
Dlouhodobé uchovávání digitálních dat vzniklých digitalizací zvukových záznamů na fonografických válečkách a šelakových deskách

Long term preservation of data produced by digitization of sound recordings on phonograph cylinders and shellac discs

Mgr. Miroslava Beňačková, Mgr. Pavlína Kočišová, Mgr. Vojtěch Kopský, Ph.D., Mgr. Natalie Ostráková / Oddělení pro standardy, Národní knihovna České republiky (Digital Preservation Standards Department, National Library of the Czech Republic), Klementinum 190, 110 00 Praha

Resumé:

Fonoválečky a šelakové gramofonové desky jsou nejstarší zvuková záznamová média. Byly vyráběny z nestabilních materiálů, a proto jsou značně ohroženy fyzickou degradací spojenou se ztrátou nesené informace. Článek popisuje vznik a historii těchto nosičů ve vztahu k ochraně jejich obsahu a přístup Národní digitální knihovny (NDK) k této ochraně s důrazem na standardy digitalizace. Je zde diskutován obsah a struktura balíčku SIP a jeho popis. Důraz je kladen především na metadata, u nichž bylo třeba respektovat specifika nosičů, například v případě intelektuální entity, jež je někdy nesená více než jedním nosičem. S ohledem na nedostupnost některých údajů byl též přehodnocen metadatový popis zvukových dokumentů, a to ve smyslu redukce povinných elementů a úplného vyřazení těch, které nelze vyplnit. Práce rovněž analyzuje formáty používané pro uložení (WAV resp. BWF) a zpřístupnění (MP3) zvukových souborů a zaměřil se nad potenciálem formátu FLAC.

Klíčová slova: fonografické válečky, šelakové gramofonové desky, digitalizace, zvukové dokumenty, metadatové standardy, dlouhodobé uchovávání, archivační souborové formáty

Summary:

The earliest audio media, phonograph cylinders and shellac discs, were made of materials prone to physical degradation and loss of information. This article describes the origins and development of those technologies with relevance to their prospective preservation and the approach of the National Digital Library (NDL) to the preservation concentrating on the standards. Content and structure of the submission information package are discussed. The emphasis is on the metadata that have to reflect specifics of the media, such as the partition of the intellectual entity to several carriers. The number of mandatory metadata was reduced due to limited accessibility of certain data and elements that can never be found and were excluded altogether. The article also discusses formats used for the preservation and distribution of preserved audio files, recognizing WAV/BWF and MP3 as respective standards, while the potential of FLAC is also considered.

Keywords: phonographic cylinders, shellac discs, digitization, audio files, metadata standards, digital preservation, archival file formats

Úvod

Zvukové dokumenty jsou významnou součástí kulturního dědictví a v posledních letech také předmětem digitalizačních aktivit, jež směřují jednak k uchování obsahu nosičů pro další generace a jednak k online zpřístupnění pro současné uživatele. Podobně jako tištěné dokumenty, jejichž digitalizace aktuálně masově probíhá, jsou zvukové dokumenty ohroženy degradací fyzických nosičů a současně kvůli zastarávání a nedostupnosti funkčních přehrávacích zařízení. Na záchranu obsahu z fonografických válečků a šelakových desek míří projekt NAKI *Nový fonograf*¹. Jeho součástí je vytvoření národního standardu pro digitalizaci těchto nosičů a příprava na jejich dlouhodobé uložení v důvěryhodném úložišti. V uplynulém období byla v Národní knihovně ČR v rámci projektu *Nový fonograf* věnována pozornost gramofonovým deskám. Výsledkem této práce vznikl standard pro digitalizaci gramofonových desek, jejichž digitalizace je nyní podporována i v podprogramu VISK 7 (Česko 2019). Vedle gramofonových šelakových desek patří mezi nejohroženější nosiče zvukových dokumentů fonografické válečky, magnetofonové pásky a některé kompaktní disky. Cílem tohoto textu je popsat problematiku související s digitalizací a dlouhodobým uchováváním digitalizovaných zvukových dokumentů, přičemž pozornost je věnována především metadatovému popisu digitálních dokumentů a vhodným archivačním formátům, jež jsou základním stavebním kamenem datových balíčků pro dlouhodobou archivaci. Text předkládá specifika metodiky digitální ochrany dokumentů podle zásad digitalizace Národní knihovny ČR.

1 Historické nosiče zvukových dokumentů

1.1 Fonografické válečky

Zvukový váleček se jako médium pro záznam zvuku objevuje v 80. letech 19. století a představuje nejstarší komerčně vyráběný zvukový nosič. Pro záznam zvuku byl používán zejména do třicátých let 20. století, okrajově se ovšem objevuje až do 60. let, než byl definitivně vytlačen modernějšími nosiči. Zvukové válečky jsou mírně kónického tvaru, s dutým středem, někdy vyztuženým kartonovou špulkou. Jejich velikost se liší v závislosti na mnoha faktorech; jako nejobvyklejší rozměry se udávají 2 a ¼ " v průměru a 4-4 a ¼" na délku. Zvuk je pak zaznamenán na vnější straně do mělkých drážek, vytlačených do povrchu válečku jehlou. Zaznamenání melodické linky nahrávky bylo docíleno prohlubněmi o různé hloubce ve stopě drážky. Podle velikosti válečku, hustoty drážek a přehrávací rychlosti, válečky obvykle obsahují záznam v délce kolem dvou minut.

Obecně se objevují dva druhy zvukových válečků, starší voskové a mladší plastové. Materiálem voskových válečků je ve skutečnosti mýdlovo-kovová směs (Brylawski et al. 2015). Podle složení směsi se voskové válečky lišily výslednou barvou. Nejznámější jsou zřejmě válečky barvy včelího vosku, ty nejstarší byly například smetanově či slonovinově bílé, později se vyskytovaly válečky tmavé, černé. Od složení směsi se také odvíjela trvanlivost záznamu, ty nejméně kvalitní válečky z hnědého vosku ztrácely schopnost věrné reprodukce zvuku již po několika desítkách přehrání.

Od roku 1888 se objevují první komerční nahrávky hudby a divadelních monologů. Komerčně vyráběné zvukové nahrávky byly obvykle opatřeny papírovou krabičkou s víčkem, na kterou později vydavatelé přidávali papírovou etiketu se svým logem. Údaje o nahrávce byly obvykle na kousku papíru, stočeném uvnitř, na krabičce pak mohlo být připsáno katalogové číslo.

1 <https://novyfonograf.cz/>

Voskové válečky jakožto zvukový nosič trpí řadou neduhů v důsledku změny teploty, dále křehnutím věkem, praskáním, opotřebením zvukových drážek. Ve vlhkém prostředí mohou být napadeny plísněmi.

Plastové válečky, vyráběné z celulóidu, se začaly objevovat po roce 1900 a postupně svého voskového předchůdce vytlačily z trhu i díky tomu, že je bylo možné vyrábět výrazně efektivnější technologií lisování, která umožnila jejich masovou produkci (mezi nejznámější značky patří Amberol či Indestructible). Obvykle byly modré, černé či fialové barvy; po roce 1908 bývaly obvykle opatřeny buď kartonovou, nebo sádrovou cívkou. Stejně jako voskové válečky jsou velice křehké, náchylné ke změnám tvaru a poškození drážek, mohou být napadeny plísněmi (Collection ID Guide; Grooved Cylinder).

1.2 Šelakové gramodesky

Šelaková gramodeska se objevuje od konce 80. let 19. století a poměrně dlouhou dobu koexistuje s fonoválečkou. Oproti fonoválečkám se jedná o plochý disk, na který je zvuk v drážkách zaznamenáván laterálně, tj. prohlubně, určující výšku a barvu tónů, jsou vytlačeny ve stranách drážky, nikoliv v jejím středu. Laterální způsob zápisu zvuku byl ve Spojených státech amerických patentován Emilem Berlinerem, který také pro systém zápisu vymyslel název „gramofon“ (přičemž vycházel z Edisonova slova „fonograf“). Disk je vyroben ze směsi šelakové pryskyřice a různých přísad (např. cihlářská hlína, břidlice nebo vápenec jakožto plnivo, uhelný prach jako barvivo či např. bavlněná vlákna pro odolnost). Šelakové gramodesky se již standardně vyráběly lisováním, což umožnilo rychlou masovou produkci a rozmach obchodu s nahrávkami již v období před první světovou válkou. Šelak jakožto termocitlivý materiál po lisování disků procházel chemickou reakcí, při které se postupně vytvrzuje. Tento proces zrání ovšem nekončí vychladnutím desky, a proto u hotových kusů docházelo ke smršťování a křivení a s postupujícím časem také ke křehnutí (Brylawski et al. 2015).

