

OPTICKÉ DATOVÉ DISKY VE FONDU DVOU NEJVĚTŠÍCH ČESKÝCH KNIHOVEN: HISTORICKÝ PŘEHLED A MOŽNOSTI JEJICH OCHRANY

OPTICAL DATA DISCS IN CZECH LIBRARIES COLLECTIONS: HISTORICAL OVERVIEW AND CURRENT STATE

Zdeněk Hruška

Moravská zemská knihovna

Abstrakt

Účel – Tento odborný článek se zabývá optickými datovými disky ve fondu dvou největších tuzemských knihoven, změnou role těchto médií v posledních desetiletích a jakým rizikům čelí. Představuje projekt CDArcha, který nabízí nástroje na ochranu obsahu datových optických disků a nastiňuje kroky, které bude do budoucna nutné přijmout, aby data uložená na nosičích zůstala dlouhodobě přístupná pro uživatele knihoven.

Design/metodologie/přístup – V první části, která se zabývá situací v Národní knihovně ČR v 90. letech a na počátku nového století, článek vychází z analýzy projektových dokumentů a dalších materiálů, jenž k tématu vznikly. Druhá část se pak věnuje situaci v Moravské zemské knihovně a popisu projektu CDArcha.

Výsledky – Článek popisuje, jaké strategie dvě největší české knihovny v minulosti užívaly v ochraně optických disků ve svých fondech. Nástroje popsané ve druhé části umožňují paměťovým institucím extrahovat data z ohrožených disků a bezpečně je uložit na centrálním serveru a zajistit jim tak ochranu na bitové úrovni.

Originalita/hodnota – Přínosem článku je shrnutí výzkumných aktivit Národní knihovny ČR v oblasti optických disků – téma bylo odbornou knihovnickou komunitou opomíjeno, samotné projektové zprávy jsou dnes dostupné již jen ve Webarchivu NK ČR. Projekt CDArcha, který zapadá do výzkumných a vývojových aktivit Moravské zemské knihovny, nabízí pro knihovny a další paměťové instituce nástroje pro ochranu dat uložených na optických discích.

Klíčová slova: optické disky, ISO image, CD-ROM, DVD, knihovny, ochrana na bitové úrovni

Abstract

Purpose – The article focuses on optical data discs in the collections of the two largest Czech libraries and examines how their role has changed in recent decades and what risks they face. It introduces the CDArcha project, which offers tools to protect the content of data optical discs and outlines the steps that need to be taken in the future to ensure that data stored on them remains accessible to library users in the long term.

Design/Methodology/Approach – The first part based on the analysis of project documents and other materials explores the situation in the National Library of Czech Republic in the 1990s and at the turn of the 21st century. The second part discusses the situation in the Moravian Library itself and describes the CDArcha project.

Results – An outline of measures taken in the past to protect optical discs in the collections of the two largest Czech libraries is presented. In the second section new software tools are introduced; they allow memory institutions to extract data from optical discs and safely store them on a central server to provide protection at the bit level are introduced.

Originality/Value – The article provides a summary of the research of the National Library of Czech Republic in the field of optical discs; the topic was long neglected by the professional library community. The CDArcha project, which forms an integral part of the research and development in the Moravian Library, offers tools for libraries and other memory institutions to protect data stored on optical discs.

Keywords: optical discs, ISO image, CD-ROM, DVD, libraries, bit-level preservation

