
MOŽNOSTI ZÁPISU TECHNICKÝCH METADÁT DIGITÁLNYCH OBJEKTOV A VYTVORENIE VLASTNEJ METADÁTOVEJ SCHÉMY PRE ELEKTRONICKÉ PUBLIKÁCIE

Mgr. Filip Pavčík, PhD.; filip.pavcik@nkp.cz; (Odbor digitálnych fondov, Národná knižnica Českej republiky – NK ČR)

Technické metadáta slúžia k zaznamenávaniu technických vlastností digitálnych objektov a umožňujú spolu s ďalšími typmi metadát uchovávať, ochraňovať a spravovať archivované objekty. Sú potrebné k monitorovaniu stavu archivovaných digitálnych objektov, k plánovaniu a vykonávaniu ochranných opatrení (migrácia) a pre správnu prezentáciu objektov používateľom digitálneho obsahu. Pre zaznamenávanie technických metadát existujú štandardizované metadátové schémy a sú dostupné nástroje, ktoré dokážu technické vlastnosti extrahovať a výstup formátovať práve do týchto štandardizovaných schém. V rámci Národní digitální knihovny sa technické vlastnosti digitálnych objektov zaznamenávajú už od začiatku projektu a sú súčasťou archivačných balíčkov. Článok sa zameriava na možnosti zápisu technických metadát v digitalizačnej praxi pamäťových inštitúcií vo svete a v druhej časti článku popisuje vytvorenie vlastnej metadátovej schémy nazvanej ndktech, ktorá vznikla pre potreby dlhodobej ochrany elektronických publikácií.

<http://doi.org/10.52036/1335793X.2022.3-4.34-41>

Medzi najdiskutovanejšie problémy dlhodobej ochrany digitálnych dokumentov patrí problematika zaznamenávania rôznych typov metadát. Pojem „metadáta“ je v odborných publikáciách už tradične definovaný ako „dáta o iných dátach“. A hoci sa medzi informačnými vedcami už dlhšiu dobu diskutuje o neužitočnosti používania tohto vágneho tvrdenia (Pomerantz 2015, s. 19), tak aj takáto nejednoznačná definícia aspoň naznačuje, že metadáta môžu mať značne široké využitie v rôznych odboroch a profesiách (Cubr 2017, s. 56). V knižniciach a pamäťových inštitúciách sa pri obsahu pojmu metadát kladie dôraz na popis zdrojov (digitálnych objektov) a na ich príbuznosť s katalogizačnými záznamami. Oproti katalogizačným záznamom sú ale metadáta vysoko štruktúrované, vďaka čomu ich možno využívať k vyhľadávaniu a správe digitálnych objektov (Hutař 2012, s. 74).

Vyššie popísaný typ metadát sa označuje aj ako popisné (alebo bibliografické) metadáta a možno ich radíť medzi najpoužívanejšie v každodennej praxi českých a slovenských knižníc. Popisné metadáta vychádzajú zo zaužívaných medzinárodných štandardov (MARC, Dublin Core, MODS a ďalšie) a sú určené na popis textových, zvukových, audiovizuálnych alebo iných dokumentov, prípadne elektronických publikácií. V súčas-

nej odbornej literatúre sa však rozlišujú aj ďalšie typy metadát, pričom najčastejšie prijímané rozlíšenie definuje tri skupiny: popisné, administratívne a štrukturálne metadáta (a v rámci nich sa rozlišujú ďalšie podskupiny). Témou tohto článku je samostatný typ metadát – technické metadáta, ktoré sa zvyčajne zaraďujú pod zastrešujúcu hlavičku administratívnych metadát, spolu s právnymi a ochrannými metadátami (Pomerantz 2015, s. 94-95; Bratková 2012, s. 15). Cieľom článku je poukázať na možnosti zapisovania technických metadát v digitalizačnej praxi vo svete a v druhej časti článku sa zameriame na popis vytvorenia vlastnej schémy na zapisovanie technických metadát, ktorú využívame pri digitalizácii elektronických publikácií v NDK.