Šelakové desky byly obvykle vyráběny v několika průměrech, od sedmpalcových přes desetipalcové, dvanáctipalcové po největší šestnáctipalcové. Původní přehrávací rychlost pak variovala mezi 30, 70 a 78 otáčkami za minutu (rpm). Ke standardizaci přehrávací rychlosti došlo až v roce 1925 – ustálila se na 78,25 rpm jakožto na kompromisu, vhodném pro většinu přehrávacích zařízení². Na jedné straně disku bylo možné zaznamenat přibližně tři minuty zvuku. První zvukové nahrávky vznikaly čistě akusticko-mechanicky pomocí kornoutu, ústíciho na citlivou membránu, z níž se vibrace přenášely na stylus, který záznam ryl do matrice. Elektroakustické nahrávání bylo představeno až ve 20. letech 20. století po vynálezu mikrofonu.

První gramofonové desky vycházely bez etikety, pouze opatřené matričním či katalogovým číslem, zabalené v jednoduché papírové pošetce. Kulatými etiketami začaly být opatřovány později, stejně jako tištěným obalem a dalšími případnými tiskovinami s údaji o nahrávce.

Kromě již zmíněného křehnutí patří mezi časté problémy šelakových gramodesek podobně jako u fonováleček změna teplot a vlhkosti, napadení plísněmi (parazitují na celulóze, obsažené v discích), při zvlhnutí dochází k rozpraskání povrchu desek či k oddělení jednotlivých vrstev disku (Collection ID Guide; Phonograph Record).

2 The history of 78RPM recordings. Dostupné z: <https://web.library.yale.edu/cataloging/music/historyof78rpm>.

1.3 Historické nosiče ve sbírkách českých paměťových institucí

V České republice se ve sbírkách paměťových institucí nachází odhadem asi 4000 fonoválečků a přibližně 300 000 šelakových gramodesek. Toto číslo se ovšem vztahuje pouze na nosiče, které již byly ve sbírkách pracovníky *Nového Fonografu* identifikovány; je tedy pravděpodobné, že se jich ve skutečnosti zachovalo mnohem více, pouze zatím nebyly v archivech a depozitářích objeveny a popsány.

Ve sbírkách Českého muzea hudby je momentálně zdokumentováno a zkatalogizováno cca 1200 fonoválečků a přibližně 50 tisíc šelakových disků. Do dnešního dne se podařilo kompletně digitalizovat přibližně 1000 ks³. Digitalizace neprobíhá pouze ze záchranných důvodů, ale také jako doplněk katalogizace, neboť digitalizované etikety mohou být vhodným doplňkem k evidenci nosičů (Gössel 2019, s. 49). Digitalizace etiket tak značně snižuje riziko nechtěného poškození nosiče (v tomto případě šelakového disku) při manipulaci katalogizátory či jinými pracovníky. Zároveň, díky elektronické formě, umožňuje v podstatě okamžité sdílení údajů o nosiči jak mezi institucemi, tak pro veřejnost (Gössel 2019, s. 51).

České muzeum hudby ovšem nedigitalizuje pouze svoje sbírky, na základě smluvních vztahů digitalizuje také sbírky jiných institucí. Sbírký historických nosičů má kromě Českého muzea hudby například Akademie věd, Národní technické muzeum, Technické muzeum v Brně a Vlastivědné muzeum v Olomouci, abychom zmínili alespoň ty nejznámější.

Sbírký historických nosičů zvuku nejsou záležitostí výlučně paměťových institucí. Pracovníci projektu *Nový Fonograf* také aktivně vyhledávají pozůstalosti a sbírky v soukromých rukách, které v případě možnosti přejímají do sbírek Českého muzea hudby (to je podmínkou budoucí digitalizace)⁴.

Pokud jde o charakterizaci obsahu fonoválečků a šelakových gramofonových desek, převažují hudební záznamy, klasická hudba, záznamy operních představení, vokální a instrumentální skladby. Dále se na nosičích objevují například záznamy divadelních a loutkových představení (Hnilička 2020) či mluvené slovo. Mnohé nahrávky jsou významným kulturně historickým dokladem; z těch nejvýznamnějších můžeme jmenovat např. záznamy písní Osvobozeného divadla, představení Emy Destinnové či Růženy Maturové (Cíl 2020) či letos objevený záznam projevu prvního československého prezidenta T. G. Masaryka ze dne 7. 3. 1930 (Vetešková 2015).

1.4 Poznámka k digitalizaci historických nosičů

Ohrožení obsahu neseného těmito nosiči nespočívá jen v degradaci samotného nosiče, ale i v zastarávání a postupném mizení potřebné technologie i odpovídajících znalostí, ale také odborníků. Digitalizace těchto nosičů je tedy v podstatě jediným způsobem, jak dochované nahrávky zpřístupnit veřejnosti či pro badatelskou činnost, a zároveň zabezpečit jejich uchování pro budoucí generace.

Pro úspěšnou digitalizaci historických médií je nezbytné, aby se proces již od samého počátku vyvaroval postupů vedoucích k nejrůznějším zkreslením a anomáliím, které mohou vznikat při přenosu analogového záznamu z originálu. Původní přístroje pro přehrávání jsou dnes muzejními exponáty a jejich technické parametry naprosto neodpovídají potřebám digitalizace. Pro digitalizaci fonoválečků se proto používají spe-

3 Dle odhadu pracovníků projektu *Nový Fonograf*.

4 Viz sbírkotvorná koncepce Národního muzea. Dotupné z: <https://www.nm.cz/o-nas/struktura-a-kontakty/dulezite-dokumenty>.

cializované přístroje využívající moderních technologií, jako je Archéophone⁵, využívající moderní přenosky, nebo zařízení Endpoint, instalované roku 2018 v Národním Muzeu v Praze, které je dokonce schopno přehrávat fonoválečky bezkontaktně pomocí laseru⁶.

U šelakových gramodesek je situace jiná, protože se vývoj gramofonů dodnes nezačal. Proto je možné využít upravených high-endových gramofonů, případně lze z dostupných součástek zkompletovat přístroj podle vlastních představ. Pro digitalizaci byl například sestaven gramofon se čtyřmi ramínky, který dovoluje současně použít čtyři různé přenosky⁷. Volba vhodné přenosky je klíčovou otázkou pro správné přehrání historických desek, a proto se při ní vychází z analýzy příslušné desky, která může zahrnovat i mikroskopickou kontrolu stavu žlábků (National Recording Preservation Board 2006). Pro nejnáročnější úlohy byl vyvinut systém optického čtení záznamu s označením Irene⁸. Jeho výhodou je, že může skenovat i disky, které jsou rozlomené.

2 Balíček SIP

Jak bylo v úvodu uvedeno, digitalizace analogových zvukových dokumentů je jedním ze způsobů, jak uchovat nesený obsah pro další generace. Převod analogové informace na informaci digitální je však jen jedním z kroků komplexního procesu digitalizace a dlouhodobé archivace. Aby byl informační obsah digitálního dokumentu použitelný a srozumitelný i v budoucnu, je třeba, aby již proces tvorby digitálních dokumentů probíhal s ohledem na jejich dlouhodobou udržitelnost. Hovoří se tedy o standardizaci procesu tvorby digitálních dokumentů a specifikuje se podoba balíčku SIP (submission information package), tj. podoba dat, v níž dlouhodobé úložiště data přijme k dlouhodobému uchování. Podobu balíčku SIP pro uložení v úložišti Národní knihovny ČR ukládá Standard NDK⁹, který definuje strukturu balíčku SIP, metadatový popis, povolené datové formáty a pravidla pro identifikátory. Z dat dodaných v podobě balíčku SIP je pak v procesu příjmu úložištěm LTP (Long-term Preservation) vytvořen balíček AIP (archival information package), tj. taková podoba dat, která je následně předmětem ochranných aktivit. V aktuální praxi LTP úložiště NK ČR a Standardu NDK je výsledný balíček AIP téměř shodný s dodaným balíčkem SIP¹⁰, AIP navíc obsahuje úložištěm vygenerovaný soubor xml sloužící ke správě dat (např. verzování).

