Formed, but not valid	Improperly formed date	pdf-hul-133	3
Well-Formed, but not valid	Invalid indirect destination	pdf-hul-149	1
Well-Formed, but not valid	Invalid destination object	pdf-hul-1	1
Not well formed	Compression method is invalid or unknown to JHOVE <sup>92</sup>	pdf-hul-109	17
Not well formed	Unexpected exception java.lang.NullPointerException		5
Not well formed	Malformed outline dictionary	pdf-hul-124	3
Not well formed	Invalid Names dictionary	pdf-hul-89	3
Not well formed	Lexical error	pdf-hul-65	2
Not well formed	Invalid object number or object stream	pdf-hul-108	2
Not well formed	Malformed filter	pdf-hul-45	1

Tabulka 29. Výsledek validace souborů PDF modulem PDF-hul (JHOVE)

<sup>90</sup> Při ověřování byla využita konfigurace validátoru veraPDF navržená Johannem van der Knijffem [204].

<sup>91</sup> Chyby jsou vysvětleny v GitHub repozitáři projektu JHOVE: [208], [209].

<sup>92</sup> Chyba souvisí s přítomností DRM, všech 17 publikací bylo zabezpečeno heslem proti úpravám.

## 8.5 Soubory ve formátu MOBI

Formát MOBI je ve zkoumaném vzorku zastoupen 66 soubory. Ve 13 případech mají soubory koncovku .mobi – tyto soubory pocházejí od 6 různých nakladatelů. Zbýlých 53 souborů s koncovkou .prc pochází od jednoho nakladatele. Vzhledem k neexistenci validačních nástrojů bylo u této části dat provedeno alespoň ověření zobrazitelnosti prostřednictvím nástroje Calibre. Ověření zjistilo u 53 souborů přítomnost DRM zabezpečení, které brání jak zobrazení obsahu, tak také konverzí do jiných formátů. Dodatečné ověření prokázalo zobrazitelnost těchto souborů v aplikaci Mobipocket. Soubory jsou tedy chráněny metodou DRM, která limituje jejich užití pouze na oficiální čtecí aplikaci výrobce. Vzhledem k tomu, že aplikace Mobipocket je již spolu s webem Mobipocket.com u výrobce nedostupná a přítomnost ochrany efektivně brání konverzi do jiného formátu, představují tyto soubory vysoký stupeň rizika.

Zbýlých 13 souborů je zobrazitelných bez omezení a je u nich také možný převod do jiného formátu. Ověřovací převod do formátu EPUB 2 byl úspěšný z části, výsledné soubory jsou zobrazitelné, nejsou však validní vůči formátové specifikaci. Příčinou mohou být chyby již ve zdrojových XHTML a CSS souborech v původních souborech MOBI.

koncovka	souborů celkem	Calibre	Mobipocket Reader	s DRM	převoditelných do EPUB 2
mobi	13	13 zobrazitelných	13 zobrazitelných	0	13
prc	53	0 zobrazitelných	53 zobrazitelných	53	0

Tabulka 30. Výsledek kontroly souborů MOBI v aplikacích Calibre a Mobipocket Reader

## 8.6 Ostatní souborové formáty

Pro formát iBooks není k dispozici validační nástroj, ostatně soulad či nesoulad s neveřejnou formátovou specifikací není zásadní. Zobrazitelnost souboru jsem ověřil prostým přejmenováním přípony z .ibooks na .epub. Po změně přípony a změně mimetype v souboru *mimetype* na hodnotu *application/epub+zip* bylo možné soubor otevřít například v aplikaci Calibre. O validní EPUB se nicméně nejedná, v důsledku užití proprietárních rozšíření formátu je z původního dokumentu zobrazitelná a jen s obtížemi čitelná pouze textová složka.

U dvou souborů MS Word 97-2003 bylo provedeno ověření zobrazitelnosti a převodu do PDF/A. Oba soubory byly úspěšně převedeny aplikací MS Word 2016 do PDF/A-3a. Výsledné soubory byly úspěšně zvalidovány aplikací veraPDF. Všechny soubory ve formátu JPG byly validátorem zhodnoceny jako *Well-Formed and valid*.

formát	EPUB	PDF	MOBI	DOC	iBooks	CELKEM
počet souborů	1015	191	66	2	1	1275
zjištěno DRM	0 / 0 %	18 / 9,4 %	53 / 80,3 %	0 / 0 %	0 / 0 %	71 / 5,57 %
nevalidní	5 / 0,5 %	38 / 19,9 %	N/A	N/A	N/A	43 / 3,37 %

Tabulka 31. Výskyt DRM a nevalidních souborů u jednotlivých zastoupených formátů

## 8.7 Dílčí závěry k 8. kapitole

Chybovost u formátu PDF odpovídá přibližně zkušenostem zahraničních institucí. Naopak jako relativně nízkou lze hodnotit úroveň chybovosti u formátu EPUB. Přesto by bylo, vzhledem k časové náročnosti, a v některých případech, i nemožnosti dodatečné opravy dat, vhodné do budoucna příjmu nevalidních souborů ve formátu EPUB, pokud možno zcela zabránit. Za závažné lze považovat zejména zjištění ochranných opatření znemožňujících úpravy u celkem 71 souborů (5,57 %). Tyto e-publikace nebude možné bez zajištění přístupových práv (hesla) od nakladatele, nebo prolomení jejich ochrany hrubou silou, do budoucna ochránit.

V dalších analýzách, jdoucích již nad rámec této práce, by bylo vhodné se podrobněji zaměřit také na vložený obsah v souborech EPUB a PDF, přítomnost použitých fontů v souborech EPUB i PDF a provést praktické ověření možností převodu souborů PDF do PDF/A. Kompletní výsledky analýzy dat jsou přiloženy k elektronické verzi práce jako Příloha 1.

## 9 Závěr

Dlouhodobé uchování digitálních dokumentů patří mezi největší výzvy, jakým paměťové instituce v současnosti čelí. Ani při nejlepší vůli není zpravidla možné dlouhodobě uchovávat vše v plném rozsahu. Proprietárnost formátů, šifrování obsahu, vkládání rizikových formátů, odkazování na externí obsah nebo chybějící fonty, to jsou jen některé z objektivních důvodů, které mohou uchování autentických a kompletních dokumentů do budoucna ohrozit. V zájmu zajištění uchování maxima možného, při současném zachování přiměřené výše nákladů, proto digitální archivy na celém světě vydávají své formátové politiky s preferovanými a akceptovanými formáty, které jim mají v jejich snažení pomoci.

Příliš benevolentní formátová politika knihovny může vést v dlouhodobém horizontu k významnému růstu rizik a nákladů na dlouhodobé uchování, v krátkodobé perspektivě nicméně pomáhá pokrýt vysoké procento vydávané produkce. Přísná formátová politika na druhou stranu náklady na dlouhodobé uchování pomáhá snižovat a v míře, v jaké si jí dokážou osvojit nakladatelé, může výrazně napomoci samotné dlouhodobé uchovatelnosti odevzdaných dat pro budoucí generace. Příliš přísná formátová politika nicméně vede nutně k nižšímu pokrytí vydávané produkce, horšímu vztahu knihoven s nakladateli a může přinášet také rizika vyplývající z nekvalitně provedených konverzí, kterými se nakladatelé mohou snažit formálně naplnit příliš přísně nastavené podmínky. V praxi je tedy potřeba hledat kompromisní řešení. Zahraniční příklady, zachycené v této práci, ukazují vesměs na volnější formátové politiky knihoven proti nakladatelům, než jak je má v současné době nastaveny NK ČR.

V předkládané práci jsem provedl určení relevantních formátů pro uchování elektronických publikací vydávaných v ČR. Jako dominantní formáty byly zjištěny EPUB a PDF, jako méně významné formáty hodné zřetele formáty MOBI, iBooks a Woody. Provedl jsem rekapitulaci základních kritérií pro výběr vhodných archivačních formátů. V kapitolách věnovaných DRM a vybraným souborovým formátům jsem definoval některá nejvýznamnější rizika pro uchování elektronických publikací a určil nejvhodnější archivační formáty. V druhé části práce jsem se ve zpracovaných případových studiích zaměřil na porovnání zjištěných poznatků s postupy v zahraničních institucích, provozujících vlastní systémy pro příjem e-publikací. V poslední kapitole práce jsem provedl zhodnocení kvality dat, dobrovolně odevzdaných nakladateli do NK ČR. Kvalita analyzovaných dat se nevymyká srovnatelným zahraničním zkušenostem, závažná je nicméně přítomnost DRM u více než 5 % souborů.

Na základě zjištěných skutečností a příkladů zahraniční praxe navrhuji v souladu se zadáním bakalářské práce následující úpravy formátové politiky Národní knihovny ČR pro příjem elektronických publikací. Za první krok považuji potřebné rozšíření formátové politiky ze dvou stávajících kategorií (preferované a ostatní formáty) na tři kategorie, tvořené formáty

preferovanými, akceptovanými a neakceptovanými. Zatímco preferovaným formátům by měla být garantována dlouhodobá logická ochrana v plném rozsahu, akceptované formáty by měly být uchovávány v původních formátech bez garance dlouhodobé logické ochrany.

Za preferované formáty navrhuji EPUB 2, EPUB 3, PDF/A-1 a PDF/A-2. V případě formátu EPUB 3 by soubory měly odpovídat definovanému profilu, který by zahrnoval zejména zákaz JavaScriptu a vložených multimediálních souborů nebo jiných cizích objektů. V případě nedodržení podmínek by byl takový soubor posuzován pouze jako akceptovaný.

Jako akceptované formáty navrhuji ostatní varianty PDF a také EPUB, neodpovídající předepsanému profilu. Všechny odevzdané soubory by měly být validní vůči formátovým specifikacím a neobsahovat DRM. V případě odevzdání nevalidních souborů by mělo být na posouzení NK ČR, zda jsou publikace ve stavu, aby je bylo možné dlouhodobě uchovat. Při příjmu akceptovaných formátů navrhuji zvážit provedení normalizace pouze mezi formáty PDF a PDF/A. Zachován by měl být vždy originální formát. Nabízí se proto možnost vyžití např. formátu PDF/A-3, který by původní soubor PDF dokázal přímo uchovat jako svou součást.

Nastavení pravidel pro budoucí povinné odevzdávání e-publikací do NK ČR by mělo zajistit v maximální možné míře odevzdávání publikací v preferovaných formátech. Teprve v případech, kdy charakter obsahu produkci v preferovaném formátu vylučuje, nebo by nakladateli způsobovala nepřiměřené náklady, mohlo by na základě souhlasu NK ČR dojít k odevzdání v akceptovaném formátu. Za rizikové lze nicméně považovat také neodborně prováděné konverze z formátu PDF do PDF/A. Snahou NK ČR by tedy mělo být těmto manipulacím předcházet a zejména se snažit nakladatele motivovat, aby PDF/A v potřebné kvalitě (nejlépe PDF/A-a, nebo PDF/A-u) dokázali přímo exportovat z používaných aplikací pro sazbu dokumentů. K tomu by jim NK ČR měla poskytovat potřebné metodické materiály, pokrývající nejčastěji používané verze těchto aplikací. Jestliže analýza trhu, provedená v rámci kapitoly 2.4, ukázala, že pouhých 10 nakladatelů je zodpovědných za plných 65 % vydaných e-knih, může být takováto úzká spolupráce NK ČR s nakladateli efektivní.