ZÁKLADNÉ DEFINÍCIE

Z hľadiska využívania medzinárodných štandardov sa pri technických metadátach dlhodobo pociťoval akútny nedostatok medzinárodných štandardizovaných schém. Situácia sa začala meniť až po roku 2010, avšak ani v súčasnosti výskum neukazuje ich výraznú rozšírenosť a využiteľnosť, ako je to napr. pri popisných metadátach (Otto 2015, s. 5). Technické metadáta pritom patria k najdôležitejším súčasťam popisu digitálnych objektov, najmä čo sa týka ich ochrany v dlhodobom

horizonte. Medzi najpoužívanejšie schémy určené pre technické metadáta patrí medzinárodný štandard MIX určený pre obrazové dokumenty, ktorý sa využíva aj v digitalizačnej praxi Národnej digitálnej knihovny (ďalej NDK). Technické metadáta sú spravidla zaznamenávané automaticky pomocou metadátových extraktorov (JHOVE, FITS, DROID, Kakadu) a popisujú informácie o technických vlastnostiach digitálnych objektov, akými sú: typ a veľkosť objektu, súborový formát, kontrolné súčty, informácie o digitalizácii a pod. Cieľom extrakcie týchto informácií je ich uchovávanie tak, aby v neskoršom období životného cyklu daného objektu bolo možné vykonávať ochranné opatrenia, t. j. formátovú migráciu, emuláciu, normalizáciu a pod. (Hutař 2012, s. 80; Cubr 2010, s. 72 – 73, 81). Podľa nedávneho výskumu britských autorov (Gerrard, Mooney a Thompson 2017) môžu byť technické metadáta používané aj k zaznamenávaniu trendov v oblasti digitalizácie obrazov v priebehu rokov a potenciálne tak môžu slúžiť k predikciám a plánovaniu budúcej situácie. Pre budúcich bádateľov a oblastí Big Data tak môžu byť informácie o digitálnych objektoch dôležitejšie ako samotné archivované digitálne objekty (v prípade, ak budú použiteľné a analyzované v rámci zodpovedajúcich nástrojov).

Za technická metadáta bývajú niekedy považované aj informácie o hardvérovom a softvérovom prostredí, ktoré sú nevyhnutné k zobrazeniu či spracovaniu digitalizovaného objektu. V odbornej literatúre však bývajú zaradované do kategórie archivačných metadát, príp. metadát o pôvode objektu (Bratková 2012, s. 15), preto sa im v predkladanom článku nebudeme venovať.

ZAZNAMENÁVANIE TECHNICKÝCH METADÁT V PRAXI NDK

Ako už bolo spomenuté, na zaznamenávanie (technických) metadát slúžia špeciálne štandardizované metadátové schémy. Ich systematické využívanie umožňuje zaistiť interoperabilitu medzi systémami (výmenu dát), konverziu dát do iných formátov a tiež efektívnu tvorbu metadát (Ostráková a Šír 2017, s. 12). Tradičná definícia podľa NISO (National Information Standards Organization) definuje štandardizovanú metadátovú schému ako „sadu metadátových elementov určenú na konkrétne účely, ako napr. k popisu konkrétneho typu informačného zdroja“ (cit. podľa Hutař 2012, s. 74). Schéma teda určuje konkrétny súbor elementov, ktoré na zaznamenávanie možno využiť a prípadne určuje alebo obmedzuje hodnoty (t. j. obsah) týchto elementov. Zapisovanie elementov je najčastejšie vyjadrované vo formáte XML, ktorý v súčasnosti patrí medzi najpoužívanejšie značkovacie jazyky (Kosek 2014).