Úvod

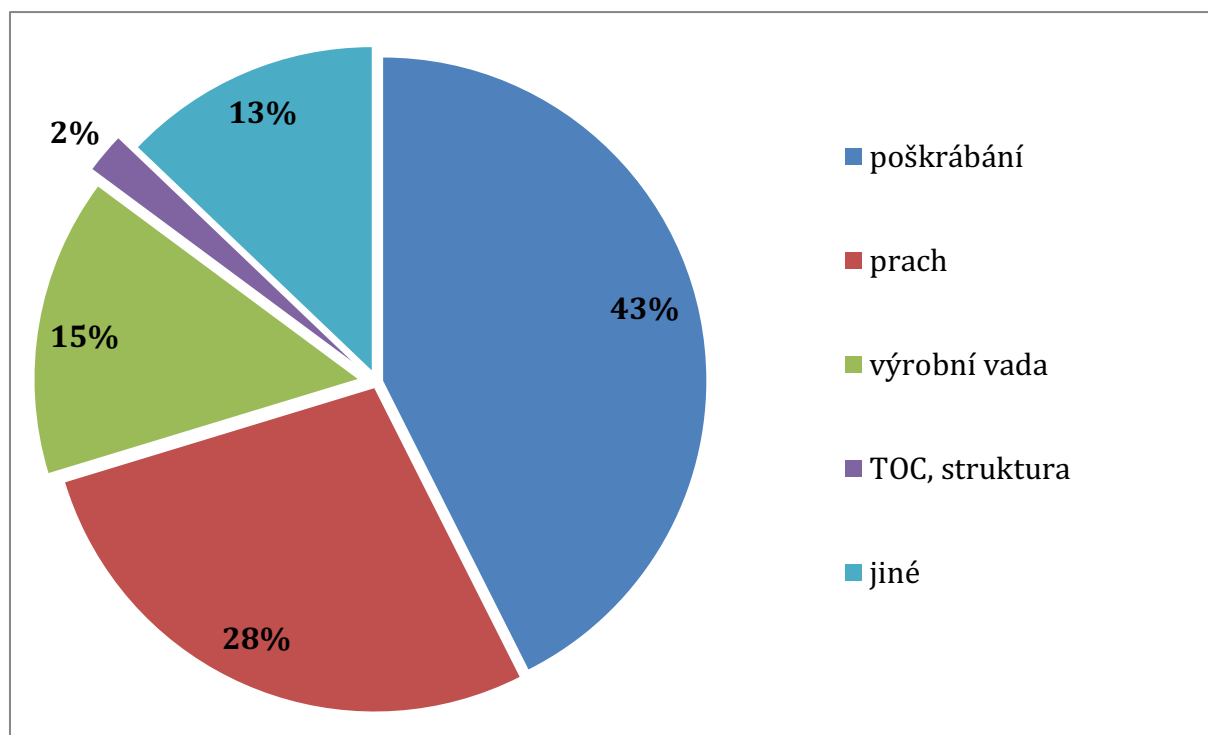
Zatímco postupy pro ochranu tradičních (papírových) dokumentů ve fondu knihoven jsou zažité a dobře ověřené, u těch nepapírových je situace složitější. Ochranou zvukových nosičů se zabývají např. projekty Virtuální národní fonotéka a Nový fonograf. Oba zmiňované projekty se zaměřují na starší nosiče (jako jsou fonografické válečky, šelakové a vinylové desky) z důvodu jejich většího ohrožení fyzickou degradací. V Moravské zemské knihovně (dále jako MZK) aktuálně probíhá příprava projektu na digitalizaci audio CD (ačkoliv se nejedná o digitalizaci v pravém smyslu slova, ale o tzv. grabování). Primárním cílem je zjednodušit přístup uživatelů ke zvukovým dokumentům pomocí importu dat do digitální knihovny Kramerius MZK, i když datové balíčky jsou koncipovány tak, aby je bylo možné v budoucnu co nejnlehčeji importovat do dlouhodobého digitálního úložiště (též LTP – z anglického Long Term Preservation). Vzhledem k tomu, že MZK zatím podobným úložištěm nedisponuje a LTP úložiště projektu Národní digitální knihovny (NDK) není v tuto chvíli schopné tyto balíčky ingestovat, budou data prozatím uložena na serverech v MZK.

Dalšími nosiči jsou datové optické disky – ty jsou zastoupeny především typy CD-ROM a DVD-ROM, ale v knihovních fondech se nalézá celá škála optických disků, na kterých jsou uložena nejrůznější data – ať už textové a obrazové dokumenty, multimédia nebo hry a jiný software. CD a DVD disky ušly za několik posledních desítek let zajímavou cestu. Od médií, která ještě relativně nedávno sloužila k ukládání a zálohování dat, až k nosičům, jejichž obsah je potřeba chránit, jinak ho můžeme navždy ztratit. Jsou tak odrazem rychlého vývoje jak v oblasti LTP, tak i v knihovnách jako takových – technologie se mění a data uložená na starších nosičích se stávají ohrožená, tudíž je potřeba podniknout akce na jejich ochranu.