Nadále by neměly být přijímány zejména publikace, vydané v proprietárních formátech, jako jsou Wooky či iBooks. V žádném případě to ale neznamená rezignaci na jejich uchování. Přestože existují a nadále budou existovat proprietární formáty, z ekonomických důvodů se v současné době jedná nejčastěji o pouhá proprietární rozšíření formátu EPUB 3. Je tomu tak také v případě Wooky i iBooks, oba tyto formáty tedy ve většině případů bude nakladatel schopen poskytnout v podobě formátu EPUB. Analogicky by mělo být postupováno také v případě formátu MOBI. Pokud byla publikace vydána také v preferovaném nebo akceptovaném formátu, od odevzdání v nepůvodním formátu by mělo být upuštěno.

## Bibliografie

- [1] STIRLING, Peter, Gildas ILLIEN, Pascal SANZ a Sophie SEPETJAN. The state of e-legal deposit in France: Looking back at five years of putting new legislation into practice and envisioning the future. *IFLA Journal* [online]. SAGE, 2012, **38**(1), 5-24 [cit. 2018-05-19]. Dostupné z: [https://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-38-1\\_2012.pdf](https://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-38-1_2012.pdf)
- [2] REITZ, Joan M. *ODLIS: Online Dictionary for Library and Information Science* [online]. Santa Barbara: ABC-CLIO, ©2004-2014 [cit. 2018-07-01]. Dostupné z: [https://www.abc-clio.com/ODLIS/odlis\\_e.aspx](https://www.abc-clio.com/ODLIS/odlis_e.aspx)
- [3] SVOBODA, Tomáš et al. *Elektronické publikace v Národní knihovně ČR* [online]. Praha: Národní knihovna České republiky, 2015, 93 s. [cit. 2018-04-29]. ISBN 978-80-7050-654-7. Dostupné z: <https://drive.google.com/file/d/0B46gpfbHV70tR28wSzYyUmxqOHc/view>
- [4] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. *Webarchiv: památník českého internetu* [online]. Praha: Národní knihovna ČR, [2001-] [cit. 2018-07-22]. Dostupné z: <https://www.webarchiv.cz/cs/>
- [5] Comparison of e-book formats. *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001-, Last edited 9 March 2018 [cit. 2018-04-01]. Dostupné z: [https://en.wikipedia.org/wiki/Comparison\\_of\\_e-book\\_formats](https://en.wikipedia.org/wiki/Comparison_of_e-book_formats)
- [6] WHIBLEY, Simon, Peter MAY a Paul WHEATLEY. EBook Summary Assessment. In: *Digital Preservation Coalition Wiki* [online]. British Library, 2017 [cit. 2018-03-30]. Dostupné z: [http://wiki.dpconline.org/images/0/04/EBook\\_Summary\\_Assessment\\_v1.0.pdf](http://wiki.dpconline.org/images/0/04/EBook_Summary_Assessment_v1.0.pdf)
- [7] E-book formats. In: *MobileRead Wiki* [online]. MobileRead, 2006, last modified 27 March 2018 [cit. 2018-04-07]. Dostupné z: [https://wiki.mobileread.com/wiki/E-book\\_formats](https://wiki.mobileread.com/wiki/E-book_formats)
- [8] PECINA, Martin. *Knihy a typografie*. Vydání třetí, rozšířené. Brno: Host, 2017. ISBN 978-80-7577-040-0.
- [9] ANTOL, Filip. *Matematika na čtečkách e-knih* [online]. Brno, 2015 [cit. 2018-05-18]. Dostupné z: <https://is.muni.cz/th/ptckg/>. Bakalářská práce. Masarykova univerzita, Přírodovědecká fakulta, Ústav matematiky a statistiky. Vedoucí práce RNDr. Roman Plch, Ph.D.
- [10] NAKLADATELSTVÍ FRAUS. Interaktivní učebnice z Nakladatelství Fraus. *Fraus* [online]. ©2018 [cit. 2018-04-04]. Dostupné z: <https://www.fraus.cz/cs/nezavisle-stranky/i-ucebnice>
- [11] HEJNÝ, Jaroslav. *Manuál pro práci s elektronickými učebnicemi* [online]. Brno: Code Creator, 2015 [cit. 2018-04-04]. Dostupné z: [http://www.soubosonohy.cz/files/Petr%20Foto%20manual\\_web.pdf](http://www.soubosonohy.cz/files/Petr%20Foto%20manual_web.pdf)
- [12] PESEK, Igor, Blaž ZMAZEL, Darko DRAKULIČ a Eva ZMAZEK. Evolution from e-learning materials to the i-textbooks in Slovenia. In: *Proceedings of the 12th international conference : the future of mathematics education in a connected world, 21-26 September 2014, Herceg Novi, Montenegro* [online]. Herceg Novi, 2014, s. 1-7 [cit. 2018-04-04]. Dostupné z: <http://directorymathsed.net/montenegro/Pesek.pdf>
- [13] End of Life Support for Adobe Flash Player. How to play video in PDF. In: *Adobe Forum Help & Resources* [online]. Adobe Systems, 2018, poslední reakce dne 17.1.2018 [cit. 2018-07-01]. Dostupné z: <https://forums.adobe.com/thread/2435724>
- [14] LEVINE, Bob. It's okay to say no to interactive PDF. *Bob Levine* [online]. January 26th, 2016 [cit. 2018-07-01]. Dostupné z: <https://www.boblevine.us/its-okay-to-say-no-to-interactive-pdf/>
- [15] TEMEX. *Elektronické učebnice: popis systému, základních funkcí a jejich cena* [online]. Střední škola stavebních řemesel Brno-Bosonohy, [2015], s. 2-4 [cit. 2018-04-04]. Dostupné z: [www.el-ucebnice.cz/ostatni/manual.pdf](http://www.el-ucebnice.cz/ostatni/manual.pdf)

- [16] TAYLOR, Martin. How to produce illustrated, multimedia ebooks. *DigitalPublishing101* [online]. Sep 30, 2012 [cit. 2018-04-04]. Dostupné z: <http://digitalpublishing101.com/how-to-produce-illustrated-multimedia-ebooks/>
- [17] CARPE DIEM. *Carpe diem online: ePub3* [online]. ©2015 [cit. 2018-04-04]. Dostupné z: <http://eshop.carpe.cz/50-epub3>
- [18] JEŘÁBEK, Antonín. ISBN a online publikace. *Ikaros* [online]. 2011, **15**(1) [cit. 2018-04-16]. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/13593>
- [19] JEŘÁBEK, Antonín. *Průručka uživatele systému ISBN* [online]. 5. vyd. Praha: Národní knihovna České republiky, 2006, revize března 2011 [cit. 2018-04-16]. Dostupné z: <https://www.nkp.cz/soubory/ostatni/prirucka-isbi-13.pdf>
- [20] SVOBODA, Tomáš, ed. *Správa elektronických publikací v síti knihoven ČR: roční periodická zpráva projektu za rok 2012* [online]. Verze 1.00. Praha: Národní knihovna ČR, 2012 [cit. 2018-05-20]. Dostupné z: interní dokument
- [21] TUREČKOVÁ, Marcela, ed. *Zpráva o českém knižním trhu 2016/2017* [online]. Praha: Svaz českých knihkupců a nakladatelů, 2017 [cit. 2018-07-01]. ISBN 978-80-907020-0-4. Dostupné z: [https://ipk.nkp.cz/docs/Zprava\\_o\\_ceskem\\_kniznim\\_trhu\\_201617.pdf](https://ipk.nkp.cz/docs/Zprava_o_ceskem_kniznim_trhu_201617.pdf)
- [22] TUREČKOVÁ, Marcela, ed. *Zpráva o českém knižním trhu 2014/2015* [online]. Praha: Svaz českých knihkupců a nakladatelů, 2015 [cit. 2018-07-01]. ISBN 978-80-905680-4-4. Dostupné z: <https://wayback.webarchiv.cz/wayback/20160427223514/http://sckn.cz/content/zpravy/file-1159.pdf>
- [23] KOHOUT, Pavel. Kdo vládne světu e-knih? A jak to bude dál?. *Kohoutp.cz* [online]. 31.10.2017 [cit. 2018-04-28]. Dostupné z: <http://www.kohoutp.cz/vladne-svetu-e-knih-bude-dal/>
- [24] PELECH, Tadeáš. Wooky začíná nabízet obsah i pro čtečky Kindle. *Computerworld* [online]. Praha: IDG Czech Republic, 01.11.2012 [cit. 2018-04-28]. Dostupné z: <https://computerworld.cz/aktuality/wooky-zacina-nabizet-obsah-i-pro-ctecky-kindle-49033>
- [25] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. Úvod. *Dobrovolné uložení elektronických publikací* [online]. Praha: Národní knihovna České republiky, ©2011 [cit. 2018-04-16]. Dostupné z: <http://edpilot.nkp.cz/wp/>
- [26] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. *Smlouva o poskytování elektronických publikací* [online]. Praha: Národní knihovna České republiky, [2011] [cit. 2018-07-22]. Dostupné z: [http://edpilot.nkp.cz/wp/smlouva\\_epublikace.pdf](http://edpilot.nkp.cz/wp/smlouva_epublikace.pdf)
- [27] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. Technická specifikace. *Dobrovolné uložení elektronických publikací* [online]. Praha: Národní knihovna České republiky, ©2011 [cit. 2018-04-16]. Dostupné z: [http://edpilot.nkp.cz/wp/?page\\_id=14](http://edpilot.nkp.cz/wp/?page_id=14)
- [28] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. Národní digitální knihovna. *Národní knihovna České republiky* [online]. Praha, 27.05.2014 [cit. 2018-07-05]. Dostupné z: <https://www.nkp.cz/o-knihovne/zakladni-informace/narodni-digitalni-knihovna>
- [29] ŽÍŽALA, Martin. Správa elektronických publikací v síti knihoven ČR. *Ikaros* [online]. 2012, **16**(4) [cit. 2018-04-16]. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/13849>
- [30] VYCHODIL, Bedřich a Václav JIROUŠEK. Technické analýzy formátů e-publikací. In: *Správa elektronických publikací v síti knihoven ČR: roční periodická zpráva projektu za rok 2012* [online]. Verze 1.00. Praha: Národní knihovna ČR, 2012, s. 25-44 [cit. 2018-04-06]. Dostupné z: Interní dokument
- [31] KVAŠOVÁ, Zuzana a Tomáš SVOBODA. Dlouhodobá ochrana elektronických publikací. *ProInflow: Časopis pro informační vědy* [online]. 2013, **5**(2), 120-127 [cit. 2018-04-16]. Dostupné z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/775>