5 https://www.archeophone.org/warcheophone_specifications.php

6 <https://novyfonograf.cz/pokrok/cislo-pet-zije-zachrana-fonovalecku/>

7 <https://archive.org/details/georgeblood>

8 <https://www.nedcc.org/audio-preservation/understanding-irene>

9 NDK – Národní digitální knihovna; <https://old.ndk.cz/standardy-digitalizace/metadata>.

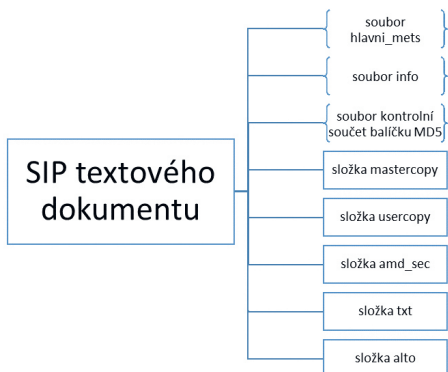
10 Zde záleží na přístupu konkrétního repozitáře, dlouhodobého úložiště. Standard NDK předepisuje podobu balíčku SIP tak, aby nebyla nutná další úprava dat a úložiště přijalo data v takové podobě, která je vhodná k uchování, a v podobě, jakou její producent zamýšlel, tj. výsledná podoba dodaných dat je pod kontrolou producenta dat; příjemce/archivátor dat ale určuje, co od takových dat očekává. Některá úložiště mohou s dodanými daty provádět složitější operace – extrakce metadat ze souborů a následné mapování do metadatových schémat, normalizace (konverze) formátů apod., tj. dlouhodobé úložiště resp. archivátor dat má pod kontrolou tvorbu archivační podoby dat z dodaných dat. Výhodou druhého přístupu je právě kontrola produkce dat, které úložiště potřebuje pro dlouhodobé uchování, nevýhodou je, že operace nad dodanými daty mohou být poměrně složité a že by ideálně měly být doplněny kontrolou správnosti takové transformace. Výhodou prvního přístupu je, že úložiště nemusí provádět úpravu dat a kontrolu správnosti této transformace, nevýhodou je, že úložiště nemá pod kontrolou produkci dat, které může při archivaci potřebovat, může však alespoň tyto informace v balíčku SIP vyžadovat či doporučovat, tak jak to činí Standard NDK.

2.1 Struktura balíčku SIP

Balíček SIP, odevzdávaný dlouhodobým úložištěm k archivaci, má ve Standardu NDK předepsanou podobu. Podoba balíčku SIP pro zvukové dokumenty vychází z podoby balíčku SIP pro textové digitalizáty, má tedy podobnou strukturu (viz obr. 1 a obr. 2). Zůstávají společné soubory mimo složky, tj. hlavní *mets*, obsahující záznam popisné vrstvy a strukturální metadata, soubor *info* a soubor s údajem *fixity* balíku (MD5). Zatímco u tištěných digitalizátů jádro balíku tvoří složky s archivační (*mastercopy*) a uživatelskou kopií (*usercopy*) obrazů a na to dále navázané složky s administrativními metadaty (*amdsec*), *alto* a *txt* (viz obr. 1), je balíček SIP u zvukových digitalizátů rozmanitější; obsahuje složky nejen pro zvukové soubory (složky *source_audio*, *master_copy_audio* a *user_copy_audio*), ale i pro soubory obrazové. Obrazové soubory tvoří zejména přílohy zvukového dokumentu, obaly (v případě gramodesky papírové pošetky či krabice v případě kolekce; v případě fonoválečku krabička a víčko), doprovodné texty či brožury, sken disku nebo fonoválečku či jeho okruží. Obrazové soubory jsou v balíčku uloženy dvakrát, v archivační a ve zpřístupňující kvalitě, tedy stejně, jako je to předepsáno Definicemi metadatových standardů pro digitalizaci tištěných dokumentů¹¹. Také zvukové soubory se v balíčku SIP vyskytují povinně v archivační a zpřístupňující kvalitě. Je však také možné (nikoliv povinně) přidat do balíčku další složku se zvukovými soubory, odvozenými z původních, neupravených zvukových souborů, jejichž kvalita byla v procesu zpracování vylepšena.¹²

Složky obsahující výstupy optického rozpoznávání textů (OCR), tj. složky se soubory *alto* a *txt*, jsou v balíčku SIP pro zvukové dokumenty nepovinné. Optické rozpoznávání textů z etiket je často nekvalitní a chybné, v některých případech není dostupný dokument, který by doprovodný text vůbec obsahoval. Z tohoto důvodu je v případě zvukových dokumentů generování souborů *alto* a *txt* nepovinné.

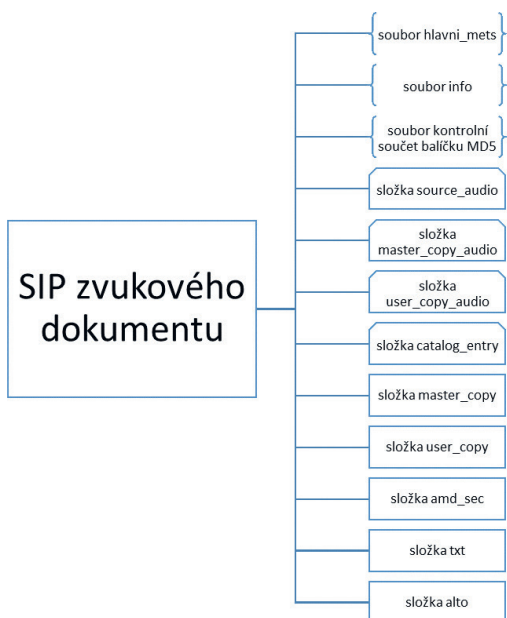
Součástí balíčku SIP pro zvukové dokumenty může být i katalogizační záznam transformovaný do formátu *marcxml* (složka *catalog_entry*). Jedná se o jakousi pojistku toho, že budou uchována všechna popisná metadata a jejich struktura, tak, jak to zatím umí zřejmě jen metadatové schéma MARC21, resp. MARCXML, a budou tak dostupná pro případně pozdější konverze.



Obr. 1 Diagram balíčku SIP textových digitalizátů (archiv autorů)

11 https://old.ndk.cz/standardy-digitalizace/DMFmonograf_14_final.pdf a https://old.ndk.cz/standardy-digitalizace/DMFperiodika_18_final.pdf.

12 Například odstranění praskání apod. Vytvoření remasterované verze je možné považovat také za autorskou tvorbu, proto je součástí balíčku. Nesmí však nahradit primární digitalizáty, které vypovídají o stavu nosiče v době, kdy byl digitalizován, a které se pak mohou stát předmětem úpravy později, kdy budou dostupné lepší nástroje na zkvalitnění zvuku.