Využívanie uznávaných štandardizovaných schém

(skrátene štandardov) má množstvo výhod: poskytuje medzinárodnú zrozumiteľnosť, a tým aj zjednodušuje migráciu medzi rôznymi metadátovými formátmi. Navyše, štandardizované schémy poskytujú pre používateľov vodidlo a pomáhajú definovať, ktoré informácie sú dôležité a mali by byť zaznamenané. Taktiež treba vyzdvihnúť ich kompatibilitu s rôznymi charakterizačnými nástrojmi na extrakciu metadát, ktoré dokážu extrahovať metadáta v konkrétnych metadátových schémach. V neposlednom rade je ich výhodou aj skutočnosť, že sú spravované spoľahlivou medzinárodne uznávanou inštitúciou (ako je napr. Kongresová knižnica USA – ďalej LOC).

Digitalizačná prax v rámci NDK využíva na tvorbu metadátových záznamov vlastný aplikačný metadátový profil, ktorý sa v koncepcnej rovine začal pripravovať od roku 2007. Pri tvorbe metadátových záznamov sa od tohto obdobia začalo pracovať s uznávanými medzinárodnými štandardmi (MODS, PREMIS, METS a ďalšie) a vzorom pri vytváraní vlastného metadátového profilu boli inštitúcie s dlhoročnou praxou v digitalizácii, ako sú LOC, Austrálska národná knižnica alebo Fínska, či Nórska národná knižnica (Hutař 2012, s. 186 – 187). V praxi NDK sa tak od roku 2012 v čo najväčšej miere využívajú nasledujúce štandardizované metadátové schémy dôležité pre popis technických metadát: MIX pre rastrové obrazy (NISO 2017), AES57 pre zvukové súbory (AES 2011), ALTO pre popis usporiadania a obsahu textových dokumentov (ALTO 2014), documentMD pre elektronické publikácie (Chou a Goethals 2009) a PREMIS, ktorý ukladá podstatné informácie k archivácii a ochrane digitálnych dokumentov (PREMIS 2015). V nasledujúcich riadkoch každú z vymenovaných schém aspoň v stručnosti popíšeme a uvedieme, akým spôsobom sa využíva v praxi NDK.

PREMIS

Jeden z najznámejších a najrozšírenejších štandardov je už spomínaný štandard PREMIS (PREservation Metadata: Implementation Strategies). Je určený na použitie pre rôzne typy digitálnych objektov (v praxi NDK sa využíva pre textové, zvukové i elektronické dokumenty) a umožňuje popísať nielen technické vlastnosti objektov, hardwarové a softwarové prostredie, ale aj informácie o pôvode, záznamy zmien na archivovaných objektoch (konverzie, zlúčenia objektov) či informácie o nástrojoch, ktoré boli pri digitalizácii použité. Podobne ako iné štandardy, aj PREMIS definuje elementy, hoci tieto elementy nie sú reprezentáciami digitálnych objektov (ako je to napr. pri bibliografických metadátach). Preto sa v rámci štandardu PREMIS používa termín sémantické jednotky, ktoré umožňujú zapísať veľkosť súborov v bytoch, informácie o formáte súboru a jeho verzii, o aplikáciách, z ktorých boli vytvorené a pod. (Caplan 2018, s. 2).

Štandard PREMIS sa využíva najmä na zaznamenávanie administratívnych metadát a technické metadáta je možné zapísať len do určitej obmedzenej miery. Jeho nespornou výhodou však je, že pomocou elementu (alebo presnejšie povedané sémantickej jednotky) `objectCharacteristicsExtension` umožňuje vložiť externú metadátovú schému, ktorá môže slúžiť na podrobnejší popis digitálnych objektov alebo k popisu vzniku či zmeny digitálnych objektov (typicky sa vkladá napr. schéma MIX alebo výstupy z charakterizačných nástrojov). Externé doplňujúce schémy sú ale závislé na konkrétnom type dát a je výrazný rozdiel v tom, či ide o zvukové, textové alebo elektronické dokumenty. PREMIS je tak spravidla nutné rozšíriť o externú schému vhodnú pre konkrétny druh digitálnych objektov (Caplan 2018, s. 6).