- [32] ČESKO. Zákon č. 37/1995 Sb. ze dne 8. února 1995 o neperiodických publikacích. In: *Sbírka zákonů*. 1995, částka 8. Dostupné také z: <https://www.zakonyprolidi.cz/cs/1995-37>
- [33] ČESKO. Zákon č. 46/2000 Sb. ze dne 22. února 2000 o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon). In: *Sbírka zákonů*. 2000, částka 17. Dostupné také z: <https://www.zakonyprolidi.cz/cs/2000-46>
- [34] ČESKO. Zákon č. 257/2001 Sb. ze dne 29. června 2001 o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon). In: *Sbírka zákonů*. 2001, částka 98. Dostupné také z: <https://www.zakonyprolidi.cz/cs/2001-257>
- [35] RICHTER, Vít. Příprava legislativy povinného výtisku elektronických publikací v České republice. In: *Infos 2017: zborník príspevkov z 39. medzinárodného informatického sympózia: 25. - 28. apríl 2017* [online]. 1. Bratislava: Spolok slovenských knihovníkov a knižníc, 2017, s. 5-12 [cit. 2018-04-28]. ISBN 978-80-89586-09-7. Dostupné z: [http://www.infolib.sk/files/infos\\_2017\\_prezentacie/sskk\\_zbornik\\_symposium-infos-2017.pdf](http://www.infolib.sk/files/infos_2017_prezentacie/sskk_zbornik_symposium-infos-2017.pdf)
- [36] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. *Návrh legislativního řešení projektu „NAKI - Správa elektronických publikací v síti knihoven ČR“* [online]. Praha: Národní knihovna ČR, 2014, verze 12. 5. 2013, doplněno na základě výsledku jednání 4. 2. 2014 [cit. 2018-06-30]. Dostupné z: [http://www.nkp.cz/soubory/ostatni/elektronickypovinnvytisk\\_navrh.docx](http://www.nkp.cz/soubory/ostatni/elektronickypovinnvytisk_navrh.docx)
- [37] ČESKO. PARLAMENT. POSLANECKÁ SNĚMOVNA. *Sněmovní tisk 1011/0, část č. 1/9 Novela z. - knihovní zákon* [online]. Praha: Parlament České republiky, 2017 [cit. 2018-04-28]. Dostupné z: <http://www.psp.cz/sqw/text/tiskt.sqw?O=7&CT=1011&CT1=0>
- [38] ŽÍŽALA, Martin. "E-deposit": elektronické publikace v Národní knihovně ČR – 1. část. *Čtenář: měsíčník pro knihovny* [online]. Kladno, 2017, **69**(7-8), 263-264 [cit. 2018-07-05]. Dostupné z: <https://ctenar.svkkk.cz/clanky/2017-roc-69/0708-2017/vyuziti-novych-technologii-„e-deposit“-elektronicke-publikace-v-narodni-knihovne-cr---1-cast-172-2552.htm>
- [39] ČESKO. MINISTERSTVO KULTURY. Závěrečná zpráva z hodnocení dopadů regulace (RIA). In: *Sněmovní tisk 1011/0, část č. 1/9 [novela knihovního zákona]* [online]. b.r. [cit. 2018-04-29]. Dostupné z: <http://www.psp.cz/doc/00/12/05/00120591.pdf>
- [40] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. Standardy pro elektronické publikace. *NDK: Národní digitální knihovna* [online]. Praha: Národní knihovna České republiky, 15.12.2017 [cit. 2018-07-18]. Dostupné z: <https://www.ndk.cz/standardy-digitalizace/standardy-pro-elektronicke-dokumenty>
- [41] CUBR, Ladislav. Dlouhodobá ochrana digitálních dokumentů. In: *TDKIV: Česká terminologická databáze knihovnictví a informační vědy* [online]. Praha: Národní knihovna ČR, ©2014 [cit. 2018-06-26]. Dostupné z: [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000014623&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000014623&local_base=KTD)
- [42] HUTAŘ, Jan, Andrea MIRANDA, Eliška PAVLÁSKOVÁ, Zdeněk VAŠEK a Zdeněk HRUŠKA. *Metodika logické ochrany digitálních dat* [online]. ARCLib, 2018 [cit. 2018-06-26]. Dostupné z: <http://hdl.handle.net/11104/0282107>
- [43] MCLELLAN, Evelyn. *Preservation planning: Normalizing file formats* [online]. [2010] [cit. 2018-06-26]. Dostupné z: <https://www.archivematica.org/download/EvelynMcLellan-PreservationPlanning.pdf>
- [44] CUBR, Ladislav. *Autenticita a digitální informace*. Praha, 2017, 227 s. Dostupné také z: <https://is.cuni.cz/webapps/zzp/detail/105596/>. Disertační práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Prof. RNDr. Jirí Ivánek, CSc.
- [45] CUBR, Ladislav. *Dlouhodobá ochrana digitálních dokumentů*. Praha: Národní knihovna ČR, 2010, 154 s. ISBN 978-80-7050-588-5.

- [46] PENNOCK, Maureen, Paul WHEATLEY a Peter MAY. Sustainability Assessments at the British Library: Formats, Frameworks, & Findings. In: *Proceedings of the 11th International Conference on Digital Preservation, iPRES 2014, Melbourne, Australia, October 6 - 10, 2014* [online]. Melbourne: State Library of Victoria, 2014, s. 141–149 [cit. 2018-03-31]. ISBN 978-0-642-27881-4. Dostupné z: <https://fedora.phaidra.univie.ac.at/fedora/get/o:378110/bdef:Content/get>
- [47] LIBRARY OF CONGRESS. *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, b.r., last updated 10-Mar-2017 [cit. 2018-04-14]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/index.shtml>
- [48] LIBRARY OF CONGRESS. *Recommended Formats Statement: 2017-2018* [online]. Washington (DC): Library of Congress, [2017] [cit. 2018-07-01]. Dostupné z: <https://www.loc.gov/preservation/resources/rfs/RFS%202017-2018.pdf>
- [49] BRITISH LIBRARY. File Formats Assessments. *Digital Preservation Coalition Wiki* [online]. British Library, b.r., last modified on 9 February 2018 [cit. 2018-04-14]. Dostupné z: [http://wiki.dpconline.org/index.php?title=File\\_Formats\\_Assessments](http://wiki.dpconline.org/index.php?title=File_Formats_Assessments)
- [50] HARVARD LIBRARY. Format Assessments. *Harvard Wiki* [online]. Harvard Library, ©2017-2018, last modified Apr 15, 2016 [cit. 2018-04-14]. Dostupné z: <https://wiki.harvard.edu/confluence/display/digitalpreservation/Format+Assessments>
- [51] KOST-CECO. Katalog archivischer Dateiformate. *KOST: Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen* [online]. Version 5.1. Bern: KOST-CECO, ©2005-2018 [cit. 2018-07-01]. Dostupné z: <https://kost-ceco.ch/wiki/whelp/KaD/index.php?ld>
- [52] TI/A STANDARD INITIATIVE. *TI/A Standard Initiative* [online]. ©2016 [cit. 2018-07-22]. Dostupné z: <http://ti-a.org/>
- [53] KING'S COLLEGE LONDON. Completed projects A-Z: Investigating the Significant Properties of Electronic Content Over Time (InSPECT). *King's College London* [online]. London, ©2018 [cit. 2018-07-01]. Dostupné z: <https://www.kcl.ac.uk/innovation/groups/cerch/research/projects/completed/inspect.aspx>
- [54] GRACE, Stephen, Gareth KNIGHT a Lynne MONTAGUE. *InSPECT: Final Report* [online]. 2009, 27 s. [cit. 2018-04-26]. Dostupné z: <https://web.archive.org/web/20120617000806/http://www.significantproperties.org.uk/inspect-finalreport.pdf>
- [55] HAVLOVÁ, Jaroslava. Správa digitálních práv. In: *TDKIV: Česká terminologická databáze knihovnictví a informační vědy* [online]. Praha: Národní knihovna ČR, ©2014 [cit. 2018-04-24]. Dostupné z: [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000015940&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000015940&local_base=KTD)
- [56] BLÄSI, Christoph a Franz ROTHLAUF. *On the Interoperability of eBook Formats* [online]. Brussels: European and International Booksellers Federation, 2013 [cit. 2018-03-30]. Dostupné z: <https://www.booksellers.org.uk/BookSellers/BizFormFiles/936121cb-a426-46da-b9aa-db8bd285d21e.pdf>
- [57] VÁCLAVÍK, Lukáš. Malá revoluce. Wooky vzala na milost sociální DRM a zasílá knihy i na Kindle. *Cnews.cz* [online]. Mladá Fronta, 2012 [cit. 2018-04-16]. Dostupné z: <https://www.cnews.cz/mala-revoluce-wooky-vzala-na-milost-socialni-drm-a-zasila-knihy-i-na-kindle/>
- [58] KOLÁŘ, Martin. *Analýza současných přístupů k DRM u elektronických knih a návrh alternativního přístupu k DRM*. Praha, 2015. Dostupné také z: <https://vskp.vse.cz/eid/44991>. Diplomová práce. Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra systémové analýzy. Vedoucí práce Mgr. Ing. Tomáš Sigmund, Ph.D.
- [59] HUNDACKER, Helge. Forensic Digital Rights Management. In: RÜDIGER, Grimm a Suzanne GUTH. *The 6th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods incorporating The 4th International ODRL Workshop, October 16-18, 2008, Poznań, Poland* [online]. Poznań: Poznań University of Economics, 2008, s. 161-175 [cit.