Obr. 2 Diagram balíčku SIP zvukových digitalizátů (archív autorů)

2.2 Metadatový popis digitálních zvukových dokumentů

2.2.1 Popisná metadata digitálních zvukových dokumentů ve Standardu NDK

Popis digitalizovaných zvukových dokumentů v Národní knihovně ČR vychází z katalogizačních záznamů analogových předloh, převáděných pomocí transformačních šablon do formátu MODS a METS. Stejně jako v případě popisné vrstvy pro textové dokumenty se vychází z katalogizačních pravidel tzv. „minimálního záznamu“, který obsahuje klíčové údaje k identifikaci díla. V případě zvukových dokumentů ovšem významné množství dostupných vlastností souvisí s technickými specifiky dokumentu, pro katalogizační popisnou vrstvu podklady poměrně často chybí nebo jsou pouze strohé. Tato strohost je dána jednak povahou nosičů (na fonoválečkách ani na deskách není dostatek fyzického prostoru pro jakékoliv údaje), a obaly mnohdy chybí, nebývají opatřeny kompletními údaji či dochází k případům, kdy je např. fonováleček uložen v obalu od jiného dokumentu. Z těchto důvodů bylo nutné zkrátit seznam povinných elementů pro popisnou vrstvu tak, aby byl praktický popis vůbec proveditelný. Abychom však zároveň umožnili co možná nejširší popis nosičů, které doprovodnými zdroji údajů v různě podrobném provedení opatřeny jsou, rozhodli jsme se také v maximální možné míře zachovat širší elementů, používaných při popisu jiných typů digitalizátů. V konečném důsledku tedy z elementů MODS byly odstraněny pouze ty, které v případě zvukových dokumentů není možné vyplnit nikdy. U ostatních elementů

byla většinou snížena povinnost¹³, a to i u těch, které spadají do kategorie minimálního záznamu (jako je například skupina elementů spadajících pod <name>). Víceméně povinnými tedy zůstaly pouze elementy pro <title>, <genre>, <typeOfResource> (údaje důležité pro vyhledávání a zobrazování ve zpřístupňujících aplikacích, v tomto případě hlavně digitální knihovna Kramerius), <originInfo> (zejména pro důležitý údaj o datu vydání a vydavateli), <physicalDescription>, <identifier> a <recordInfo> pro zaznamenání údajů o samostatném záznamu.

Popis zvukových dokumentů vychází z popisu tištěných digitalizovaných dokumentů, ze stejných zásad a principů¹⁴. Plnění všeobecných sekcí (info.xml, metsHdr, strukturální mapy...) a pokyny k vyplňování jmen apod. jsou stejné. Popis nejvyšší úrovně (sound collection) vychází z katalogizačního záznamu, nekopíruje jej ale úplně. V některých případech je naopak popisování zvukových dokumentů velmi specifické, čemuž jsme naše Definice metadatových formátů pro digitalizaci zvukových dokumentů¹⁵ museli přizpůsobit.

Zásadní rozdíl je zejména v chápání intelektuální entity¹⁶ a rozdělení úrovní popisu. U digitalizátů tištěných předloh, např. monografií, je intelektuální entitou svazek monografie (tzv. volume v textu Standardu NDK), u periodik se jedná o jedno číslo periodika, tj. jedná se o jednotlivé fyzické nosiče, jež odpovídají intelektuální entitě. Dle sdělení pracovníků z Českého muzea hudby u gramodesek ani fonováleček nemusí jeden nosič obsahovat jedno ukončené dílo, ani samostatné dílčí dílo (tj. nosič není jako takový nejvyšším celkem); po konzultaci s pracovníky projektu *Nový Fonograf*¹⁷ byl intelektuální entitou stanoven samotný zvukový záznam (nikoliv gramofonová deska, jak stanovil prvotní návrh Standardu NDK). Zvukový záznam, například opera, může být nahrán na celou kolekci gramofonových desek, a přitom se stále jedná o jedno dílo. Jednotlivé desky tedy nemohou vystupovat jakožto intelektuální entita. Tento princip byl užit i u fonováleček, ačkoliv v jejich případě je většinou na jednom nosiči obsaženo pouze jedno dílo (typicky jedna skladba). Intelektuální entitu tedy tvoří zvukový záznam (např. opera), který může mít několik skladeb (analogicky ke kapitolám tištěných monografií). I pro úroveň skladby probíhala diskuse nosič (strana gramofonové desky) vs. dílo a ve

13 V rámci Definic metadatových formátů se používají pro vyjádření povinnosti elementů/atributů hodnoty *mandatory* (M), *mandatory if available* (MA), *recommended* (R), *recommended if available* (RA) a *optional* (O). Pod zmíněným „snížením povinnosti“ máme na mysli, že element/atribut, který byl v předpisu pro tištěné dokumenty povinný (M), může mít u zvukových dokumentů povinnost MA nebo nižší, nalikoliv informace nemusí být vzhledem k strohému katalogizačnímu popisu k dispozici.

14 Všechny aktuální a neaktuální, oficiálně vydané Definice metadatových formátů pro konkrétní typy dokumentů jsou volně k dispozici na <https://old.ndk.cz/standardy-digitalizace/metadata> a případy popsané níže/výše jsou v nich dohledatelné.

15 https://old.ndk.cz/standardy-digitalizace/DMFzvuk03_web.pdf a https://old.ndk.cz/standardy-digitalizace/dmf_fonovalecky_draft_0.1.pdf

16 Intelektuální entitou (*Intellectual Entity*) rozumíme „jednotlivý intelektuální nebo umělecký výtvor (*creation*), který je považován za relevantní pro cílovou komunitu v kontextu digitální archivace“. (Definice dle PREMIS EDITORIAL COMMITTEE. PREMIS Data Dictionary for Preservation Metadata [online]. Version 3.0. Washington (DC): Library of Congress, June 2015, rev. Nov 2015, viii, 273 s. [cit. 2017-10-10]. Dostupné z: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.)

17 Pracovní tým projektu *Nový Fonograf* je složen z odborníků z Národního muzea, Národního technického muzea, Akademie věd ČR a dalších institucí. Metadatovou část jsme vyvíjeli zejména s úzkou pracovní skupinou Filipa Šíra, DiS., který je zároveň projektovým manažerem *Nového Fonografu*.

spolupráci s Českým muzeem hudby bylo opět upřednostněno dílo nad nosičem. Volba skladby jakožto úroveň popisu byla upřednostněna z ryze praktického hlediska, vycházejícího ze specifika popisovaného. Záznamy skladeb na šelakových gramodeskách totiž nemusí být uspořádány hierarchicky podle pořadí stran (tj. např. disk 1, strana A, strana B, disk 2, strana A, strana B etc.), ale podle způsobu přehrávání může být pořadí třeba následující: disk 1, strana A, disk 2, strana A, disk 1, strana B, disk 2, strana B etc.

Zvuková stopa samotná (tj. track) se pod názvem „část skladby“ potom nachází na hierarchicky nejnižší úrovni popisu a slouží k zaznamenání údajů o nižších podcelcích skladby – například větách ze symfonie.

Oproti textovým digitalizátům se v případě zvukových dokumentů liší také skladba strukturálních map, do kterých se zaznamenávají dva typy digitalizačních výstupů – tj. zvukové i obrazové soubory¹⁸.

Z důvodu přítomnosti zvukových a obrazových souborů a zápisu technických metadat ke každému z nich vyvstala potřeba rozlišit mezi nimi i v názvu souborů, přestože číslování mohou mít stejné. U technických metadat zvukových dokumentů se do názvu vloží slovo „audio“. Příkladem může být název následujícího souboru: amd_mets_audio_mb15c210-d3f7-11e6-a1de-001018b3eb4c_0001.

U obrazových dokumentů pak název stejného souboru bude: amd_mets_mb15c210-d3f7-11e6-a1de-001018b3eb4c_0001.

Při plnění některých polí (např. *mods:genre*, *mods:typeOfResource*, *mods:form*) jsou pevně předepsané hodnoty jiné než v případě tištěných dokumentů – to je podmíněno jiným typem dokumentu. Jako příklad můžeme uvést hodnotu v *mods:typeOfResource*: u tištěných dokumentů je dána hodnota „text“, zatímco u zvukových dokumentů se může podle potřeby zvolit z hodnot „sound recording“, „sound recording-musical“ nebo „sound recording-nonmusical“¹⁹. V některých případech se tyto hodnoty mohou ze stejného důvodu lišit i mezi gramofonovými deskami a fonografickými válečky.