V rámci aplikačného metadátového profilu NDK sa v súčasnosti nepopisuje hardwarové a softwarové prostredie, ktoré sa používa pri práci so súbormi, výhľadovo však je záujem evidovať aj tieto informácie. V budúcnosti sa so štandardom PREMIS počíta aj pri digitalizácii audiovizuálnych dokumentov. Pre zápis špecifických technických vlastností daných digitálnych objektov sa taktiež počíta s vložením externej metadátovej schémy.

MIX

Najpoužívanejším štandardom na popis technických metadát je štandard MIX (NISO 2017). Využíva sa na popis obrazových dokumentov a jeho prednosťou je, že nie je viazaný na konkrétny súborový formát. Štandard MIX taktiež definuje množinu elementov, ktoré sú potrebné na popis digitálnych obrazov a veľkú väčšinu týchto elementov využíva aj aplikačný metadátový profil NDK. Pomocou štandardu MIX je takto možné popísať nielen vlastnosti digitálnych obrazov, ale i okolnosti vzniku digitálnych súborov, ako typ skenovacieho zariadenia, výrobcu skeneru, jeho model a sériové číslo, skenovací software a i. (Ostráková a Šír 2017, s. 15). Štandard MIX je taktiež implementovaný v charakterizačných nástrojoch ako JHOVE alebo FITS, ktoré dokážu vygenerovať metadáta priamo v požadovanom formáte. Digitalizačná prax NDK využíva na ukladanie metadát vo formáte MIX rozšírenie schémy METS, ktorá umožňuje ukladať aj štrukturálne, autorско-právne a iné typy metadát. Príkladom dobrej praxe je však aj ich ukladanie do PREMIS-u ako externé doplňujúce schémy, ako s tým pracuje napr. Národný archív Švédska či iné inštitúcie (Riksarkivet 2012).

AES57

Pre digitálne zvukové dokumenty sa od roku 2011 využíva štandard AES57 (AES 2011). Ten vznikol, aby nahradil starší a jednoduchší štandard s názvom audioMD, vytvorený kongresovou knižnicou USA v ro-

koch 2000 – 2003. Štandard AES57 bol vytvorený Spoločnosťou zvukových inžinierov – Audio Engineering Society – a jeho používanie odporúča aj medzinárodná inštitúcia ARSC – Association for Recorded Sound Collections (Brylawski et al. 2015). Tá vo svojej príručke odporúča používať okrem AES57 aj štandard PBCore. PBCore sa síce primárne využíva na zaznamenávanie popisných metadát, no umožňuje zapisovať aj technické vlastnosti zvukových dokumentov (Ostráková a Šír 2017, s. 14).

V minulosti sa počas prípravy vlastného štandardu pre zvukové dokumenty v rámci NDK, zameraného na gramofónové platne, uvažovalo aj nad uplatnením jednoduchšej schémy audioMD, no od tohto zámeru sa dodatočne upustilo, keďže schéma nevyhovovala potrebám NDK. Schéma AES57 je implementovaná aj v charakterizačnom nástroji FITS, ktorý tak vie generovať technické metadáta v zhode so štandardom. Jeho nevýhodou ale je, že na rozdiel od štandardu MIX nedokáže popísať okolnosti vzniku digitálnych zvukových dokumentov (Ostráková a Šír 2017, s. 16).