- 2018-04-24]. Dostupné z:  
<https://pdfs.semanticscholar.org/16ba/5f9362bb75ef43a51b18e4cb5e540a48341a.pdf>
- [60] HEIN, Stefan a Tobias STEINKE. DRM and digital preservation: A use case at the German National Library. In: *Proceedings of the 11th International Conference on Digital Preservation, iPRES 2014, Melbourne, Australia, October 6 - 10, 2014* [online]. Melbourne: State Library of Victoria, 2014, s. 228-231 [cit. 2018-04-07]. ISBN 978-0-642-27881-4. Dostupné z:  
[http://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version\\_1.pdf](http://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version_1.pdf)
- [61] LÉR, Martin. EReading odstartoval prodej digitálních knih. *Lupa.cz* [online]. Internet Info, 9. 9. 2010 [cit. 2018-06-25]. Dostupné z: <https://www.lupa.cz/clanky/ereading-odstartoval-prodej-digitalnich-knih/>
- [62] KAUR, Kirnn, Stefan HEIN, Sabine SCHRIMPF a Manuela HOLZMAYER. *Report on DRM preservation* [online]. Version 1.4. APARSEN, 2014, 85 s. [cit. 2018-04-07]. Dostupné z:  
[http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D31\\_1-01-1\\_4\\_incURN.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/sites/7/downloads/2014/06/APARSEN-REP-D31_1-01-1_4_incURN.pdf)
- [63] CARR, Miles, ed. *Information Security: Principles and Practices* [online]. New York: Library Press, 2018 [cit. 2018-04-24]. ISBN 978-1-9789-1262-5. Dostupné z:  
<http://ebooks.wtbooks.com/static/wtbooks/ebooks/9781978912625/9781978912625.pdf>
- [64] ADOBE SYSTEMS. Adobe® Digital Editions 1.7.1 Release Notes. *Adobe* [online]. San Jose: Adobe Systems, 7 April 2009 [cit. 2018-04-24]. Dostupné z: <https://helpx.adobe.com/digital-editions/release-note/digital-editions-1-7-1.html>
- [65] SCOTT, Jared. The End of Microsoft Reader. *EBook Reader Software* [online]. August 16, 2011 [cit. 2018-04-24]. Dostupné z: <https://ebookreadersoftware.wordpress.com/2011/08/16/the-end-of-microsoft-reader/>
- [66] WILLIAMS, Anastasiia. Epubor DRM Removal tool that helps you to read your ebooks anywhere. *TechPcVipers* [online]. TechPcVipers, June 26, 2017, last updated on: April 17, 2018 [cit. 2018-04-25]. Dostupné z: <https://www.techpcvipers.com/epubor-drm-removal-tool-that-helps-you-to-read-your-ebooks-anywhere/>
- [67] RING, Bennett. How to remove ebook DRM with Calibre. *TechRadar* [online]. TechRadar, July 31, 2014 [cit. 2018-04-25]. Dostupné z: <https://www.techradar.com/news/software/how-to-remove-ebook-drm-with-calibre-1291960>
- [68] ZHANG, Xiao. *A Survey of Digital Rights Management Technologies* [online]. Last modified November 28, 2011 [cit. 2018-04-25]. Dostupné z: <http://www.cse.wustl.edu/~jain/cse571-11/ftp/drm/#sec2.3>
- [69] KIRCHHOFF, Amy a Sheila MORRISSEY. *Preserving eBooks: DPC Technology Watch Report 14-01* [online]. Great Britain: Digital Preservation Coalition, 2014 [cit. 2018-03-30]. DOI: <http://dx.doi.org/10.7207/twr14-01>. ISSN 2048-7916. Dostupné z:  
<https://www.dpconline.org/docs/technology-watch-reports/1230-dpctw14-01/file>
- [70] INTERNATIONAL DIGITAL PUBLISHING FORUM. Open eBook Publication Structure Specification History. *International Digital Publishing Forum* [online]. International Digital Publishing Forum, [2006] [cit. 2018-04-08]. Dostupné z:  
<http://web.archive.org/web/20101212183133/http://idpf.org/oebps/history.htm>
- [71] LIBRARY OF CONGRESS. OEBPS (Open Ebook Forum Publication Structure) 1.0. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, b.r., last updated: 21-Feb-2017 [cit. 2018-04-08]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000054.shtml>
- [72] OPEN EBOOK AUTHORIZING GROUP. *Open eBook™ Publication Structure 1.0* [online]. Open eBook Authoring Group, September 16, 1999 [cit. 2018-04-02]. Dostupné z:  
<http://web.archive.org/web/20101212183603/http://idpf.org/oebps/oebps1.0/download/oeb1.pdf>

- [73] LIBRARY OF CONGRESS. EPUB, Electronic Publication, Version 2. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, Last Updated 06/18/2018 [cit. 2018-06-23]. Dostupné z: <http://www.loc.gov/preservation/digital/formats/fdd/fdd000278.shtml>
- [74] RUSSELL, Bob. IDPF invites input on new e-book standard file format (OPS 2.0) [příspěvek v diskuzním fóru]. In: *MobileRead Forum* [online]. MobileRead, 04-16-2007, 07:35 PM [cit. 2018-04-02]. Dostupné z: <https://www.mobileread.com/forums/showthread.php?t=10408>
- [75] INTERNET ARCHIVE. Internet Archive: Wayback Machine. *Internet Archive* [online]. San Francisco: Internet Archive, [2001-] [cit. 2018-07-22]. Dostupné z: <https://archive.org/web/>
- [76] Looking for OEB reader and converter [příspěvek v diskuzním fóru]. In: *MobileRead Forums* [online]. MobileRead, 04-27-2013, 02:30 PM, poslední aktualizace 08-01-2013, 12:21 PM [cit. 2018-04-08]. Dostupné z: <https://www.mobileread.com/forums/showthread.php?p=2583654>
- [77] DAY, Michael, David RUSSO a Simon WHIBLEY. Mobipocket Format Preservation Assessment. In: *Digital Preservation Coalition Wiki* [online]. British Library, 2018 [cit. 2018-03-30]. Dostupné z: [http://wiki.dpconline.org/images/4/4c/Mobipocket\\_Assessment\\_v1.pdf](http://wiki.dpconline.org/images/4/4c/Mobipocket_Assessment_v1.pdf)
- [78] HOFFELDER, Nate. RIP: Mobipocket 2000-2011. *The Digital Reader* [online]. 2 November, 2011 [cit. 2018-04-08]. Dostupné z: <https://the-digital-reader.com/2011/11/02/rip-mobipocket-2000-2011/>
- [79] HOFFELDER, Nate. Mobipocket to Turn Off Its Website and Servers on 31 October. *The Digital Reader* [online]. 23 October, 2016 [cit. 2018-06-19]. Dostupné z: <https://the-digital-reader.com/2016/10/23/mobipocket-turn-off-website-servers-31-october/>
- [80] AZW. *MobileRead Wiki* [online]. MobileRead, 2007, last modified 26 March 2018 [cit. 2018-04-21]. Dostupné z: <https://wiki.mobileread.com/wiki/AZW>
- [81] KF8. *MobileRead Wiki* [online]. MobileRead, 2012, last modified 10 October 2017 [cit. 2018-04-22]. Dostupné z: <https://wiki.mobileread.com/wiki/KF8>
- [82] MOBIPOCKET. What is the Mobipocket file format. *Mobipocket.com* [online]. ©2000-2007 [cit. 2018-04-02]. Dostupné z: <https://web.archive.org/web/20160414103204/http://www.mobipocket.com/dev/article.asp?BaseFolder=prcgen&File=mobiformat.htm>
- [83] PalmDOC. *MobileRead Wiki* [online]. MobileRead, 2007, last modified 9 January 2018 [cit. 2018-04-22]. Dostupné z: <https://wiki.mobileread.com/wiki/PalmDOC>
- [84] MOBI. In: *MobileRead Wiki* [online]. MobileRead, 2007, last modified 26 March 2018 [cit. 2018-04-08]. Dostupné z: <https://wiki.mobileread.com/wiki/Mobi>
- [85] KindleUnpack. *MobileRead Wiki* [online]. MobileRead, 2012, last modified 9 August 2017 [cit. 2018-04-08]. Dostupné z: <https://wiki.mobileread.com/wiki/KindleUnpack>
- [86] MOBIPOCKET. Security/Encryption Options. *Mobipocket.com* [online]. ©2000-2007 [cit. 2018-04-08]. Dostupné z: <https://web.archive.org/web/20150907235354/http://www.mobipocket.com/dev/article.asp?BaseFolder=creatorpublisher&File=security.htm>
- [87] DAY, Michael, Peter MAY, Paul WHEATLEY, Simon WHIBLEY a Johan VAN DER KNIJFF. EPUB Format Preservation Assessment. In: *Digital Preservation Coalition Wiki* [online]. Version 1.2. British Library, 2015 [cit. 2018-03-30]. Dostupné z: [http://wiki.dpconline.org/images/a/a9/EPUB\\_Assessment\\_v1.2.pdf](http://wiki.dpconline.org/images/a/a9/EPUB_Assessment_v1.2.pdf)
- [88] INTERNATIONAL DIGITAL PUBLISHING FORUM. Older Versions of EPUB. *IDPF* [online]. 2017 [cit. 2018-04-21]. Dostupné z: <http://idpf.org/epub-older-versions>

- [89] INTERNATIONAL DIGITAL PUBLISHING FORUM. EPUB 2.0.1. *IDPF* [online]. International Digital Publishing Forum, 2017 [cit. 2018-04-21]. Dostupné z: <http://idpf.org/epub/201>
- [90] INTERNATIONAL DIGITAL PUBLISHING FORUM. Open Container Format (OCF) 2.0.1 v1.0.1. *IDPF* [online]. 2010 [cit. 2018-04-21]. Dostupné z: [http://www.idpf.org/doc\\_library/epub/OCF\\_2.0.1\\_draft.doc](http://www.idpf.org/doc_library/epub/OCF_2.0.1_draft.doc)
- [91] INTERNATIONAL DIGITAL PUBLISHING FORUM. Open Packaging Format (OPF) 2.0.1 v1.0.1. *IDPF* [online]. 2010 [cit. 2018-04-21]. Dostupné z: [http://www.idpf.org/epub/20/spec/OPF\\_2.0.1\\_draft.htm#Section2.3.1.1](http://www.idpf.org/epub/20/spec/OPF_2.0.1_draft.htm#Section2.3.1.1)
- [92] INTERNATIONAL DIGITAL PUBLISHING FORUM. Open Publication Structure (OPS) 2.0.1 v1.0.1. *IDPF* [online]. 2010 [cit. 2018-04-21]. Dostupné z: [http://www.idpf.org/epub/20/spec/OPS\\_2.0.1\\_draft.htm](http://www.idpf.org/epub/20/spec/OPS_2.0.1_draft.htm)
- [93] LIBRARY OF CONGRESS. ZIP File Format, Version 6.3.3 (PKWARE). In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, Last Updated 27-Feb-2017 [cit. 2018-06-23]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000362.shtml>
- [94] KNIJFF, Johan. *EPUB for archival preservation* [online]. Version for external distribution. National Library of the Netherlands, 20 July 2012 [cit. 2018-04-21]. Dostupné z: <http://openpreservation.org/system/files/epubForArchivalPreservation20072012ExternalDistribution.pdf>
- [95] WHEATLEY, Paul. EPUB Format Assessment. In: *Harvard Wiki* [online]. V04. Harvard Library, [2014] [cit. 2018-03-30]. Dostupné z: [https://wiki.harvard.edu/confluence/download/attachments/204385883/Format profile - ePUB v04.docx?version=1&modificationDate=1459873793000&api=v2](https://wiki.harvard.edu/confluence/download/attachments/204385883/Format%20profile%20-%20ePUB%20v04.docx?version=1&modificationDate=1459873793000&api=v2)
- [96] OPEN PRESERVATION FOUNDATION. Flint. *Open Preservation Foundation* [online]. [2014] [cit. 2018-04-22]. Dostupné z: <http://flint.openpreservation.org/>
- [97] SCAPE PROJECT TEAM. *SCAPE: Scalable Preservation Environments* [online]. [2011-2014] [cit. 2018-07-22]. Dostupné z: <http://scape-project.eu/>
- [98] GITHUB. Openpreserve / flint: Releases. *GitHub* [online]. [San Francisco]: GitHub, ©2018 [cit. 2018-07-21].
- [99] KNIJFF, Johan van der. Policy-based assessment of EPUB with Epubcheck. In: *KB RESEARCH: Research at the National Library of the Netherlands* [online]. Koninklijke Bibliotheek, March 13, 2015 [cit. 2018-04-16]. Dostupné z: <http://blog.kbresearch.nl/2015/03/13/policy-based-assessment-of-epub-with-epubcheck/>
- [100] INTERNATIONAL DIGITAL PUBLISHING FORUM. EPUB. *IDPF* [online]. 2017 [cit. 2018-04-22]. Dostupné z: <http://idpf.org/epub>
- [101] INTERNATIONAL DIGITAL PUBLISHING FORUM. EPUB 3.0 Published as ISO Technical Specification. *IDPF* [online]. December 21, 2014 [cit. 2018-05-19]. Dostupné z: <http://idpf.org/news/epub-30-published-as-iso-technical-specification>
- [102] INTERNATIONAL DIGITAL PUBLISHING FORUM. EPUB 3.0. *IDPF* [online]. International Digital Publishing Forum, 2017 [cit. 2018-04-22]. Dostupné z: <http://idpf.org/epub/30>
- [103] INTERNATIONAL DIGITAL PUBLISHING FORUM. EPUB 3.0.1 Changes from EPUB 3.0. *IDPF* [online]. International Digital Publishing Forum, 26 June 2014 [cit. 2018-04-22]. Dostupné z: <http://www.idpf.org/epub/301/spec/epub-changes.html>
- [104] LIBRARY OF CONGRESS. EPUB, Electronic Publication, Version 3. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of