Zvukové dokumenty také používají vlastní identifikátory. Kromě UUID a URN:NBN, které jsou užívány v NDK, jde jmenovitě o objednáací nakladatelské číslo (*issue number*), matriční číslo (*matrix number*), číslo vydavatele (*music-publisher*) či lokální identifikátory.²⁰

Elementy jako *mods:abstract*, *mods:url* nebo *mods:part* také v současnosti nemají při popisu zvukových dokumentů své využití. Na druhou stranu se ale častěji k jednotlivým metadatovým elementům přidává atribut *type* a je tak možné popsat další vlastnosti popisovaného dokumentu. Například v případě elementu *mods:note* je možné do atributu *type* zapsat více hodnot, které dále specifikují typ uváděné poznámky. V případě zvukových dokumentů bude jistě užitečný nejen typ poznámky „*statement of responsibility*“²¹ (vyjádření odpovědnosti za dílo), ale například i typ poznámky „*performers*“ (jména vystupujících), „*venue*“ (místo konání) nebo „*creation/production credits*“ (detaily o produkci).

18 V důsledku toho se logická strukturální mapa stává složitější. Je totiž možné pod každou stranu desky přiřadit všechny nahrávky, které se na ní nacházejí, stejně jako odlišit každý typ přílohy.

19 Tyto hodnoty se odvíjejí z kódů zapsaných v katalogizačním záznamu (Leader/06). V rámci DMF jsme je převzali tak, jak je do MODS převádí Library of Congress.

20 Tyto identifikátory se mohou vázat jak k celému válečku, tak ke konkrétní nahrávce; neobjevují se ale vždy všechny. V katalogizačním záznamu je nalezneme v poli 028, přičemž jejich typ je určený hodnotou prvního indikátoru.

21 V datech, resp. v metadatovém záznamu se zapíše: *mods:note type = „statements of responsibility“* a odpovídající text z poznámky z katalogizačního záznamu.

2.2.2 Rozdíly mezi popisem gramofonových desek a fonografických válečků

Popis gramofonových desek²² a fonografických válečků²³ je v NDK v zásadě stejný. Přesto v rámci metadatového popisu můžeme najít několik rozdílů, které vycházejí z povahy samotných nosičů jak na úrovni popisných metadat, tak i na úrovni metadat technických. Liší se některé z pevně předepsaných hodnot polí.

Popis pro fonografické válečky vychází z popisu pro gramofonové desky. Jak již bylo zmíněno, popis fonografických válečků je ale stručnější, a proto v mnoha případech došlo oproti popisu gramodesek ke snížení množství povinných elementů a atributů (v řeči Standardu NDK byla u elementů a atributů snížena povinnost a nově jsou nepovinné).

Nejokřivnější rozdíl je již zmíněný rozdíl v použití jednotlivých úrovní. Popis fonografických válečků nepočítá s úrovní pro část skladby (*sound part*), jelikož se na tomto nosiči vyskytují pouze krátké nahrávky, jejichž dělení na menší části je tudíž velmi nepravděpodobné.

Jako úroveň „příloha“ (*supplement*) se v obou případech popisují skeny samotného nosiče. V případě gramofonových desek však rozlišujeme více typů přílohy: skenuje se samotná deska (*imgdisc*), obal desky (*cover*) a případný *booklet*, tj. brožura ke zvukovému nosiči. U fonografických válečků se skenuje pouze obal válečku (včetně víčka a spodní části obalu, protože i ty mohou nést důležité informace) a okruží samotného válečku (*imgdisc* – často se skenuje zvlášť, spolu se zátkou). V metadatech je následně zaznamenán typ přílohy, a to v atributu typu elementu *mods:genre*. V případě válečků se typ přílohy vyplní také rovnou do *mods:title*; u gramofonových desek je to jen do *mods:partName* (nicméně se tu může vyskytovat i název přílohy jako takové).

Změna nastala i při vyplňování atributu *eventType* v elementu *<mods:originInfo>*. Praxe ukázala, že v případě fonografických válečků jsou často dostupné jenom informace o nakladateli (*publication*) nebo distributorovi, prodejci (*distribution*), případně o obou. Ostatní hodnoty, běžně používané u tištěných dokumentů (*production*, *manufacture*, *copyright*), se nevyskytují, a nejsou tedy součástí popisných metadat v návrhu pro popis fonografických válečků.

2.2.3 Administrativní metadata

Součástí archivačního balíčku jsou vedle popisných a autorskoprávních metadat dále administrativní metadata. Ta zahrnují informace jednak o technických vlastnostech jednotlivých počítačových souborů, v případě zvukových dokumentů i analogových nosičů, a také o aktivitách, jež jsou s digitálními objekty spojeny – zvláště důležité jsou aktivity týkající se vzniku archivovaného digitálního dokumentu. Tyto informace pak slouží k provádění ochranných aktivit nad daty (např. migrace dat a kontrola výsledku migrace) a ke kontrole a doložení autenticity uchovávaného obsahu. Cílem tak je v co největší míře zaznamenat signifikantní vlastnosti digitálních dokumentů a popsat dostatečně všechny události, které archivované objekty nějak mění. Konečný seznam signifikantních vlastností digitálních zvukových a obrazových dokumentů ještě nebyl stanoven, resp. odvíjí se do určité míry od potřeb cílových uživatelů a tedy i přístupů jednotlivých archivních institucí. Standard NDK aktuálně počítá se zaznamenáním minimálně těch vlastností, které jsou ze souborů schopny extrahovat dostupné softwarové nástroje. U zvukových dokumentů se vyžaduje zápis následujících vlastností: délka nahrávky (*duration*), velikost souboru (*file size*), bitová hloubka (*bit depth*), vzorkovací frekvence (*sample rate*),

22 https://old.ndk.cz/standards-digitalizace/DMFzvuk03_web.pdf

23 https://old.ndk.cz/standards-digitalizace/dmf_fonovalecky_draft_0.1.pdf

počet kanálů (*number of channels*), formát souboru (*format*), druh komprese, metoda převodu dat do digitální podoby, pořadí bytů (*endianita*). U obrazových souborů se zapisuje: velikosti souboru (*file size*), bitová hloubka (*bit depth*), rozlišení, rozměry v pixelech, druh komprese, barevný prostor, formát a jeho verze.

V archivačním balíčku se pro zápis výše uvedených vlastností používají metadatová schémata PREMIS, MIX, AES57 a NK.

Metadatové schéma MIX slouží k popisu obrazových dokumentů, jež vznikají při digitalizaci obalů, etiket i samotných nosičů (tj. fotografie desky či válečku). Popis těchto souborů je stejný jako v případě digitalizace periodik a monografií. Metadatové schéma MIX umožňuje popsat vlastnosti digitálního obrazu (viz výše) i okolnosti jejich vzniku (výrobce skeneru, název a sériové číslo skeneru apod.).

Pro digitální zvukové dokumenty je obdobně použito metadatové schéma AES57. AES57 umožňuje popsat nejen digitální zvukový soubor, ale i analogový nosič. Ve Standardu NDK je v současné době povinné popsat jen digitální zvukový soubor (konkrétní vlastnosti viz výše).

Popis analogového nosiče je ve Standardu NDK volitelný, a to z toho důvodu, že je časově velmi pracný a náročný na znalosti zpracovatele (popisuje se materiál, vlastnosti drážky apod.). Záleží tedy na zhodnocení uživatelů standardu, zda usoudí, že popis nosiče může být pro zamýšlené uživatele užitečný, zajímavý. U nosiče je možné popsat jeho fyzické vlastnosti, tj. použité materiály, tloušťku jednotlivých materiálových vrstev, rozměry nosiče a jeho obalu, směr přehrávání, rychlost přehrávání, orientaci drážky apod. Metadatové schéma AES57 neposkytuje prostor pro zápis informací o okolnostech vzniku digitálního souboru, tak jak to známe ze schématu MIX. Schéma MIX u digitálních obrazů umožňuje popsat typ skenovacího zařízení, výrobce skeneru, model a sériové číslo skeneru, skenovací software apod. Částečně tuto funkci plní metadatové schéma PREMIS, které k tomuto účelu bylo ve Standardu NDK užito (viz dále).