ALTO XML

Štandard ALTO XML sa síce primárne nevyužíva na zaznamenávanie technických metadát, no umožňuje ukladať informácie potrebné k popisu usporiadania a obsahu textového dokumentu. Najčastejšie sa využíva ako externá schéma v rámci rozšírenia štandardu METS (v sekcii administratívnych metadát), ale je možné ho používať aj samostatne (ALTO 2016). V aplikačnom metadátovom profile NDK slúži štandard ALTO XML na popis vzhľadu a obsahu textového dokumentu a – podobne ako iné štandardy – je naviazaný na schému METS (Hutař 2012, s. 209). ALTO XML je úplne závislý na procese rozoznávania znakov pomocou aplikácie OCR – Optical character recognition. Aplikácia OCR spracovuje text z digitalizovaného dokumentu a tento čistý text následne prevedie do formátu ALTO XML. Efektivita OCR býva v českej digitalizačnej praxi značne vysoká a úspešnosť rozoznávania textu z obrazu sa pohybuje minimálne na úrovni okolo 95 % (Cubr 2017, s. 55). V súčasnosti je aktuálna verzia ALTO 4.3. Digitalizačná prax NDK však pripúšťa aj používanie verzií 2.0 a 3.0.

documentMD

DocumentMD je metadátový štandard, pomocou ktorého sa v NDK popisujú technické vlastnosti elektronických publikácií. Štandard vyvíja a spravuje knižnica Harvardskej univerzity a bol vytvorený špeciálne pre elektronické knihy, PDF dokumenty a dokumenty z aplikácií Microsoft (Dappert, Guenther a Peyrard 2016, s. 191). Štandard bol pôvodne navrhnutý na kombináciu so štandardom PREMIS a taktiež umožňuje vložiť aj externé metadátové schémy pre zápis

ďalších vlastností, čo je jeho nespornou prednosťou (Chou a Goethals 2009).

CHARAKTERIZAČNÉ NÁSTROJE

Vlastné zdokumentované a rozšírené metadátové schémy je možné nájsť aj vo výstupoch charakterizačných nástrojov, akými sú Jpylyzer, JHOVE, FITS, VeraPDF a ďalšie. Aj tieto je možné teoreticky použiť na uloženie technických vlastností digitálnych objektov a priebežne sa o tom v projekte NDK uvažuje. Najmä v tých prípadoch, ak neexistuje iný vhodný metadátový štandard, prípadne neobsahuje elementy pre podrobný popis všetkých vlastností, alebo ak by mapovanie do vybraného štandardu mohlo byť príliš náročné s vysokým rizikom vzniku chýb. Nejde však o preferované riešenie, ale skôr len o doplnok – napr. zahrnutie výstupu z nástrojov do rozšírenia schémy PREMIS. Nástroje ako FITS či JHOVE totiž dokážu generovať metadáta aj v medzinárodných štandardizovaných schémach, čo je ich výhodou, keďže inštitúcia sa tak nemusí uspokojiť s predvoleným formátom, ktorý vygeneruje daný nástroj. Problematickou otázkou ale je, ako sú formáty v rámci týchto nástrojov spravované, keďže – na rozdiel od štandardov spravovaných LOC – tieto nástroje podliehajú vo výrazne vyššej miere zmenám a častým aktualizáciám. Pre bežného používateľa sú navyše výstupy charakterizačných nástrojov ťažšie čitateľné a pochopiteľné (napr. veraPDF). V prípade využívania týchto nástrojov by kurátori obsahu NDK museli taktiež uchovávať schémy daných nástrojov a sledovať, či ich nové verzie nezmenili aj svoje schémy, a teda, či sa nezmenil aj metadátový popis pri dátach uložených k archivácii.

V každom prípade, v praxi NDK platí pravidlo, že ak existujú vhodné metadátové štandardy, tak sú tieto preferovanými formami zápisu technických metadát a výstupy z nástrojov sú akceptované len v krajnom prípade. Pri dátach z týchto nástrojov by sa počítalo s ich neskoršou migráciou do vhodnej štandardizovanej schémy. Jeden z príkladov dobrej praxe zahraničných pamäťových inštitúcií: špecifikácia E-ARK pre AIP používa metadátovú schému PREMIS a do rozšírenej schémy pripúšťa aj uloženie výstupov z nástroja JHOVE (DILCIS 2021).