- Congress, Last Updated 06/18/2018 [cit. 2018-06-21]. Dostupné z: <http://www.loc.gov/preservation/digital/formats/fdd/fdd000308.shtml>
- [105] KNIJFF, Johan van der. EPUB for archival preservation: an update. In: *Open Preservation Foundation* [online]. Open Preservation Foundation, 23rd May 2013 [cit. 2018-05-19]. Dostupné z: <http://openpreservation.org/blog/2013/05/23/epub-archival-preservation-update/>
- [106] BJARNASON, Baldur. EPUB javascript security. In: *Baldur Bjarnason* [online]. 27 July 2012 [cit. 2018-06-21]. Dostupné z: <https://www.baldurbjarnason.com/notes/epub-javascript-security/>
- [107] BOTT, Ed. How Apple is sabotaging an open standard for digital books. In: *ZDNet* [online]. CBS Interactive, January 22, 2012 [cit. 2018-04-01]. Dostupné z: <http://www.zdnet.com/article/how-apple-is-sabotaging-an-open-standard-for-digital-books/>
- [108] CALDWELL, Serenity. Hands on: iBooks Author effortless to use, but iPad-only. In: *Macworld* [online]. IDG Communications, Jan 19, 2012 [cit. 2018-05-20]. Dostupné z: <https://www.macworld.com/article/1164895/software-graphics/hands-on-ibooks-author-effortless-to-use-but-ipad-only.html>
- [109] GLAZMAN, Daniel. IBooks Author, a nice tool but.. In: *<Glazblog/>* [online]. 20 January 2012 [cit. 2018-04-01]. Dostupné z: <http://www.glazman.org/weblog/dotclear/index.php?post/2012/01/20/iBooks-Author-a-nice-tool-but>
- [110] BJARNASON, Baldur. The iBooks 2.0 textbook format. In: *Baldur Bjarnason* [online]. 19 January 2012 [cit. 2018-04-01]. Dostupné z: <http://www.baldurbjarnason.com/notes/the-ibooks-textbook-format/>
- [111] APPLE. Add video or audio to your books in iBooks Author. *Apple Support* [online]. Apple, January 17, 2017 [cit. 2018-06-21]. Dostupné z: <https://support.apple.com/en-us/HT202374>
- [112] APPLE. Formats Supported by QuickTime. *Final Cut Pro Help* [online]. Apple, ©2010 [cit. 2018-06-21]. Dostupné z: <https://documentation.apple.com/en/finalcutpro/usermanual/index.html#chapter=103%26section=4%26tasks=true>
- [113] APPLE. IBooks Author: Widgets overview. *Apple Support* [online]. Apple, Dec 7, 2016 [cit. 2018-06-20]. Dostupné z: [https://support.apple.com/kb/ph2789?locale=en\\_US](https://support.apple.com/kb/ph2789?locale=en_US)
- [114] BJARNASON, Baldur. The iBooks 2.0 built-in widgets. In: *Baldur Bjarnason* [online]. 20 January 2012 [cit. 2018-04-22]. Dostupné z: <http://www.baldurbjarnason.com/notes/the-ibooks-built-in-widgets/>
- [115] APPLE. About ePubs created with iBooks Author. *Apple Support* [online]. Apple, June 21, 2017 [cit. 2018-06-20]. Dostupné z: <https://support.apple.com/en-us/HT204884>
- [116] CALDWELL, Serenity. IBooks Author 2.3 adds support for ePub 3, opens up ebook creation to anyone. In: *IMore* [online]. Mobile Nations, 14 Jul 2015 [cit. 2018-04-22]. Dostupné z: <https://www.imore.com/ibooks-author-23-opens-ebook-creation-anyone-support-epub-3>
- [117] URBÁNEK, Vladimír. Wooky jedná se strategickým investorem, posouvá vstup na burzu. In: *Kurzy.cz* [online]. Kurzy.cz, 10.07.2012 [cit. 2018-06-21]. Dostupné z: <https://www.kurzy.cz/zpravy/330846-wooky-jedna-se-strategickym-investorem-posouva-vstup-na-burzu/>
- [118] POKORNÝ, Lukáš. Popularita e-knižních formátů. *Inflow: informaion journal* [online]. 5. 10. 2012 [cit. 2018-04-28]. ISSN 1802–9736. Dostupné z: <http://www.inflow.cz/popularita-ekniznich-formatu>
- [119] VYTLAČIL, Petr. Wooky se již kamarádí se čtečkami e-knih PocketBook. In: *Ebooky.cz* [online]. 7. 11. 2014 [cit. 2018-04-28]. Dostupné z: <http://www.ebooky.cz/wooky-se-jiz-kamaradi-se-cteckami-e-knih-pocketbook/>

- [120] THE NATIONAL ARCHIVES. The technical registry PRONOM. *The National Archives* [online]. [2002] [cit. 2018-07-19]. Dostupné z: <http://www.nationalarchives.gov.uk/PRONOM/>
- [121] ADOBE SYSTEMS. PDF: tři písmena, která změnila svět. *Adobe* [online]. Adobe Systems, ©2018 [cit. 2018-06-17]. Dostupné z: <https://acrobat.adobe.com/cz/cs/acrobat/about-adobe-pdf.html>
- [122] VYCHODIL, Bedřich. *Produkce digitálních obrazových dat a jejich kontrola* [online]. Praha, 2013, 324 s. [cit. 2018-06-17]. Dostupné z: <https://is.cuni.cz/webapps/zzp/detail/102364/>. Disertační práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce Doc. RNDr. Jiří Souček, DrSc.
- [123] PDFLIB. *A Technical Introduction to PDF/VT: PDFlib Whitepaper* [online]. München: PDFlib, [2011] [cit. 2018-06-17]. Dostupné z: [https://www.pdflib.org/wp-content/untill2016\\_uploads/2011/08/Technical-Introduction-to-PDF-VT.pdf](https://www.pdflib.org/wp-content/untill2016_uploads/2011/08/Technical-Introduction-to-PDF-VT.pdf)
- [124] NOVÁK, Jiří. *Formát PDF a jeho využití* [online]. Praha, 2014 [cit. 2018-06-17]. Dostupné z: [https://is.ambis.cz/th/g3n9/format\\_pdf\\_a\\_jeho\\_vyuziti.pdf](https://is.ambis.cz/th/g3n9/format_pdf_a_jeho_vyuziti.pdf). Bakalářská práce. Bankovní institut vysoká škola, Katedra informatiky a kvantitativních metod. Vedoucí práce Ing. Bohuslav Růžička, CSc.
- [125] PDF. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- , last edited on 25 May 2018 [cit. 2018-06-18]. Dostupné z: <https://en.wikipedia.org/wiki/PDF>
- [126] ADOBE SYSTEMS INCORPORATED. *PDF Reference: Adobe portable document format version 1.4* [online]. Third edition. Boston: ADDISON–WESLEY, b.r. [cit. 2018-07-19]. ISBN 0-201-75839-3. Dostupné z: [https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdfs/pdf\\_reference\\_archives/PDFReference.pdf](https://www.adobe.com/content/dam/acom/en/devnet/pdf/pdfs/pdf_reference_archives/PDFReference.pdf)
- [127] ARMS, Caroline, Don CHALFANT, Kevin DEVORSEY, Chris DIETRICH, Carl FLEISCHAUER, Butch LAZORCHAK, Sheila MORRISSEY a Kate MURRAY. *The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions: an NDSA report* [online]. National Digital Stewardship Alliance, February 2014 [cit. 2018-06-17]. Dostupné z: <http://hdl.loc.gov/loc/gdc/lcpub.2013655115.1>
- [128] FANNING, Betsy A. *Preservation with PDF/A (2nd Edition): DPC Technology Watch Report 17-01* [online]. 2. Great Britain: Digital Preservation Coalition, 2017. DOI: <http://dx.doi.org/10.7207/twr17-01>. ISSN 2048 - 7916. Dostupné také z: <https://www.dpconline.org/docs/technology-watch-reports/1707-twr17-01-revised/file>
- [129] PALMER, Jörg. *Improved PDF/A-1b* [online]. PDF/A Competence Center, 2007-09-22 [cit. 2018-06-17]. Dostupné z: [https://www.pdflib.org/wp-content/untill2016\\_uploads/2011/08/improved\\_pdfa-2b.pdf](https://www.pdflib.org/wp-content/untill2016_uploads/2011/08/improved_pdfa-2b.pdf)
- [130] WHEATLEY, Paul, Peter MAY, Maureen PENNOCK a Simon WHIBLEY. PDF Format Preservation Assessment. In: *Digital Preservation Coalition Wiki* [online]. Version 1.3. British Library, 2015 [cit. 2018-06-17]. Dostupné z: [http://wiki.dpconline.org/images/e/e8/PDF\\_Assessment\\_v1.3.pdf](http://wiki.dpconline.org/images/e/e8/PDF_Assessment_v1.3.pdf)
- [131] LIBRARY OF CONGRESS. PDF (Portable Document Format) Family. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, Last Updated 28-Feb-2017 [cit. 2018-06-17]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000030.shtml>
- [132] LIBRARY OF CONGRESS. PDF/A, PDF for Long-term Preservation. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, Last Updated 07/27/2017 [cit. 2018-06-17]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000318.shtml>