Metadatové schéma PREMIS je u zvukových dokumentů použito obdobně jako u digitalizovaných tištěných dokumentů, tj. k popisu technických vlastností a především k popisu procesu vzniku digitální verze dokumentu. Umožňuje zapsat, jaké události při vzniku digitálního dokumentu proběhly (digitalizace, konverze, optické rozpoznávání textu apod.) a jaké zařízení (SW a HW) bylo přitom použito. Pro případ, že je nutný podrobnější popis, než dovolují dedikované elementy standardu PREMIS, je umožněno tento popis doplnit jiným metadatovým standardem, jehož elementy se vloží do elementu *agentExtension*. Pro zvukové dokumenty bylo tedy vytvořeno lokální metadatové schéma s názvem NK pro údaje, které nebylo možné zapsat do schémat AES57, PREMIS a dalších, jež byly v té době prověřovány. Schéma NK ČR umožňuje blíže popsat hardware použitý v procesu vzniku digitálních dokumentů – tj. byly vytvořeny elementy pro zápis výrobce zařízení, sériového čísla, nastavení zařízení a role v procesu digitalizace.

Pro vyplňování některých údajů se používají kontrolované slovníky, jejichž cílem je zajistit konzistentní zápis hodnot elementů, což by mělo přispět ke srozumitelnosti uchovávaného obsahu pro budoucí uživatele a umožnit automatizovanou práci s daty. Standard NDK pracuje s kontrolovanými slovníky, které používají metadatová schémata, jež Standard NDK implementuje (např. u schémat AES57 a PREMIS), v některých

případech byl kontrolovaný slovník vytvořen lokálně pro potřeby Standardu NDK.²⁴ Lokální kontrolovaný slovník je předepsán u elementu pro zápis role zařízení v procesu digitalizace nosiče (role: Reproducer, Amplifier, ADC, Recorder, Digital audio editor) a pro další elementy jsou předepsané konkrétní hodnoty, které se mají vyplnit (formát, metoda převodu do digitální podoby apod.). Některé kontrolované slovníky jsou vytvářeny na základě doporučených hodnot či přímo existujících slovníků (např. od Kongresové knihovny), další vznikají na základě průzkumu praxe jiných institucí. Výše zmíněný kontrolovaný slovník rolí zařízení v procesu digitalizace vznikl průzkumem praxe Britské knihovny, knihovny Harvardovy univerzity a Univerzity v Indianě, kdy jsme získali přehled různých rolí a jejich anglických označení a návrh byl následně konzultován s odborníky v Českém muzeu hudby.

2.2.4 Porovnání metadatového popisu pro zvukové dokumenty Britské knihovny a návrhu pro řešení NDK

V zahraničí v současnosti zvukové dokumenty zpracovává několik institucí, z těch významných např. Indiana University, University of Santa Barbara (USBC Audio Archive), Berkeley University nebo například Britská knihovna (The British Library²⁵), a v dalších se jejich digitalizace a následný popis chystá. Všeobecně lze říci, že zájem o digitalizaci zvukových dokumentů roste, přičemž každá instituce vyvíjí své originální řešení.

Díky spolupráci s odborníky z Českého muzea hudby, kteří jsou v kontaktu s pracovníky Britské knihovny, jsme měli možnost porovnat naše řešení popisu zvukového dokumentu s tím, které používají ve Velké Británii. Obecně lze říci, že metadatový popis zvukového dokumentu podle Britské knihovny z velké části odpovídá našemu návrhu pro NDK. Popisují ale jen zvukové nahrávky; skeny válečku součástí balíčku nejsou. Jednotlivé úrovně popisu jsou také jinak členěny. K popisu využívají kontejnerový standard METS, stejně jako je tomu v NDK. Popisná metadata se ale v balíčku neuvádějí, na katalogizační záznam konkrétního dokumentu je pouze odkázáno prostřednictvím identifikátorů ARK, unikátních pro Britskou knihovnu. Technická metadata jsou zapsána pomocí standardů PREMIS, AES 60 a lokálně vytvořených schémat mediaMD a blaph. Metadatový standard PREMIS je v Britské knihovně použit k zápisu přibližně stejných informací jako ve Standardu NDK a k extrakci technických metadat používáme stejné nástroje (DROID, JHOVE a jiné), ve speciálně vytvořených standardech uvádí Britská knihovna i informace o zařízení, na kterém byl fonováček přehrán či migrován, informace o jeho technických parametrech a navíc také, kdo záznam zpracoval. Ve schématu mediaMD jsou vyjádřeny základní informace o dané nahrávce, jako např. velikost, formát, délka, bitová frekvence (data o souboru) nebo typ komprese, kodek, počet kanálů, bitová hloubka nebo vzorkovací frekvence (ve Standardu NDK uloženo ve schématu AES57). Balíček obsahuje také části fileSec, logickou strukturální mapu, fyzickou struk-

24 Standard NDK tak například u elementu `preservationLevelValue` ze schématu PREMIS vyžaduje hodnotu „*deleted*“ nebo „*preservation*“, u elementu „*formatName*“ odkazuje na kontrolovaný slovník MIME type a u elementu „*formatRegistryName*“ vyžaduje hodnotu „PRONOM“. U plnění elementů ze schématu AES57 předepisuje Standard NDK kontrolovaný slovník u elementů a atributů: *format*, *schemaVersion*, *byteOrder*, *useType*, *identifierType*, *direction*, *channelAssignment*, *soundField*, a pro jednotky měření, označení rolí jednotlivých materiálových vrstev fyzického nosiče, směr přehrávání nosiče, orientace drážky a typ drážky.

25 <https://www.bl.uk/>

turální mapu a structLink.²⁶ Ve fyzické strukturální mapě v Britské knihovně navíc oproti Standardu NDK vymezují i jednotlivé úseky nahrávky. Autorskoprávní metadata jsou v Britské knihovně popsána pomocí Open Digital Rights Language (World Wide Web Consortium 2018), který umožňuje uvádět tvrzení o užití obsahu a služeb. V rámci něj se pak využívá i Dublin Core pro zapsání statutu a provenience dokumentu.

2.3 Souborové formáty vhodné pro dlouhodobou archivaci a pro zpřístupnění

V letech 2016–2017 byl Národní knihovnou ČR proveden průzkum souborových formátů s cílem určit vhodný archivační formát pro dlouhodobé uložení digitálních zvukových dat. Na základě mezinárodně uznávaných doporučení IASA (IASA Technical Committee 2017), ARSC (Brylawski 2015) a LOC²⁷ byl v roce 2017 Národní knihovnou ČR vybrán souborový formát WAVE, který se stal součástí národního standardu pro digitalizaci zvukových dokumentů. Formát WAVE je zřejmě i v současnosti nejrozšířenějším formátem pro archivaci zvukových dokumentů v paměťových institucích, případně formát BWF, který je od formátu WAVE odvozen (je rozšířen o metadatové segmenty určené k zaznamenávání údajů specifických pro rozhlasová vysílání). Formát BWF a zmíněné metadatové segmenty využívají například na Harvardově univerzitě a na univerzitě v Indianě.

Dalším formátem, který se u paměťových institucí v souvislosti s archivací zvukových dokumentů vyskytuje, je formát FLAC, který byl v letech 2016–2017 pro Standard NDK také prověřován. Tehdy jako konečný archivační formát vybrán nebyl, protože ve srovnání s formátem WAVE/BWF byl stále málo rozšířen, neexistoval validační nástroj pro formát a významné dokumenty jednoznačně doporučovaly formát WAVE/BWF. Nicméně i nadále je formát FLAC průběžně sledován. Britská knihovna formát FLAC označila ve svém hodnocení v roce 2018 za formát vhodný pro dlouhodobou archivaci, ale upozorňuje, že jeho široké rozšíření mezi uživateli (jedna z požadovaných vlastností pro archivační formáty) je spíše málo pravděpodobné (Russo 2018). V současnosti se ale formát FLAC z domény takzvaných audiofilů přesouvá do běžné spotřeby. Vydavatelé hudby, včetně Supraphonu, ho již využívají v distribuci a nabízejí v tomto formátu některá alba. Pro přehrávání i úpravy formátu FLAC existuje řada programů, a to jak pro počítače, tak pro mobilní zařízení. Je rovněž v nabídce kodeků u multimediální spotřební elektroniky a je využíván některými streamovacími službami (např. DEEZER a TIDAL). Mezi archivními institucemi se formát FLAC začíná prosazovat teprve v poslední době a k jeho prevalenci nedojde zřejmě proto, že potřeba komprese už není tak vysoká. Formát WAW se již také začíná objevovat i v online prodeji.²⁸ Formát FLAC však přinejmenším zůstane formátem pro mobilní zařízení. Pro Standard NDK a lokální digitalizační projekty tak formát FLAC tedy spíše připadá v úvahu ne jako cílový formát při digitalizaci analogových zvukových dokumentů, ale jako přijatelný formát pro archivaci

26 Zmíněné sekce slouží na vyjádření struktury a dokumentu a vztahů mezi jednotlivými soubory. Část fileSec obsahuje výčet všech souborů v balíčku SIP podle jejich typu. Logická strukturální mapa se využívá k správnému řazení dokumentu v zobrazovacích aplikacích, zatímco fyzická strukturální mapa obsahuje vazby každé stránky ke konkrétním souborům. Sekce structLink se v rámci NDK využívá pro výčet jednotlivých stránek a jejich přiřazení k dokumentu, příloze, případně ke konkrétní kapitole článku či obrázku a zápis hyperlinků.