Vývoj v Národní knihovně ČR

Historii optických disků v českém knihovním prostředí poměrně dobře mapují aktivity Národní knihovny ČR (NK ČR) v minulých desetiletích, protože část digitalizovaných dat byla v této instituci ukládána v 90. letech a na počátku nového století na optická média (data z digitalizace Memoriae Mundi Series Bohemica (Knoll, 1998), později i zálohy databází (Knoll, Polišenský a Uhlíř, 2004). Z tohoto důvodu se NK ČR poměrně detailně věnovala testování optických disků ve svém fondu. Jednak kvůli vybírání vhodných nosičů pro uložení dat, jednak pro ověřování podmínek uložení a později následoval také průzkum disků ve fondu získaných v rámci povinného výtisku, nákupů a dalšími způsoby. Celkem byla data archivována na zhruba 10 000 CD-R a tyto archivní kopie byly uloženy v ideálních podmínkách bez přístupu světla, v teplotách do 20 °C a max. 30% vzdušné vlhkosti (Psohlavec 2005).

Během testování v letech 2000 a 2001 nebyly na médiích využívaných pro ukládání dat Memoriae Mundi Series Bohemica zjištěny žádné problémy. To koresponduje s pečlivým výběrem a podmínkami uložení – alespoň u archivních kopií, u uživatelských disků byly podmínky o něco volnější (Knoll a Psohlavec, 2002). Zajímavé jsou rovněž výsledky testování disků z běžného fondu NK ČR a procentuální zastoupení zjištěných vad. Přestože není uveden počet testovaných jednotek, je to dobrá ukázka, jakým rizikům CD disky čelily. Zatímco prach a poškrábání jsou problémy, které se dají řešit vhodným uložením a zacházením (např. vytvářením uživatelských kopií), 15% zastoupení vad z výroby je něco, co knihovna ovlivní jen těžko.



Graf 1 Složení nalezených vad (Knoll a Psohlavec, 2002)

Testování probíhalo i v dalších letech a také docházelo k vytváření a ukládání ISO obrazů CD a DVD disků ve fondu (bitová kopie disku, dle normy ISO 9660), jako způsob ochrany jejich obsahu. Takto bylo vytvořeno zhruba 800 ISO obrazů, které jsou dosud stále uloženy na serverech v NK ČR (Ljubka, diskuze na Semináři k digitalizaci (nejen) pro krajské knihovny VIII., MZK, Brno 2019). V další fázi bylo plánováno ve spolupráci s MZK a Vědeckou knihovnou v Olomouci zpracovat dostupné optické disky z povinného výtisku ve fondech zmíněných knihoven, přičemž pro evidenci zpracovaných jednotek se uvažovalo o využití Registru digitalizace (Hutař, Melichar 2010). Identifikátorem zpracovaných jednotek mělo být číslo České národní bibliografie (čČNB) spolu s URN:NBN (Knoll, Polišenský a Uhlíř, 2009). K tomuto společnému zpracování disků však nikdy nedošlo a vývoj v této oblasti se na několik let úplně zastavil. Jednotlivé knihovny nadále řešily problém ohrožení optických disků ve svých fondech vlastními silami – vytvářením uživatelských kopií, nebo zálohováním dat na externí disky, ale žádné centrální řešení se v následujících letech neobjevilo.

Vývoj v Moravské zemské knihovně

Ve fondu MZK se nachází cca 20 až 25 000 nosičů. Přesné číslo není možné určit kvůli několika faktorům – část záznamů je převzata z jiných knihoven a záznam nemusí vždy odpovídat realitě (např. některé knihy mohou být zakoupeny do MZK bez nosiče, ale v záznamu je tento nosič uveden). Dalším aspektem je nedostatečný popis, kdy z katalogizačního záznamu není možné určit, o jaký disk se jedná – může jít o CD i DVD, CD může být audio či datové (např. CD obsahující pouze soubory ve formátu mp3 je z jednoho pohledu datové CD, zatímco z pohledu katalogizátora mohlo jít o disk zvukový). Metadatový popis se v MZK zlepšil až s příchodem katalogizačních pravidel RDA, od té doby jsme schopni některé údaje vyčíst z příslušných polí katalogizačního záznamu pro elektronické dokumenty:

- pole 300\$a – typ nosiče, např.:
 - 1 DVD-ROM
 - 2 elektronické optické disky (CD-ROM)
 - 1 CD-R :
- pole 347\$a a \$b – charakteristika digitálního souboru; např.:
 - 347 \$a textový soubor; \$b PDF
- pole 538\$a – požadavky na systém; např.:
 - Windows 95 a vyšší; Adobe Acrobat Reader (na disku), Word, PowerPoint, přehrávač videa (wmv); rozlišení 1024x768; mechanika DVD-ROM
 - Osobní počítač s procesorem Pentium III, 64 MB RAM (doporučeno 128), grafický adaptér VGA, mechanika CD-ROM nebo DVD-ROM, operační systém Windows, operační systém Unix s X Window, prohlížeč Internet Explorer 5.0 nebo vyšší, applet JAVA 1.4 nebo vyšší

Vzhledem k tomu, že starší záznamy nemají výše uvedená pole vždy vyplněná, anebo je popis nekonkrétní, budeme mít kompletní přehled o discích až poté, co všechny fyzicky prozkoumáme a zjistíme, jaký typ obsahu je na nich uložen.