- [133] KNIJFF, Johan van der. *Adobe Portable Document Format: Inventory of long-term preservation risks* [online]. Version 0.2. Koninklijke Bibliotheek, 20-10-2009 [cit. 2018-06-17]. Dostupné z: [http://openpreservation.org/system/files/PDFInventoryPreservationRisks\\_0\\_2\\_0.pdf](http://openpreservation.org/system/files/PDFInventoryPreservationRisks_0_2_0.pdf)
- [134] JOHNSON, Duff. The 8 most popular document formats on the web. In: *Duff Johnson* [online]. 17 FEB [2014] [cit. 2018-06-19]. Dostupné z: <http://duff-johnson.com/2014/02/17/the-8-most-popular-document-formats-on-the-web/>
- [135] ISO/TC 171/SC 02/WG 08. *ISO/TC 171/SC 2/WG 8 N 603: Meeting report* [online]. Salt Lake City (UT, USA), May 2011 [cit. 2018-06-19]. Dostupné z: [https://web.archive.org/web/20130810233608/http://pdf.editme.com/files/pdfREF-meetings/ISO-TC171-SC2-WG8\\_N0603\\_SC2WG8\\_MtgRept\\_SLC.pdf](https://web.archive.org/web/20130810233608/http://pdf.editme.com/files/pdfREF-meetings/ISO-TC171-SC2-WG8_N0603_SC2WG8_MtgRept_SLC.pdf)
- [136] JOHNSON, Duff. Achieving Canonical PDF Validation. In: *Proceedings of the 11th International Conference on Digital Preservation, iPRES 2014, Melbourne, Australia, October 6 - 10, 2014* [online]. Melbourne: State Library of Victoria, 2014, s. 39-43 [cit. 2018-06-17]. ISBN 978-0-642-27881-4. Dostupné z: <http://www.ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf>
- [137] JOHNSON, Duff. Is PDF an Open Standard?. In: *Talking PDF* [online]. Lansdowne (Pennsylvania, U.S.A.): Appligent, May 14, 2010 [cit. 2018-06-17]. Dostupné z: <https://talkingpdf.org/is-pdf-an-open-standard/>
- [138] MORRISSEY, Sheila M. The Network is the Format: PDF and the Long-term Use of Digital Content. In: *IS&T Archiving Conference 2012, Copenhagen, Denmark, 12-15 June 2012* [online]. Copenhagen, Denmark: Society for Imaging Science & Technology, 2012, s. 200-203 [cit. 2018-06-17]. ISBN 978-1-63266-641-3. Dostupné z: <https://web.archive.org/web/20161224065550/http://www.portico.org/digital-preservation/wp-content/uploads/2012/11/TheNetworkIsTheFormat.pdf>
- [139] TUNNAT, Yvonne. JHOVE – the one and only PDF validator. In: *Open Preservation Foundation* [online]. Open Preservation Foundation, 22nd December 2017 [cit. 2018-06-17]. Dostupné z: <http://openpreservation.org/blog/2017/12/19/jhove-the-one-and-only-pdf-validator/>
- [140] LINDLAR, Michelle, Yvonne TUNNAT a Carl WILSON. A PDF Test-Set for Well-Formedness Validation in JHOVE - The Good, the Bad and the Ugly. In: *Proceedings of the 14th International Conference on Digital Preservation, iPRES 2017, Kyoto, Japan, September 25 – 29, 2017* [online]. Kyoto: iPRES, 2017 [cit. 2018-06-17]. Dostupné z: <https://ipres2017.jp/wp-content/uploads/35.pdf>
- [141] PREFORMA TEAM. *PREFORMA* [online]. [2014-2018] [cit. 2018-07-22]. Dostupné z: <http://www.preforma-project.eu/>
- [142] OPEN PRESERVATION FOUNDATION. VeraPDF (PDF/A validation). *Open Preservation Foundation* [online]. [2015] [cit. 2018-06-17]. Dostupné z: <http://openpreservation.org/about/projects/verapdf/>
- [143] WOLF, Julia. *OMG WTF PDF: [PDF Ambiguity and Obfuscation]* [online]. 2010 March 31 [cit. 2018-06-17]. Dostupné z: [https://www.troopers.de/media/filer\\_public/13/51/135142fe-0768-4b53-92a3-fd2fc7248ed2/tr11\\_wolf\\_omg\\_pdf.pdf](https://www.troopers.de/media/filer_public/13/51/135142fe-0768-4b53-92a3-fd2fc7248ed2/tr11_wolf_omg_pdf.pdf)
- [144] KLINDT, Marco. PDF/A considered harmful for digital preservation. In: *Proceedings of the 14th International Conference on Digital Preservation, iPRES 2017, Kyoto, Japan, September 25 – 29, 2017* [online]. Kyoto: iPRES, 2017 [cit. 2018-06-17]. Dostupné z: <https://ipres2017.jp/wp-content/uploads/15Marco-Klindt.pdf>
- [145] MITCHAM, Jenny a Paul WHEATLEY. PDF to PDF-A conversion. In: *OPF Knowledge Base Wiki* [online]. Open Preservation Foundation, Sep 27, 2011, last edited on May 09, 2012 [cit. 2018-06-17]. Dostupné z: <http://wiki.opf-labs.org/display/REQ/PDF+to+PDF-A+conversion>



- [146] BROWN, Geoffrey a Kam WOODS. Born Broken: Fonts and Information Loss in Legacy Digital Documents. *The International Journal of Digital Curation* [online]. 2011, 6(1), s. 5-19 [cit. 2018-06-18]. Dostupné z: <http://www.ijdc.net/article/view/159/243>
- [147] ZARNDT, Frederick, Dorothy CARNER a Edward MCCAIN. Born Digital Legal Deposit Policies and Practices. In: *IFLA 2017 News Media Satellite Conference, Dresden, Germany, 16.-18. August 2017* [online]. Dresden: SLUB, 2017, s. 1-33 [cit. 2018-03-30].
- [148] NĚMECKO. Verordnung über die Pflichtablieferung von Medienwerken an die Deutsche Nationalbibliothek (Pflichtablieferungsverordnung - PflAV) vom 17. Oktober 2008. In: *Bundesgesetzblatt*. 2008 [Zuletzt geändert durch Art. 2 G v. 1.9.2017 I 3346], Teil I., Nr. 47. ISSN 0341-1095. Dostupné také z: <https://www.gesetze-im-internet.de/pflav/BJNR201300008.html>
- [149] NĚMECKO. Gesetz über die Deutsche Nationalbibliothek (DNBG) vom 22. Juni 2006. In: *Bundesgesetzblatt*. 2006 [Zuletzt geändert durch Art. 2 G v. 1.9.2017 I 3346], Teil I., Nr. 29. ISSN 0341-1095. Dostupné také z: <https://www.gesetze-im-internet.de/pflav/BJNR201300008.html>
- [150] BÍLÝ, Michal. *Trendy v legislativách povinného odevzdávání online publikací v evropských zemích*. Praha, 2016. Dostupné také z: <https://is.cuni.cz/webapps/zzp/detail/166123/>. Bakalářská práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí práce PhDr. Eva Bratková, Ph.D.
- [151] Müssen Kindle-E-Books an die Deutsche Nationalbibliothek abgeliefert werden? [Aktualisierung]. *Literaturcafe.de* [online]. 28. Juli 2011 [Aktualisierung vom 15. Januar 2013] [cit. 2018-03-31]. Dostupné z: [www.literaturcafe.de/muessen-kindle-e-books-an-die-deutsche-nationalbibliothek-abgeliefert-werden/](http://www.literaturcafe.de/muessen-kindle-e-books-an-die-deutsche-nationalbibliothek-abgeliefert-werden/)
- [152] DEUTSCHE NATIONALBIBLIOTHEK. Frequently asked questions (FAQ). *Deutsche Nationalbibliothek: Online publications* [online]. Leipzig: Deutsche Nationalbibliothek, Last update 14.02.2018, 14.02.2018 [cit. 2018-03-31]. Dostupné z: <http://www.dnb.de/EN/Header/Hilfe/netzpublikationenFaq.html>
- [153] HEIN, Stefan a Karlheinz SCHMITT. Risk Management for Digital Long-Term Preservation Services. In: *Proceedings of the 10th International Conference on Digital Preservation of Digital Objects, iPRES 2013, Lisbon, Portugal, 3-5 September 2013* [online]. Lisbon: Biblioteca Nacional de Portugal, 2013 [cit. 2018-04-19]. ISBN 978-972-565-493-4. Dostupné z: <http://purl.pt/24107/1/>
- [154] DEUTSCHE NATIONALBIBLIOTHEK. *Spezifikation der Dateiformat-Policy für die Sammlung von Netzpublikationen der Deutschen Nationalbibliothek: Erläuterungen zur Handhabung* [online]. Version 1.0. Leipzig: Deutsche Nationalbibliothek, 24.10.2012 [cit. 2018-04-19]. Dostupné z: <https://d-nb.info/1027296343/34>
- [155] DIEBEL, Cornelia. How to handle the masses – automated workflows as a solution for the collection and preservation of e-books in the German National Library. In: *IFLA WLIC 2014, 16-22 August 2014, Lyon, France* [online]. Lyon: IFLA, 2014 [cit. 2018-07-08]. Dostupné z: <http://library.ifla.org/843/1/087-diebel-en.pdf>
- [156] DERROT, Sophie a Clément OURY. Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France. In: *IFLA WLIC 2014, 16-22 August 2014, Lyon, France* [online]. Lyon: IFLA, 2014 [cit. 2018-04-18]. Dostupné z: <http://library.ifla.org/830/1/087-derrot-en.pdf>
- [157] DERROT, Sophie, Jean-Philippe MOREUX, Clément OURY a Stéphane REECHT. Preservation of ebooks: from digitized to born-digital. In: *Proceedings of the 11th International Conference on Digital Preservation, iPRES 2014, Melbourne, Australia, October 6 - 10, 2014* [online]. Melbourne: State Library of Victoria, 2014 [cit. 2018-03-30]. ISBN 978-0-642-27881-4. Dostupné z: <https://hal-bnf.archives-ouvertes.fr/hal-01088755>
- [158] CAMILE, Géraldine. *BnF experiences with harvesting content beyond paywalls: NAS Workshop, Vienna, 27 April 2017* [online presentace]. Vienna, 27 April 2017 [cit. 2018-04-18]. Dostupné z:

- <https://sbforge.org/download/attachments/23101500/BnF-2017%20Content%20paywall.pptx?version=1&modificationDate=1493300835131&api=v2>
- [159] BIBLIOTHÈQUE NATIONALE DE FRANCE. List of data formats accepted in SPAR (end of 2015). *BnF* [online]. Paris: Bibliothèque nationale de France, 2015 [cit. 2018-04-18]. Dostupné z: [http://www.bnf.fr/documents/spar\\_formats\\_techniques\\_en.pdf](http://www.bnf.fr/documents/spar_formats_techniques_en.pdf)
- [160] BIBLIOTHÈQUE NATIONALE DE FRANCE. Data formats for digital preservation purposes. *BnF* [online]. Paris: Bibliothèque nationale de France, April 7, 2016 [cit. 2018-04-18]. Dostupné z: [http://www.bnf.fr/en/professionals/digital\\_preservation\\_data\\_metadata\\_bnf/a.data\\_formats\\_bnf.html](http://www.bnf.fr/en/professionals/digital_preservation_data_metadata_bnf/a.data_formats_bnf.html)
- [161] LEDOUX, Thomas. SPAR: From Design to Operations. In: *Presentations from the PASIG meetings in Austin, Texas, January 2012* [online]. Austin (USA): Stanford University, 2012 [cit. 2018-04-18]. Dostupné z: <http://c.bnf.fr/fyL>
- [162] BIBLIOTHÈQUE NATIONALE DE FRANCE. Technical metadata for digital preservation. *BnF* [online]. Paris: Bibliothèque nationale de France, 2016 [cit. 2018-07-08]. Dostupné z: [http://www.bnf.fr/en/professionals/digital\\_preservation\\_data\\_metadata\\_bnf/a.technical\\_metadata\\_bnf.html#SHDC\\_\\_Attribute\\_BlocArticle1BnF](http://www.bnf.fr/en/professionals/digital_preservation_data_metadata_bnf/a.technical_metadata_bnf.html#SHDC__Attribute_BlocArticle1BnF)
- [163] COUFAL, Libor. *Digitální sbírky a jejich dlouhodobé uchování v Národní knihovně Austrálie* [online prezentace]. První část. Praha, 22.3.2017 [cit. 2018-04-20]. Dostupné z: <https://www.slideshare.net/dp-blog-cz/libor-coufal-australsk-nrodn-knihovna-pednka-praha-2232017>
- [164] NATIONAL LIBRARY OF AUSTRALIA. Legal Deposit frequently asked questions. *National library of Australia* [online]. Canberra: National library of Australia, b.r. [cit. 2018-04-20]. Dostupné z: <https://www.nla.gov.au/legal-deposit-faq>
- [165] NATIONAL LIBRARY OF AUSTRALIA. Edeposit Home Page. *National Library of Australia* [online]. Canberra: National Library of Australia, b.r. [cit. 2018-07-22]. Dostupné z: <https://www.nla.gov.au/edeposit>
- [166] COUFAL, Libor. *Digitální sbírky a jejich dlouhodobé uchování v Národní knihovně Austrálie* [online prezentace]. 2. část. Praha, 22.3.2017 [cit. 2018-04-20]. Dostupné z: <https://www.slideshare.net/dp-blog-cz/libor-coufal-australsk-nrodn-knihovna-2-pednka-praha-2232017>
- [167] SLOVINSKO. Zakon [č. 2977] o obveznem izvodu publikacij (ZOIPub). In: *Uradni list Republike Slovenije*. 2006, Št. 69. ISSN 1318-0576. Dostupné také z: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO3606>
- [168] SLOVINSKO. Pravilnik [č. 4422] o vrstah in izboru elektronskih publikacij za obvezni izvod. In: *Uradni list Republike Slovenije*. 2007, Št. 90. ISSN 12033-12034. Dostupné také z: <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina?urlid=200790&stevilka=4422>
- [169] NARODNA IN UNIVERZITETNA KNJIŽNICA. *SVAROG: sistem za varno arhiviranje oddanega gradiva* [online]. [Ljubljana]: Narodna in univerzitetna knjižnica, b.r. [cit. 2018-07-22]. Dostupné z: <http://svarog.nuk.uni-lj.si/>
- [170] NARODNA IN UNIVERZITETNA KNJIŽNICA. Obvezni izvod spletnih publikacij. *Narodna in univerzitetna knjižnica* [online]. Ljubljana: Narodna in univerzitetna knjižnica, b.r. [cit. 2018-04-21]. Dostupné z: <http://www.nuk.uni-lj.si/informacije/obvezni-izvod-elektronski-spletni-nosilci-zapisa>
- [171] KONINKLIJKE BIBLIOTHEEK, DUTCH PUBLISHERS' ASSOCIATION. *Arrangement for depositing electronic publications at the Deposit of Netherlands Publications in the Koninklijke Bibliotheek* [online]. The Hague: Koninklijke Bibliotheek, 1999 [cit. 2018-05-20]. Dostupné z: <https://www.kb.nl/sites/default/files/docs/overeenkomst-nuv-kb-en.pdf>

- [172] INTERNATIONAL PUBLISHERS ASSOCIATION. Digital Legal Deposit: An IPA Special Report. *IPA* [online]. Geneva: International Publishers Association, 24th July 2014 [cit. 2018-05-20]. Dostupné z: <https://www.internationalpublishers.org/images/news/2014/digital-legal-deposit-2014.pdf>
- [173] KONINKLIJKE BIBLIOTHEEK. Deposit Library for Dutch Publications. *KB* [online]. The Hague: Koninklijke Bibliotheek, b.r. [cit. 2018-05-20]. Dostupné z: <https://www.kb.nl/en/organisation/for-publishers/deposit-library-for-dutch-publications>
- [174] KONINKLIJKE BIBLIOTHEEK. Depositing individual digital publications. *KB* [online]. The Hague: Koninklijke Bibliotheek, b.r. [cit. 2018-05-20]. Dostupné z: <https://www.kb.nl/en/organisation/for-publishers/depositing-publications/depositing-individual-digital-publications>
- [175] KNIJFF, Johan van der. Identification of PDF preservation risks: analysis of Govdocs selected corpus. In: *Open Preservation Foundation* [online]. Open Preservation Foundation, 27th Jan 2014 [cit. 2018-07-07]. Dostupné z: <http://openpreservation.org/blog/2014/01/27/identification-pdf-preservation-risks-analysis-govdocs-selected-corpus/>
- [176] ROG, Judith a Caroline van WIJK. *Evaluating File Formats for Long-term Preservation* [online]. The Hague: National Library of the Netherlands, 2008 [cit. 2018-03-31]. Dostupné z: [https://www.kb.nl/sites/default/files/docs/KB\\_file\\_format\\_evaluation\\_method\\_27022008.pdf](https://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf)
- [177] NATIONAL LIBRARY OF NEW ZEALAND. Country report: New Zealand. *CDNL-AO meeting 2007 [Conference of Directors of National Libraries in Asia and Oceania: The 15th meeting]* [online]. Tokyo: National Diet Library, 2007 [cit. 2018-05-18]. Dostupné z: [http://www.ndl.go.jp/en/cdnla0/meetings/pdf/CR2007\\_New\\_Zealand.pdf](http://www.ndl.go.jp/en/cdnla0/meetings/pdf/CR2007_New_Zealand.pdf)
- [178] NATIONAL LIBRARY OF NEW ZEALAND. Submitting your publication. *National Library* [online]. Wellington: National Library of New Zealand, b.r. [cit. 2018-05-18]. Dostupné z: <https://natlib.govt.nz/publishers-and-authors/legal-deposit/submitting-your-publication>
- [179] NATIONAL LIBRARY OF NEW ZEALAND. Preserving digital objects. *National Library* [online]. Wellington: National Library of New Zealand, b.r. [cit. 2018-05-18]. Dostupné z: <https://natlib.govt.nz/publishers-and-authors/legal-deposit/preserving-digital-objects>
- [180] NATIONAL LIBRARY OF NEW ZEALAND. National Library of New Zealand Collections Policy. *National Library* [online]. Wellington: National Library of New Zealand, 26 October 2015 [cit. 2018-05-18]. Dostupné z: <https://natlib.govt.nz/about-us/strategy-and-policy/collections-policy/national-library-of-new-zealand-collections-policy>
- [181] KING, Peter a Heather MATHIE. List of ISBNs of electronic books submitted by electronic legal deposit the library. In: *FYI.org.nz* [online]. Open New Zealand, March 12, 2018, poslední aktualizace April 18, 2018 [cit. 2018-05-18]. Dostupné z: <https://fyi.org.nz/request/7434-list-of-isbns-of-electronic-books-submitted-by-electronic-legal-deposit-the-library>
- [182] CAPIAU, Sam. *Deponering van digitale publicaties uit Vlaanderen: Een haalbaarheidsonderzoek door de Vlaamse Erfgoedbibliotheek in opdracht van de Vlaamse overheid* [online]. Antwerpen: Vlaamse Erfgoedbibliotheek, 2015, 104 s. [cit. 2018-05-20]. Dostupné z: <http://www.vlaamse-erfgoedbibliotheek.be/en/node/3510>
- [183] SPOJENÉ KRÁLOVSTVÍ. *The Legal Deposit Libraries (Non-Print Works) Regulations* [online]. London: TSO, 2013 [cit. 2018-07-13]. ISBN 9780111538203. Dostupné také z: [http://www.legislation.gov.uk/uksi/2013/777/pdfs/uksi\\_20130777\\_en.pdf](http://www.legislation.gov.uk/uksi/2013/777/pdfs/uksi_20130777_en.pdf)
- [184] SPOJENÉ KRÁLOVSTVÍ. *Legal Deposit Libraries Act 2003* [online]. London: TSO, 2003 [cit. 2018-07-14]. Dostupné také z: [http://www.legislation.gov.uk/ukpga/2003/28/pdfs/ukpga\\_20030028\\_en.pdf](http://www.legislation.gov.uk/ukpga/2003/28/pdfs/ukpga_20030028_en.pdf)
- [185] Depositing electronic publications. *British Library* [online]. London: British Library, b.r. [cit. 2018-07-14]. Dostupné z: <https://www.bl.uk/aboutus/legaldeposit/websites/elecpubs/>

- [186] BRITISH LIBRARY. *Enter your username and password [Publisher Submission Portal]* [online]. British Library, b.r. [cit. 2018-07-22]. Dostupné z: <https://publisherdeposit.idls.org.uk/>
- [187] MAY, Peter. Testing JHOVE PDF Module: the good, the bad, and the not well-formed. In: *Open Preservation Foundation* [online]. Open Preservation Foundation, 10th Mar 2017 [cit. 2018-07-14]. Dostupné z: <http://openpreservation.org/blog/2017/03/10/testing-jhove-pdf-module-the-good-the-bad-and-the-not-well-formed/>
- [188] SMITH, Caylin. There's Ample to Sample: Content Sampling at the British Library. In: *Open Preservation Foundation* [online]. Open Preservation Foundation, 23rd Jan 2017 [cit. 2018-07-14]. Dostupné z: <http://openpreservation.org/blog/2017/01/23/theres-ample-to-sample-content-sampling-at-the-british-library/>
- [189] DIGITAL PRESERVATION COALITION. *Non-Print Legal Deposit Digital Preservation Review: Final Report* [online]. Digital Preservation Coalition, 15/12/2017 [cit. 2018-07-14]. Dostupné z: <https://www.bl.uk/britishlibrary/~media/bl/global/digital%20preservation/non-print-legal-deposit-digital-preservation-review.pdf>
- [190] WEBB, Colin, David PEARSON a Paul KOERBIN. 'Oh, you wanted us to preserve that?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections. *D-Lib Magazine* [online]. 2013, **19**(12) [cit. 2018-06-19]. Dostupné z: <http://www.dlib.org/dlib/january13/webb/01webb.html>
- [191] JIROUŠEK, Václav a Natalie OSTRÁKOVÁ. Výběr vhodných formátů pro dlouhodobé uchování elektronických publikací. *Bulletin SKIP* [online]. 2017, **26**(3) [cit. 2018-07-01]. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/node/222>
- [192] MATUŠÍK, Zdeněk. Právní podmínky sběru povinného depozitu elektronických publikací, jejich zpracování a zpřístupnění v knihovnách. *Knihovna* [online]. 2014, **25**(1), 55-83 [cit. 2018-07-08]. ISSN 1801-3252. Dostupné z: <http://oldknihovna.nkp.cz/knihovna141/141055.htm>
- [193] BERNAS, Jiří. Národní digitální archiv. *Knihovna* [online]. 2009, **20**(1), 22-29 [cit. 2018-07-08]. ISSN 1802-8772. Dostupné z: <http://oldknihovna.nkp.cz/knihovna91/bernas.htm>
- [194] LIBRARY OF CONGRESS. Sustainability Factors. In: *Sustainability of Digital Formats: Planning for Library of Congress Collections* [online]. Washington (DC): Library of Congress, Last Updated 01/ 5/2017 [cit. 2018-06-26]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>
- [195] TODD, Malcolm. *File formats for preservation: Technology Watch Report* [online]. DPC Technology Watch Series Report 09-02. Digital Preservation Coalition, 2009 [cit. 2018-07-01]. Dostupné z: <https://www.dpconline.org/docs/technology-watch-reports/375-file-formats-for-preservation/file>
- [196] Bezpečnost elektronických publikací. *E-deposit* [online]. Praha: Národní knihovna České republiky, 2018 [cit. 2018-06-30]. Dostupné z: <http://edeposit.nkp.cz/cs/bezpecnost-elektronickych-publikaci>
- [197] WORLD WIDE WEB CONSORTIUM. Comparison of e-book formats: W3C working draft. *W3C* [online]. b.r. [cit. 2018-06-21]. Dostupné z: <https://www.w3.org/AudioVideo/ebook/>
- [198] MCLELLAN, Evelyn Peters. *Selecting Digital File Formats for Long-Term Preservation: InterPARES 2 Project General Study 11 Final Report* [online]. Version 1.1. The InterPARES 2 Project, March 2007 [cit. 2018-06-26]. Dostupné z: [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_file\\_formats\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf)
- [199] BROWN, Adrian. *Selecting File Formats for Long-Term Preservation: Digital Preservation Guidance Note 1* [online]. National Archives, 2003 [cit. 2018-06-26]. Dostupné z: [http://webarchive.nationalarchives.gov.uk/+http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://webarchive.nationalarchives.gov.uk/+http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)