27 Kongresová knihovna vydává každý rok doporučení ohledně vhodných archivačních formátů, viz <https://www.loc.gov/preservation/resources/rfs/>.

28 <https://www.supraphonline.cz/napoveda-hi-res.html>

digitálních zvukových dokumentů, které byly vytvořeny a uloženy do formátu FLAC, tj. není u nich důvod je konvertovat do aktuálně preferovaného formátu WAVE/BWF. Nicméně je na tomto místě ještě třeba uvést, že v hodnocení formátů od švýcarské skupiny expertů KOST-CECO si formát FLAC vede lépe než formát WAVE²⁹, oba formáty jsou ale považovány za velmi vhodné pro dlouhodobou archivaci (KOST-CECO 2019).

Pro zpřístupňující kopie je celosvětově využíván formát MP3; je v tomto smyslu využíván i institucemi zabývajícími se digitalizací fonováleček, přičemž pro některé je přijatelný i v kontejneru Quicktime, pro snazší zpřístupnění uživatelům počítačů Apple. Kombinaci formátů WAVE jako archivního formátu a MP3 jako zpřístupňujícího formátu využívá i projekt digitalizace fonováleček Kalifornské Univerzity v Santa Barbaře.³⁰

Přehled formátových politik dalších institucí je uveden v tabulce 3. Je zde vidět výrazná preference formátu WAV, ale zajímavá je též poměrně silná pozice formátu AIFF, která je nejspíše důsledkem oblíbenosti streamovacích platform firmy Apple. Ze srovnání názvů označení kategorií vhodnosti pro archivaci je vidět, že v této oblasti panuje nejednotný přístup a terminologie. Je rovněž nutné upozornit, že ne všechny politiky jasně odlišují, zda se jedná o archivační nebo zpřístupňující formát, a že především v případě MP3 se v některých případech může jednat o jeho uznání jako zpřístupňujícího formátu spíše než archivačního, resp. je pro archivaci přijatelný, pokud obsah existuje jen ve formátu MP3 a nikoliv ve formátu WAVE/BWF. V tabulce jsou uvedeny pouze dva stupně ze tří, tedy high a medium preference. Řada formátových politik těchto institucí má ještě třetí kategorii (low preference, not accepted, unsupported apod.), ale v případě zvukových dokumentů nebyly žádné formáty v této kategorii.

Závěr

Jedním z cílů práce Národní knihovny ČR v projektu *Nový fonograf* v uplynulém období bylo vytvořit národní standard pro digitalizaci vybraných zvukových dokumentů tak, aby odpovídal mezinárodní praxi a lokálním zkušenostem při dlouhodobém uchovávání digitálních dokumentů. Byl navržen metadatový popis, jehož cílem je co nejkompaktněji popsat uchovávané digitální soubory. Vedle toho Standard NDK umožňuje popsat analogový nosič (materiály, tvar) pro případ, že digitalizující instituce zhodnotí, že se jedná o údaje důležité pro jejich určené uživatele. Pro účely popisné vrstvy záznamu byl adaptován již existující soubor elementů ve formátech MODS a METS tak, aby vyhovoval omezeným možnostem abstrahování údajů z nosičů zvukových dokumentů. Ohledně souborových formátů vhodných pro archivační účely je stále doporučován formát WAVE, resp. BWF, jak bylo rozhodnuto již v roce 2017 na základě průzkumu situace v paměťových institucích a významných doporučení (viz výše, IASA, LOC, ARSC, KOST-CECO).

Zdá se tedy, že Standard NDK pro vybrané zvukové dokumenty je připraven pro pilotní provoz, ze kterého budou patrně získány další poznatky pro optimalizaci standardu. Proto počítáme, že v budoucnosti bude Standard NDK v této oblasti nadále vyvíjen v souladu s dalším vývojem poznatků v oblasti dlouhodobého uchovávání, a také upravován na základě poznatků z praxe jednotlivých digitalizujících institucí.

29 Formáty jsou ohodnoceny počtem bodů, tj. bodově si lépe vede formát FLAC, nutno však podotknout, že formát FLAC získal body i za použití komprese, tj. za umožnění uložení dat ve větší hustotě, a je otázka, nakolik je toto kritérium důležité pro hodnocení vhodného archivačního formátu v době, kdy přestává být úložný prostor finančně nákladný.

30 <http://cylinders.library.ucsb.edu/overview.php>

Tab. 3 Přehled formátových politik vybraných institucí. Fialové podbarvené letopočty ve sloupci roku vydání označují údaje, které byly dohledány ve zdrojovém kódu dokumentu a nebyly zobrazeny v dokumentu samotném. Řádky označené na konci hvězdičkou odpovídají institucím, které mají ještě třetí kategorii přijetí. (archiv autorů)

Instituce	rok vydání	BWF		WAV		FLAC		MP3		AIFF		Kategorie přijetí	
		HI	ME	HI	ME	HI	ME	HI	ME	HI	ME	Vysoká (High)	Střední (Medium)
Alberta University Lib. ¹	2018	Rec		Rec								Recommended	Bit-level preservation
Archivemata ²	2019			PrZ				acc				Preservation/access	není
Canada Library and Archives ³	2015	Pre			Acc				Acc		Acc	Preferred	Accepted
Connecticut Uni. Lib. ⁴	2018				Kno					Sup		Supported	Known
Cornell University Lib. ⁵	2019			Hi			Me.		Me	Hi		High	Medium
Deep Blue (Michigan Uni.) ⁶	2011			L1				L2	L1			Level 1	Level 2
Florida Uni. Libraries ⁷	2012			Hi			Me.		Me	Hi		High	Medium
Hawai'i Uni. ⁸	2019									Sup		Supported	Known
Houston University ⁹	2018			HP					MP		MP	High preference	Medium preference
Library Of Congress ¹⁰	2016	Pre		Pre								Preferred	Acceptable
McMaster, Ontario, CAN ¹¹	2019			NoN		NoN			Bit			No Normalization	Bit-level preservation
Minnesota University ¹²	2014			Full						Lim		Full support	Limited support
National Archives (USA) ¹³	2019	Pre			Acc	Pre			Acc		Acc	Preferred	Acceptable
North Carolina State Archives ¹⁴	2012	Rec		Rec					Acc		Acc	Recommended	Acceptable
North Carolina State Uni. Lib. ¹⁵	2018				Par					Par		Supported	Partially supported
Northwestern University ¹⁶	nezn.			Hi						Hi		Highly recommended	Moderately recom.
Purdue University Libraries ¹⁷	2012			Sus							Sup	Sustainable	Supported
Simon Fraser Uni. Archives ¹⁸	2017			Pre			Acc		Acc		Acc	Preferred	Acceptable
Smithsonian Archives ¹⁹	2017	Pre			Acc		Acc			Acc		Preferred	Acceptable
Southern Illinois Uni. ²⁰	2008				Kno					Sup		Supported	Known
Tasmanian Archives ²¹	2015	Rec		Rec		Rec			Acc		Acc	Recommended	Acceptable
Texas A&M Uni. ²²	2014	Pre		Pre			Acc					Preferable	Acceptable
UK Data service ²³	2014				Acc	Rec			Acc		Acc	Recommended	Acceptable
Uni. of Washington Libraries ²⁴	2014	Hi										Highest confidence	Medium confidence