VLASTNÁ METADÁTOVÁ SCHÉMA NDKTECH A JEJ POUŽÍVANIE V RÁMCI PROJEKTU NDK

Všeobecné odporúčania k digitalizácii odporúčajú používať medzinárodne uznávané štandardizované schémy. Na druhej strane ale medzinárodné schémy nie vždy zodpovedajú potrebám pamäťových inštitúcií. V takom prípade si inštitúcie navrhujú vlastné lokálne metadátové štandardy. Využívanie vlastných štandardov nie je v praxi pamäťových inštitúcií ničím nezvyčajným, keďže lokálne schémy využívajú mnohé

európske knižničné inštitúcie. Ich využívaním je známa napr. Nemecká národná knižnica (DNB – Deutsche National Bibliothek), ktorá aj pri popisných metadátach využíva upravený Dublin Core pod názvom DNB Casual a na technické metadáta využíva lokálny štandard LMER (Long-Term Preservation Metadata for Electronic Resources). Spolu so štandardom LMER ale používa ako doplnok aj schémy MIX alebo textMD. Zámerom DNB pri vytváraní lokálneho štandardu bolo nájsť „hlavné typy metadát, ktoré by mohli reprezentovať všetky typy súborov“ (Steinke 2005, s. 5). Na elektronické univerzitné publikácie (záverečné a vedecké práce) používa DNB taktiež vlastnú metadátovú schému XMetaDissPlus, ktorú vyvinuli v spolupráci s Library Service Centre Baden-Württemberg (Netzpublikationen 2020). Podobne vlastné schémy využívajú aj ďalšie knižnice, napr. Nórska či Fínska národná knižnica a i.

Aj v digitalizačnej praxi NDK bolo nutné doplniť existujúce medzinárodné štandardy vlastnými lokálne vytvorenými, pretože pre niektoré údaje vhodné na zaznamenanie chýbali zodpovedajúce elementy. Pre dlhodobú ochranu elektronických publikácií tak bola v rámci NDK vytvorená schéma ndktech, ktorá sa ukladá ako rozšírenie do štandardu PREMIS a v rámci neho následne do schémy documentMD. Týmto postupom sme sa snažili o zápis všetkých (technických) vlastností, ktoré je možné automatizovane zistiť, a zároveň sa v čo najväčšej miere podarilo využiť aj medzinárodné rozšírené schémy (PREMIS a documentMD). Lokálna schéma ndktech sa v súčasnosti využíva len pre formáty PDF a EPUB a zaznamenávajú sa do nej údaje, ktoré vychádzajú z charakterizačných nástrojov (VeraPDF, Epubcheck).

Keďže štandardy PREMIS a documentMD neumožňovali zapísať všetky dôležité údaje v rámci svojej sady elementov, došlo v rámci projektu NDK k vytvoreniu vlastnej lokálnej schémy. Schéma ndktech síce vznikla pre konkrétny typ dát (elektronické publikácie) a formátov (PDF a Epub), no z praktických príčin sa naďalej spravuje a rozvíja ako – pokiaľ možno – čo najviac formátovo a dátovo nezávislá. Zámerom totiž je rozširovať ju aj o elementy dôležité pre popis technických vlastností aj iných formátov a iných typov dát.

SIGNIFIKANTNÉ VLASTNOSTI ELEKTRONICKÝCH PUBLIKÁCIÍ V PRAXI NDK

Technické metadáta sa v projekte NDK mnohokrát označujú aj za súčasť tzv. signifikantných (alebo významných) vlastností digitálnych objektov. Pojem signifikantných vlastností je mimoriadne komplexný a zložitý a v súčasnosti patrí k najdiskutovanejším konceptom dlhodoobej ochrany digitálnych dokumentov. Jeho najčastejšie citovaná definícia uvádza, že ide o charakteristiky objektu, ktoré je potrebné v čo najväčšej miere zachovať počas

archivačných ochranných opatrení (Konečný 2018, s. 1 – 2). Určovanie signifikantných vlastností spočíva nielen v analýze archivovaných dát (t. j. vlastností digitálneho objektu), ale zameriava sa aj na producentov a zberateľov dát a taktiež na konečných používateľov (tzv. stakeholderov) a potreby danej používateľskej komunity (Beňačková et al. 2020, s. 26).