- [200] DIGITAL PRESERVATION COALITION. File formats and standards. *Digital Preservation Handbook* [online]. 2nd Edition. Glasgow: Digital Preservation Coalition, 2015 [cit. 2018-06-26]. Dostupné z: <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>
- [201] Flash & The Future of Interactive Content. *Adobe Blog* [online]. San Jose: Adobe Systems, 2017 [cit. 2018-04-28]. Dostupné z: <https://theblog.adobe.com/adobe-flash-update/>
- [202] NÁRODNÍ KNIHOVNA ČESKÉ REPUBLIKY. Standardy pro obrazová data. *NDK: Národní digitální knihovna* [online]. Praha: Národní knihovna České republiky, 20.03.2017 [cit. 2018-07-18]. Dostupné z: <https://www.ndk.cz/standardy-digitalizace/standardy-pro-obrazova-data>
- [203] LIBRARY OF CONGRESS. Text: Quality and Functionality Factors. *Sustainability of Digital Formats* [online]. Washington (DC): Library of Congress, Last Updated 05-Jan-2017 [cit. 2018-07-18]. Dostupné z: [https://www.loc.gov/preservation/digital/formats/content/text\\_quality.shtml](https://www.loc.gov/preservation/digital/formats/content/text_quality.shtml)
- [204] KNIJFF, Johan van der. Policy-based assessment with VeraPDF – a first impression. *Open Preservation Foundation* [online]. Open Preservation Foundation, 1st Jun 2017 [cit. 2018-07-18]. Dostupné z: <http://openpreservation.org/blog/2017/06/01/policy-based-assessment-with-verapdf-a-first-impression/>
- [205] ROG, Judith. *PDF Guidelines: Recommendations for the creation of PDF files for long-term preservation and access* [online]. Version 1.7. The Hague: Koninklijke Bibliotheek, 31-05-2007 [cit. 2018-07-07]. Dostupné z: [https://www.kb.nl/sites/default/files/docs/pdf\\_guidelines.pdf](https://www.kb.nl/sites/default/files/docs/pdf_guidelines.pdf)
- [206] CLIFF, Peter a Johan van der KNIJFF. Policy-Driven Identification of Preservation Risks in Electronic Document Formats. In: *SCAPE Public Wiki* [online]. SCAPE Project, Jan 28, 2013, last edited Dec 09, 2013 [cit. 2018-07-19]. Dostupné z: <http://wiki.opf-labs.org/display/SP/Policy-Driven+Identification+of+Preservation+Risks+in+Electronic+Document+Formats>
- [207] KNIJFF, Johan van der. Portable Document Format. In: *OPF Knowledge Base Wiki* [online]. Open Preservation Foundation, last edited on Jul 09, 2014 [cit. 2018-07-19]. Dostupné z: <http://wiki.opf-labs.org/display/TR/Portable+Document+Format>
- [208] WILSON, Carl. Openpreserve / jhove: PDF hul Messages. *GitHub* [online]. [San Francisco]: GitHub, 9 May 2018, edited 17 May 2018 [cit. 2018-07-18]. Dostupné z: <https://github.com/openpreserve/jhove/wiki/PDF-hul-Messages>
- [209] WILSON, Carl. Openpreserve / jhove: PDF hul Messages 2. *GitHub* [online]. [San Francisco]: GitHub, 17 May 2018 [cit. 2018-07-18]. Dostupné z: <https://github.com/openpreserve/jhove/wiki/PDF-hul-Messages-2>
- [210] GITHUB. Epubcheck / src / test / resources / com / adobe / epubcheck / test / command\_line / listSeverities\_expected\_results.txt. *GitHub* [online]. [San Francisco]: GitHub, ©2018 [cit. 2018-07-19]. Dostupné z: [https://github.com/IDPF/epubcheck/blob/master/src/test/resources/com/adobe/epubcheck/test/command\\_line/listSeverities\\_expected\\_results.txt](https://github.com/IDPF/epubcheck/blob/master/src/test/resources/com/adobe/epubcheck/test/command_line/listSeverities_expected_results.txt)

## Seznam počítačových programů

### **Použité počítačové programy**

GOYAL, Kovid. *Calibre: ebook management* [software]. Verze 3.21 z 19. dubna 2018. Kovid Goyal, ©2006. Dostupné také z: <https://calibre-ebook.com/>

INTERNATIONAL DIGITAL PUBLISHING FORUM. *EpubCheck* [software]. Version 4.0.2. IDPF, 29 Nov 2016. Dostupné také z: <https://github.com/IDPF/epubcheck/>

MOBIPOCKET. *Mobipocket Reader* [software]. Version 6.2. Paris: Mobipocket.com, ©1999-2007.

MICROSOFT. *Microsoft® Word 2016* [software]. Redmond: Microsoft, 2015.

NATIONAL ARCHIVES. *DROID: file format identification tool* [software]. Version 6.4 z 5.3.2018. Signature File V93. The National Archives, ©2016. Dostupné také z: <https://github.com/digital-preservation/droid>

OPEN PRESERVATION FOUNDATION. *JHOVE* [software]. Version 1.20.1 z 2018-03-29. PDF-hul v 1.11. JPEG-hul v1.4. Open Preservation Foundation, [2004-2018]. Dostupné také z: <http://jhove.openpreservation.org/>

VERAPDF CONSORTIUM. *VeraPDF: PDF/A Conformance Checker* [software]. Version 1.12.1. VeraPDF Consortium, ©2015-2018. Dostupné také z: <http://verapdf.org>

### **Ostatní počítačové programy zmiňované v textu**

ADOBE SYSTEMS. *Adobe Acrobat Professional* [software].

ADOBE SYSTEMS. *Adobe Acrobat Reader DC* [software].

ADOBE SYSTEMS. *Adobe Distiller* [software].

APACHE SOFTWARE FOUNDATION. *Apache PDFBox®* [software].

APPLE. *iBooks* [software].

ARTIFEX SOFTWARE. *Ghostscript®* [software].

FARQUHAR, Adam et al. *Format Identification for Digital Objects (FIDO)* [software].

HARVARD LIBRARY. *The File Information Tool Set (FITS)* [software].

HENDRICKS, Kevin a Doug MASSAY. *FlightCrew* [software].

HENDRICKS, Kevin et al. *KindleUnpack* [software].

MARKOVIĆ, Strahinja et al. *Sigi: Editor EPUB* [software].

SCAPE PROJECT TEAM. *FLINT* [software].

## Seznam tabulek

Tabulka 1. Souborové formáty e-knih dle báze ohlášených e-knih (2010 až září 2012) .....	19
Tabulka 2. Souborové formáty e-knih dle báze ohlášených e-knih (2010 až červen 2018).....	19
Tabulka 3. Souborové formáty e-knih dle báze ohlášených e-knih (rok 2017).....	19
Tabulka 4. Výskyt formátů MOBI v kombinaci s dalšími formáty .....	20
Tabulka 5. Přehled vybraných údajů o formátu OEBPS .....	34
Tabulka 6. Přehled vybraných údajů o formátu MOBI .....	36
Tabulka 7. Přehled vybraných údajů o formátu EPUB 2 .....	39
Tabulka 8. Přehled vybraných údajů o formátu EPUB 3 .....	43
Tabulka 9. Přehled vybraných údajů o formátu iBooks .....	46
Tabulka 10. Přehled vybraných údajů o formátu Wooky .....	48
Tabulka 11. Přehled vybraných údajů o formátu PDF .....	50
Tabulka 12. Digital e-Legal Deposit Survey – akceptované formáty e-publikací.....	58
Tabulka 13. Digital e-Legal Deposit Survey – další akceptované formáty e-publikací .....	58
Tabulka 14. Hodnocení rizik e-publikací dle DNB .....	59
Tabulka 15. Seznam nástrojů využívaných DNB pro práci s přijímanými e-publikacemi .....	60
Tabulka 16. Struktura kvality EPUB a PDF odevzdaných do DNB mezi 12/2012 a 4/2013...60	
Tabulka 17. Struktura kvality EPUB a PDF odevzdaných do DNB v roce 2013.....	61
Tabulka 18. Kategorizace formátů podle míry rizika pro uchovávání formátu dle BnF .....	62
Tabulka 19. Seznam nástrojů využívaných BnF pro práci s přijímanými e-publikacemi.....	62
Tabulka 20. Seznam nástrojů využívaných NLA pro práci s přijímanými e-publikacemi.....	64
Tabulka 21. Seznam nástrojů využívaných KB pro práci s přijímanými e-publikacemi .....	67
Tabulka 22. Struktura formátů e-publikací odevzdaných do NLNZ .....	68
Tabulka 23. Ověření validity 2759 souborů PDF z fondu BL.....	69
Tabulka 24. Seznam nástrojů využívaných BL pro práci s přijímanými e-publikacemi.....	70
Tabulka 25. Přehled formátů, přijímaných sledovanými knihovnami.....	70
Tabulka 26. Výsledek identifikace souborových formátů nástrojem DROID.....	73
Tabulka 27. Výsledek validace souborů EPUB validátorem EPUBCheck .....	73
Tabulka 28. Chyby zjištěné validátorem EPUBCheck u souborů EPUB.....	74
Tabulka 29. Výsledek validace souborů PDF modulem PDF-hul (JHOVE) .....	74
Tabulka 30. Výsledek kontroly souborů MOBI v aplikacích Calibre a Mobipocket Reader...75	
Tabulka 31. Výskyt DRM a nevalidních souborů u jednotlivých zastoupených formátů.....	76

## **Seznam obrázků**

Obrázek 1. Počet ohlášených ISBN u 50 nejaktivnějších producentů e-knih v roce 2017.....20

Obrázek 2. Grafické vyjádření vztahů mezi formáty e-knih, centrováno na EPUB.....42



## **Seznam příloh**

Příloha č. 1: analýza dat e-publikací dobrovolně odevzdaných do NK ČR v letech 2011–2018  
(příloha je součástí elektronické verze bakalářské práce)