¹ https://docs.google.com/document/d/1YQDOzjOObgGN2QiSoJ_X_zPuAHTJmFTfSEnFCCTliu8/edit

² https://wiki.archivemata.org/Format_policies

³ <https://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/Pages/guidelines-file-formats-transferring-information-resources-enduring-value.aspx#m>

⁴ <https://guides.lib.uconn.edu/opencommons/fileformats>

⁵ <https://guides.library.cornell.edu/economics/formats>

⁶ <https://deepblue.lib.umich.edu/static/about/deepbluepreservation.html>

⁷ <https://libraries.flvc.org/documents/181844/502298/Recommended+Data+Formats/0b25496f-33ac-4f56-9550-12c34f3d5d7c>

⁸ <http://manoa.hawaii.edu/library/wp-content/uploads/2019/06/ddr-policies.pdf>

⁹ <https://libraries.uh.edu/wp-content/uploads/DigitalPreservationPolicyExternal-2018-02-revision.pdf>

¹⁰ <https://www.loc.gov/preservation/resources/rfs/stilling.html#ogdigital>

¹¹ <https://digitalarchive.mcmaster.ca/node/58>

¹² <https://conservancy.umn.edu/pages/policies/>

¹³ <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#digitalaudio>

¹⁴ https://files.nc.gov/dncr-archives/documents/files/file_formats_transfer_requirements.pdf

¹⁵ <https://www.lib.ncsu.edu/faq/what-file-formats-are-accepted-and-supported-spr-content>

¹⁶ <https://www.library.northwestern.edu/about/administration/policies/file-format-recommendations.html>

¹⁷ <https://purr.purdue.edu/legal/file-format-recommendations>

¹⁸ <https://www.sfu.ca/content/dam/sfu/archives/DigitalPreservation/FormatPolicyRegistry.pdf>

¹⁹ <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>

²⁰ https://opensiuc.lib.siu.edu/format_support.pdf

- ²¹ <https://www.informationstrategy.tas.gov.au/Records-Management-Principles/Document%20Library%20%20Tools/Guideline%2019%20Digital%20Preservation%20formats.pdf>
- ²² https://library.tamu.edu/services/scholarly_communication/deposit/policies/file-formats.html
- ²³ <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>
- ²⁴ https://www.lib.washington.edu/preservation/preservation_services/digitization-and-digital-preservation/preferred-file-formats

Poděkování

Tento článek byl připraven v souvislosti s plněním dotačního programu MK ČR, konkrétně NAKI II č. DG18P02OVV032 s názvem „Nový fonograf: naslouchejme zvuku historie. Vytvoření postupů a nástrojů pro evidenci, digitalizaci, zpřístupnění a dlouhodobou ochranu zvukových záznamů na historických nosičích v paměťových institucích“.

Použitá literatura:

BRYLAWSKI, Sam, Maya LERMAN, Robin PIKE a Kathlin SMITH. *Guide to Audio Preservation, Council on Library and Information Resources* [online]. 2015. [cit. 2020-09-17] ISBN 978-1-932326-50-5. Dostupné z: <http://www.clir.org/pubs/reports/pub164>.

Cíl. *Nový Fonograf* [online]. [cit. 2020-09-17]. Dostupné z: <https://novyfonograf.cz/o-projektu/cil>.

Collection ID Guide: Grooved Cylinder. *Preservation Self-Assessment Program* [online]. University of Illinois at Urbana-Champaign [cit. 2020-08-13]. Dostupné z: <https://psap.library.illinois.edu/collection-id-guide/cylinder>.

Collection ID Guide: Phonograph Record. *Preservation Self-Assessment Program* [online]. University of Illinois at Urbana-Champaign [cit. 2020-08-13]. Dostupné z: <https://psap.library.illinois.edu/collection-id-guide/phonodisc#shellacdisc>.

ČESKO, Ministerstvo kultury, Odbor umění, literatury a knihoven. Veřejné informační služby knihoven (VISK): podprogram č. 7: národní program ochrany a digitalizace dokumentů ohrožených degradací kyselého papíru – KRAMERIUS. In: *Veřejné informační služby knihoven* [online]. Praha: Národní knihovna ČR, 2019 [cit. 2020-08-13]. Dostupné z: <https://visk.nkp.cz/dokumenty/visk7/2020/VISK7-podm2020.docx>

GÖSSEL, Gabriel, Martin MEJZR, Michal STUDNIČNÝ a Filip ŠÍR. Digitalizace etiket standardních šelakových desek jako nástroj pro efektivní muzejní evidenci: výsledky průzkumu ve sbírce fonotéky Národního muzea – Českého muzea hudby. *ProInflow: časopis pro informační vědy* [online]. 2019, 11(2) [cit. 2020-09-17]. ISSN 1804-2406. Dostupné z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/2019-2-4>.

HNILÍČKA, Přemysl. Tradiční čeští loutkáři na deskách z roku 1945. *Nový Fonograf* [online]. 2020 [cit. 2020-09-17]. Dostupné z: <https://novyfonograf.cz/herci-v-drazkach/tradicni-cesti-loutkari-na-deskach-z-roku-1945/?fbclid=IwAR1cupW1Y45utGTXBgA97V0qAbB8CaQWE7rYsV9RUcgl6rfgNBSyg1wvHI8>.

IASA TECHNICAL COMMITTEE. *The Safeguarding of the Audiovisual Heritage: Ethics, Principles and Preservation Strategy* [online]. Version 4. International Association of Sound and Audiovisual Archive, 2017 [cit. 2020-06-06]. Dostupné z: <https://www.iasa-web.org/tc03/ethics-principles-preservation-strategy>.

KOORDINATIONSSTELLE FÜR DIE DAUERHAFT E ARCHIVIERUNG ELEKTRONISCHER UNTERLAGEN. *Kriterienkatalog zur Bewertung der Archivtauglichkeit von Dateiformaten KOST-CECO*, Version 6.0, Juli 2019. [online]. 2019 [cit. 2020-10-06]. Dostupné z: https://kost-ceco.ch/wiki/whelp/Bewertungsmatrix/Bewertungsmatrix_v6.pdf.

NATIONAL RECORDING PRESERVATION BOARD. *Capturing Analog Sound for Digital Preservation: Report of a Roundtable Discussion of Best Practices for Transferring Analog Discs and Tapes*. Library of Congress [online]. 2006 [cit. 2020-10-06]. ISBN 1-932326-25-1. Dostupné z: <https://www.loc.gov/static/programs/national-recording-preservation-plan/publications-and-reports/documents/pub137.pdf>.

Phonograph. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2020-08-13]. Dostupné z: <https://en.wikipedia.org/wiki/Phonograph>.

RUSSO, David. FLAC Format Preservation Assessment. DPC [online]. 2018 [cit. 2020-06-19]. Dostupné z: https://wiki.dpconline.org/images/f/fe/FLAC_Assessment_v1.0.pdf.

VETEŠKOVÁ, Michaela. Nadčasové teze i projev: Národní muzeum objevilo dosud neznámou nahrávku prezidenta Masaryka. *iRozhlas* [online]. 29. 6. 2015 [cit. 2020-09-17]. Dostupné z: https://www.irozhlas.cz/kultura/tomas-garrigue-masaryk-narodni-muzeum-nahravka_2003031706_ada.

WORLD WIDE WEB CONSORTIUM (W3C). ODRL Information Model 2.2: W3C Recommendation 15 February 2018. W3C [online]. World Wide Web Consortium, 2018. Dostupné z: <https://www.w3.org/TR/odrl-model/>.

BEŇAČKOVÁ, Miroslava, Pavlína KOČIŠOVÁ, Vojtěch KOPSKÝ a Natalie OSTRÁKOVÁ. Dlouhodobé uchování digitálních dat vzniklých digitalizací zvukových záznamů na fonografických válečcích a šelakových deskách. *Knihovna: knihovnická revue*. 2020, **31**(2), 45–61. ISSN 1801-3252.