V digitalizačnej praxi NDK sa za signifikantné vlastnosti považujú všetky elementy, ktoré sú definované v aplikačnom metadátovej profile, konkrétne v tzv. Definíciách metadátovej formátov (ďalej DMF). Keďže

medzinárodný štandard PREMIS na popis signifikantných vlastností pre potreby NDK nepostačoval, tak sa v rámci PREMIS-u pri elektronických publikáciách využíva rozšírenie schémy v rámci kontajnerovej sémantickej jednotky `objectCharacteristicsExtension`, ktorá ukladá výstupy z ďalších schém alebo z charakterizačných nástrojov (VeraPDF, FITS a ďalšie). Pomocou tohto rozšírenia tak bolo možné vložiť vlastnú metadátovej schému, ako to bolo použité v projekte NDK pri tvorbe tzv. DMF pro elektronické publikace – Monografie, verze 2.3 (Beňačková, Ostráková a Mally 2019).

Element	Atribút	Popis
<code><objectCharacteristicsExtension></code>		kontajnerový element, pomocou ktorého sa vloží externá metadátovej schéma
<code><ndktech:ndktech></code>		koreňový element
	<code>xmlns:ndktech</code>	odkaz na webovú stránku metadátovej schémy: "https://standards.ndk.cz/ndk/standards-digitalizace/ndktech"
	<code>xsi:schemaLocation</code>	odkaz na aktuálnu verziu metadátovej schémy: "https://standards.ndk.cz/ndk/standards-digitalizace/ndktech-v1-1.xsd"
<code>x<ndktech:filters></code>		zoznam použitých filtrov v PDF
<code>xx<ndktech:filter></code>		použité filtre (napr. <i>FlateDecode</i> alebo <i>JPXDecode</i>); element sa opakuje podľa počtu použitých filtrov
<code>x<ndktech:profiles></code>		zoznam použitých profilov v PDF
<code>xx<ndktech:profile></code>		profil (napr. <i>Linearized PDF</i>)
<code>x<ndktech:colorspaces></code>		obsahuje zoznam použitých farebných priestorov v PDF
<code>xx<ndktech:colorspace></code>		názov farebného priestoru; element sa opakuje podľa počtu farebných priestorov
<code>x<ndktech:iccprofile></code>		
<code>xx<ndktech:iccprofilename></code>		meno farebného priestoru (napr. <i>RGB</i> , <i>Adobe RGB</i> , <i>CIE</i>)
<code>xx<ndktech:iccprofileversion></code>		verzia profilu (napr. <i>sRGB IEC61966-2.1</i>)
<code>x<ndktech:imagesCount></code>		počet obrázkov v dokumente
<code>x<ndktech:indirectObjectsNumber></code>		celkový počet objektov v dokumente

Tabuľka č. 1 Ukážka metadátovej schémy `ndktech`, ktorá sa využíva pri digitalizácii elektronických publikácií.
Tabuľka je prevzatá z tzv. DMF pro elektronické publikace – Monografie, verze 2.3.

STANOVENIE SIGNIFIKANTNÝCH VLASTNOSTÍ ELEKTRONICKÝCH PUBLIKACÍ PRI NÁVRHU VLASTNEJ LOKÁLNEJ SCHÉMY NDKTECH

Pre popis elektronických publikácií sme pri vytváraní lokálnej schémy `ndktech` definovali signifikantné vlastnosti, ktoré nebolo možné zaznamenať medzinárodnými metadátovej schémami (napr. pomocou

PREMIS-u), ale ich zapísanie bolo nutné pre dlhodobú ochranu elektronických publikácií. V nižšie uvedenej tabuľke sú vymenované vlastnosti digitálnych objektov (pre formát PDF), ktoré si náš výskum určil ako signifikantné. Keďže výber signifikantných vlastností môže byť značne široký a rozsiahly, náš výber bol zúžený iba na dáta-centrické vlastnosti a iba na úroveň