Disky ve fondu MZK dělíme pro naše potřeby do několika základních skupin podle jejich obsahu a typu katalogizačního záznamu:

1. Zvuková CD (CD-DA, formát je definován tzv. Červenou knihou (angl. Red Book) a mezinárodní normou IEC 60908). Tyto nosiče jsou mimo záběr tohoto článku, jejich zpracováním se zabývá jiný projekt v MZK.
2. Datové disky (převážně CD-ROM a DVD-ROM, ale i další typy nosičů jako jsou CD-R, CD-RW, DVD+-R, DVD-RW a další).
 - A. Disky, které jsou přílohou jiného dokumentu (nejčastěji knih a časopisů) a v metadatovém záznamu jsou zmíněny pouze v polích patřících příloze. Obsah je velmi pestrý, jedná se jak o zvukové disky, tak i disky obsahující multimédia, software a další.
 - B. Disky se samostatným katalogizačním záznamem. Jsou uloženy v MZK pod signaturou CDR a CDR-2, do fondu se dostávají v rámci povinného výtisku, ale i jinými cestami. Často obsahují šedou literaturu (sborníky z konferencí, výstupy z projektů, skripta z vysokých škol), ale též multimédia a nejrůznější software.

Ještě v roce 2010 nebylo zpracování optických disků v MZK důležitým tématem: “Například Moravská zemská knihovna (dle sdělení vedoucí odboru revize a ochrany fondů) neprovádí ani jednoduché zálohování CD či DVD ze svého fondu.” (Lukšů, 2010). Vzhledem k tomu, že vývoj v této oblasti probíhal již několik let v NK ČR, byl to ze strany MZK pochopitelný přístup. Po konci prací v NK ČR se začal pomalu měnit přístup i v brněnské knihovně. MZK šla cestou zjednodušení přístupu uživatelů k obsahu disků a záloha dat byla až vedlejším výsledkem této činnosti. Datové disky ze signatur CDR a CDR-2 byly manuálně zpracovávány na PC, jejich obsah zkopírován na externí disk a pokud to formát dat umožňoval, tak byla data importována do digitální knihovny. V případě, že byl celý obsah importován do Krameria MZK, disk se již uživatelům nepůjčuje. V opačném případě zůstává stále k dispozici.

Vzhledem k tomu, že digitální knihovna Kramerius umí zobrazovat vedle obrazových a zvukových souborů i formát PDF, byly vybrané soubory do tohoto formátu konvertovány. To jde velmi dobře u obrazových formátů jako je JPG, TIFF, PNG, textových formátů DOC, DOCX, RTF, ODT, formátů prezentací PPT a ODP (pokud neobsahují zvukové nebo video stopy či jiné interaktivní prvky). V omezených případech i u formátů tabulkových (pokud převodem nedochází ke ztrátě interaktivity nebo vzorců), jako je XLS, XLSX a ODS. Digitální knihovna Kramerius neuměla ve starších verzích strukturovat jednotlivé PDF soubory do kapitol či jednotlivých článků, takže bylo nutné ze všech vhodných souborů na disku vytvořit jeden jediný PDF soubor. Ve výsledku to někdy vedlo ke vzniku PDF o několika tisících stranách a velikosti přes 500 MB. To s sebou často přinášelo problémy při

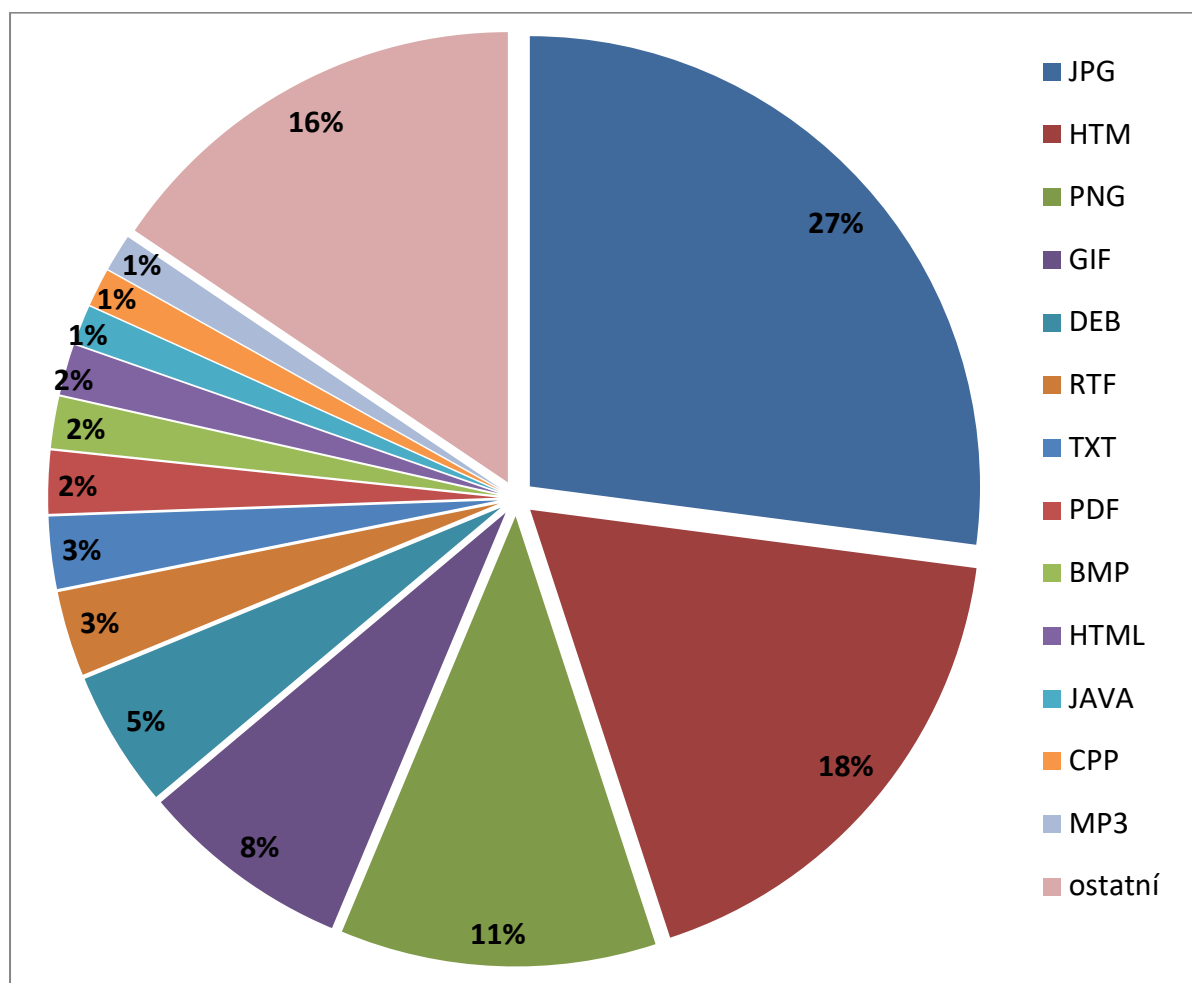
importu do Krameria, indexaci i při samotném zobrazování. Soubor se načítal dlouho, jeho struktura nemusela odpovídat struktuře dat na disku, uživatel nemusel mít k dispozici kompletní funkcionalitu disku, chyběl tvůrci disku zamýšlený uživatelský zážitek. Tyto nevýhody však byly vyváženy zjednodušením přístupu k informacím – uživatel již nemusel čekat na vyhledání disku ve skladu, jeho doručení a nemusel poté hledat PC s optickou mechanikou.

Zpracování disků nepatřilo k hlavním činnostem knihovny, ročně se podařilo zpracovat pouze několik desítek až stovek disků. Aktuálně v roce 2019 je to cca 2 000 kusů, tedy ani ne 10 % z celkového předpokládaného počtu disků ve fondu MZK.

Postupně byly podniknuty kroky k pokročilejší ochraně obsahu optických disků. Data zálohovaná na externí disk byla opatřena kontrolními součty MD5 pro zajištění alespoň základní bitové ochrany. Pro další ochranné kroky bylo ale nejdříve nutné získat více informací o obsahu optických disků ve fondu knihovny. Bylo proto rozhodnuto o formátové analýze náhodně vybraného vzorku nosičů. Průzkum byl proveden v rámci studentské stáže ve spolupráci s Kabinetem informačních studií a knihovnictví FF MU. Za použití nástroje DROID, který slouží k identifikaci formátů souborů, byl analyzován obsah 200 náhodně vybraných disků a díky tomuto a obdobnému průzkumu disků z roku 2013 ze signatury CDR jsme si vytvořili představu o tom, s jakými daty budeme v budoucnu pracovat.

| | Formát | Počet | | Formát | Počet | | Formát | Počet |
|-----|--------|-------|-----|--------|-------|-----|--------|-------|
| 1. | JPG | 54083 | 11. | JAVA | 2793 | 21. | ZIP | 1164 |
| 2. | HTM | 35755 | 12. | CPP | 2731 | 22. | ICO | 865 |
| 3. | PNG | 22649 | 13. | MP3 | 2669 | 23. | JS | 798 |
| 4. | GIF | 15150 | 14. | PAS | 1926 | 24. | WAV | 792 |
| 5. | DEB | 9788 | 15. | DLL | 1498 | 25. | DSP | 789 |
| 6. | RTF | 6104 | 16. | EXE | 1323 | 26. | DPR | 781 |
| 7. | TXT | 5163 | 17. | SFL | 1308 | 27. | RC | 771 |
| 8. | PDF | 4489 | 18. | DCU | 1236 | 28. | PY | 697 |
| 9. | BMP | 3699 | 19. | RES | 1211 | 29. | CLASS | 689 |
| 10. | HTML | 3645 | 20. | DFM | 1200 | 30. | XML | 606 |

Tabulka 1 Třicet nejčastějších formátů na CD a DVD v MZK



Graf 2 Nejčastější formáty na CD a DVD v MZK

Projekt CDArcha

Vzhledem k pomalému manuálnímu zpracování fondu optických disků se další vývoj zaměřil na zjednodušení a automatizaci celého procesu. Byl tak položen základ projektu CDArcha. První prototyp spatřil světlo světa na konci roku 2017 a jednalo se o modifikovanou aplikaci projektu Obálky knih, která slouží ke skenování obálek knih v zapojených knihovnách a odesílání dat na server. Software CDArcha byl dále v roce 2018 rozvíjen díky financím z programu VISK3 (pod názvem Vytvoření nástrojů pro kooperativní zpracování a bit-level ochranu obsahu datových optických disků ve fondu knihoven). Za projektem stojí idea společného zpracování a ochrany na bitové úrovni datového obsahu optických disků ve fondu českých knihoven. Výsledkem byly dvě aplikace CDArcha-klient a CDArcha-server, obě jsou volně dostupné na platformě GitHub. Aplikaci CDArcha-klient lze jednoduše instalovat na běžné kancelářské PC, opatřené optickou mechanikou a vybavené skenerem. Každá zapojená instituce je pak zaregistrována a obdrží jednoznačný identifikátor, kterým jsou následně opatřeny všechny datové balíčky pocházející od tohoto producenta. Aby se zabránilo případnému duplicitnímu zpracování disků v různých institucích je využívána dvojitá kontrola – jednak pomocí popisných metadat, jednak pomocí kontrolního součtu ISO obrazu disku. Jedním z cílů je vytvořit databázi všech zpracovaných disků, jako vhodnou

platformou se jeví Registr digitalizace, který by bylo nutné nejdříve upravit, aby vyhovoval všem potřebám pro evidenci optických disků.

Zpracování probíhá tak, že nejprve je pomocí čárového kódu z katalogu ALEPH stažen metadatový záznam nosiče a aplikace CDArcha-klient vytvoří z optického disku ISO image, poté je naskenovaná vrchní strana disku (i vizuální podobu nosiče je nutné zachovat) a pokud existuje, digitalizuje se i obal a booklet. Na závěr jsou data spolu s kontrolními součty odeslána na server k dalšímu zpracování. Zde pak v procesu pokračuje aplikace CDArcha-server, která doplní výstup z nástroje DROID, vytvoří OCR k naskenovaným obrazům disku, bookletu a obalu a vše uloží do balíčku ve formátu BagIt (standard vytvořený Library of Congress pro transfer a uložení dat, též jako RFC 8493). Vnitřní datová a metadatová struktura je inspirovaná standardy projektu Národní digitální knihovny. Datový a metadatový balíček je ve finále uložen na serveru a díky kontrolním součtům je možné provádět pravidelné automatické kontroly, zda nedošlo k poškození dat.

Balíček BagIt ve verzi 0.97 (identifikátorem je řetězec UUID)

- soubor *manifest-md5.txt*
 - obsahuje kontrolní součty ve formátu md5 pro všechny soubory ve složce data
- soubor *bagit.txt*
 - obsahuje verzi BagItu a kódování znakové sady (vždy UTF-8)
- složka “data”
 - složka “iso_img”
 - obsahuje ISO image
 - složka “master_copy_scan”
 - skeny vrchní strany disku, bookletu a přebalu ve formátu JPEG2000 v archivní verzi
 - složka “user_copy_scan”
 - skeny vrchní strany disku, bookletu a přebalu ve formátu JPEG2000 v uživatelské kopii
 - složka “alto”
 - alto vytvořené při OCR skenů
 - složka “txt”
 - txt vytvořené při OCR skenů
 - složka “amdsec”
 - administrativní metadata, vedlejší METS proč každou naskenovanou stranu
 - soubor *info_uuid.xml*
 - popis obsahu balíčku (manifest)
 - soubor *main_mets_uuid.xml*
 - hlavní mets záznam

- soubor *checksum_uuid.md5*
 - kontrolní součet balíčku

Schéma 1 Model datového a metadatového balíčku

V roce 2019 probíhá další rozvoj systému, a to směrem k lepší optimalizaci celého procesu (např. vyřešení problému s nadměrnou zátěží serveru ze strany nástroje DROID), úprava grafického rozhraní pro snazší a intuitivnější používání, možnosti importovat obsah flash disků (které se také nachází ve fondu knihoven a které jsou mnohem zranitelnější než optické disky, jelikož jejich obsah lze snadno smazat). Probíhá také právní rozbor, za jakých podmínek lze ukládat bitové kopie optických disků z různých knihoven na jednom serveru a jak těmto právním požadavkům vyhovět po technické stránce.

MZK se rovněž připojila k projektové žádosti do Programu na podporu aplikovaného výzkumu a vývoje národní a kulturní identity na léta 2016 až 2022 (NAKI II) na rozvoj open source LTP systému ARCLib a to za účelem, aby tento systém byl schopný importovat datové balíčky vzniklé v projektu CDArcha. Nicméně ani zde snahy o zajištění dlouhodobé ochrany nekončí – dříve či později bude nutné vyřešit logickou ochranu obsahu optických disků. Jako nejvhodnější řešení se jeví softwarová emulace (Espenschied et al, 2013). Ve světě běží několik projektů, mající ambice vytvořit rozsáhlé emulační prostředí, které bude možné využít pro dlouhodobou logickou ochranu a zpřístupnění interaktivního obsahu optických disků. Jako perspektivní se jeví např. projekt EaaS (The Scaling Emulation and Software Preservation Infrastructure), který nabízí infrastrukturu a nástroje pro emulaci softwaru (Rechet, 2019).

Závěr

Článek shrnuje historický vývoj přístupu k optickým diskům v českých knihovnách na příkladu Národní knihovny ČR a Moravské zemské knihovny v Brně, jakožto dvou největších knihoven v ČR. V druhé části představuje projekt CDArcha, jeho záměr, technické řešení i plánovaný rozvoj.

Optické datové disky tvoří početnou skupinu nosičů dat v knihovních fondech. Jejich obsah je často unikátní, nemá, a mnohdy ani nemůže mít, svůj analogový protějšek. Bohužel je ohrožuje řada rizik, proto je nutné podniknout akce na ochranu jejich obsahu. Moravská zemská knihovna v Brně, jako příjemce povinného výtisku, má ve svém fondu velké množství těchto unikátních dokumentů a jejich nosičů. Má také dostatek zkušeností s vědeckým výzkumem i vývojem softwaru. V rámci projektu CDArcha si proto dala za úkol vytvořit nástroje pro kooperativní zpracování a dlouhodobou ochranu obsahu optických disků ve fondu českých knihoven. Do budoucna čeká na knihovny ještě mnoho práce, první kroky k zajištění ochrany této vzácné části fondu již ale byly učiněny.

Dedikace

Článek vznikl v rámci Institucionální podpory na dlouhodobý koncepční rozvoj výzkumné organizace – Moravská zemská knihovna v Brně.

Použité zdroje

CDArcha-client [Online]. (2019). Dostupné z: <https://github.com/moravianlibrary/CDArcha-client>

CDArcha-server [Online]. (2019). Dostupné z: <https://github.com/moravianlibrary/CDArcha-server>

Espenschied, D., Rechert, K., von Suchodoletz, D., Valizada, I. & Russler, N. (2013) Large-Scale Curation and Presentation of CD-ROM Art. *iPres 2013*. Dostupné z: http://purl.pt/24107/1/iPres2013_PDF/Large-Scale%20Curation%20and%20Presentation%20of%20CD-ROM%20Art.pdf

Hutař, J., Melichar, M. (2010). *Návrh opatření pro prodloužení životnosti obsahu CD a DVD disků ve sbírkách NK*. Praha: Národní knihovna ČR. Dostupné z: https://wayback.webarchiv.cz/wayback/20170130203300/http://digit.nkp.cz/projekty/VZ-2004_2010/2010/PDF/10_cd-archive_END.pdf

Knoll, A. (1998) Memoriae Mundi Series Bohemica: Program digitálního zpřístupnění vzácných fondů. *Ikaros*. 2(7). urn:nbn:cz:ik-11113. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/11113>

Knoll, A. & Psohlavec, S. (2002) *Zpráva o řešení projektu výzkumu a vývoje Optimalizace archivace a zpřístupnění digitálních dat: závěrečná zpráva za léta 2001-2002*. Praha: Národní knihovna ČR. Dostupné z: <https://wayback.webarchiv.cz/wayback/20140316071406/http://digit.nkp.cz/knihcin/digit/vav23/zpravaweb.doc>

Knoll, A., Polišenský, J. & Uhlíř, Z. (2004). *Vytvoření virtuálního badatelského prostředí pro zpřístupnění a ochranu digitálních dokumentů: Zpráva o řešení výzkumného záměru za rok 2004*. Praha: Národní knihovna ČR. Dostupné z: https://wayback.webarchiv.cz/wayback/20100428151510/http://digit.nkp.cz/projekty/VZ-2004_2010/2004/Zprava2004.pdf

Knoll, A., Polišenský, J. & Uhlíř, Z. (2009). *Vytvoření virtuálního badatelského prostředí pro zpřístupnění a ochranu digitálních dokumentů: dílčí zpráva o řešení za r. 2009*. Praha: Národní knihovna ČR. Dostupné z: https://wayback.webarchiv.cz/wayback/20100428200212/http://digit.nkp.cz/projekty/VZ-2004_2010/2009/Hlavn%C3%ADZpravaWeb.pdf

Kunze, J., Boyko, A., Littman, J., Madden, L., & Vargas, B. (2012). The BagIt File Packaging Format (V0.97) [Online]. In *IETF Tools*. Dostupné z: <https://tools.ietf.org/html/draft-kunze-bagit-08>

Ljubka, Ivan. Národní knihovna České republiky. *Diskuze na Semináři k digitalizaci (nejen) pro krajské knihovny VIII.*, pořádaném Moravskou zemskou knihovnou v Brně, 3. - 4. dubna 2019.

Lukšů, A. (2010). *Dlouhodobé uchování a zpřístupňování dokumentů zaznamenaných na optických discích*. Brno: Masarykova univerzita. Filozofická fakulta. Ústav české literatury a knihovnictví. Bakalářská diplomová práce. Dostupné z: https://is.muni.cz/th/lxafa/Bakalarska_prace_Alzbeta_Luksu.pdf

Psohlavec, S. Praktické zkušenosti s dlouhodobou archivací dat na CD-R discích. (2005). *Automatizace knihovnických procesů 2005 (AKP 2005), 10. ročník semináře, Liberec, 3. a 4. květen 2005*. Dostupné z: <https://wayback.webarchiv.cz/wayback/20140321010326/http://www.akvs.cz/akp-2005/11-psohlavec.pdf>

Rechet, K. (2019) Scaling Emulation and Software Preservation Infrastructure (EaaS). *ILIDE International Conference 2019*. Dostupné z: http://www.schk.sk/wordpress/wp-content/themes/MichelleSCHK/digilib/Presentations/April19/ILIDE2019_Rechet.pptx

Poznámka o autorovi

Jméno: Zdeněk Hruška

Pracoviště: Moravská zemská knihovna, Oddělení skenování

E-mail: zdenek.hruska@mzk.cz