Signifikantná vlastnosť	Vysvetlivka	Použité charakterizačné nástroje na plnenie	Hodnoty z charakterizačných nástrojov	Prítomnosť v tzv. DMF
Súborový formát	údaje o formáte súboru a o jeho verziách	identifikácia formátu z nástrojov: VeraPDF, DROID	slovná hodnota, stačí uviesť príponu súboru (napr. pdf)	element <format> v PREMIS Object
Veľkosť súboru	údaj o veľkosti súboru v bytoch	VeraPDF alebo JHOVE	číselný údaj	<size> v PREMIS Object
Farebnosť	obsahuje zoznam farebných priestorov	VeraPDF	slovná hodnota - názov použitých farebných priestorov v PDF	<colorspace> v schéme ndktech
Kódovanie	technické údaje o súbore	VeraPDF	číselný údaj	<objectCharacteristics> v PREMIS Object
Počet strán	údaj o počte strán	VeraPDF	číselný údaj - preberá sa z elementu <pageOrderNumber>	<PageCount> v documentMD
Grafické vyobrazenia	obsahuje počet obrázkov v dokumente PDF	VeraPDF	preberá sa z elementu <xobject type="image">	<imagesCount> v schéme ndktech
Vložené tabuľky	údaj o počte tabuliek	je možné zistiť len pri katalogizácii	číselný údaj	<TableCount> v schéme ndktech
Filtre	údaj o použitých filtroch	VeraPDF	hodnoty: FlateDecode a JPXDecode	<filters> v schéme ndktech
Profily	údaje o názve a verziách profilu	VeraPDF	hodnoty: RGB, Adobe RGB, CIE	<profiles> v schéme ndktech

Tabuľka č. 2 Stanovenie významných vlastností elektronických publikácií

objektu. Stanovenie signifikantných vlastností pre ďalšie úrovne, pre iné typy formátov ako PDF alebo na základe potrieb rôznych pamäťových inštitúcií nebolo predmetom nižšie uvedenej analýzy.

Pri každej signifikantnej vlastnosti zobrazenej v tabuľke 2 je uvedené, akým spôsobom boli hodnoty získané (napr. ktorým charakterizačným nástrojom) a či bolo možné pre zápis danej vlastnosti využiť medzinárodnú schému, alebo bolo nutné použiť vlastnú schému – ndktech (tieto vlastnosti sú v tabuľke zvýraznené tučným písmom). Taktiež je v tabuľke uvedené, či sú dané vlastnosti vyžadované aj pri používaní súčasných štandardov pre elektronické publikácie – v spomínaných DMF pro elektronické publikace – Monografie, verze 2.3.

ZÁVER

Odborná literatúra sa zhoduje, že pri technických metadátach sa dlhodobo pociťujú nedostatky v rámci existujúcich štandardizovaných schém, ktoré nie vždy vyhovujú potrebám pamäťových inštitúcií. Vytváranie vlastných metadátových štandardov je tak bežnou praxou v knižničných inštitúciách v Európe a vo svete. Pre vytvorenie vlastnej metadátovej schémy sme v rámci projektu NDK najprv analyzovali výstupy z charakterizačných nástrojov (v danom prípade konkrétne výstupy z nástroja VeraPDF) a následne si určili signifikantné vlastnosti, ktoré je potrebné zaznamenávať kvôli dlhodobej ochrane elektronických publikácií. Dôležitou skutočnosťou pri vytváraní vlastnej schémy tu bola možnosť rozšírenia metadátového zápisu v štandarde PREMIS pomocou kontajnerovej sémantickkej jednotky objectCharacteristicsExtension. Dané rozšírenie umožnilo do medzinárodného štandardu PREMIS implementovať vlastnú schému ndktech a zachytiť tak signifikantné vlastnosti, ktoré nebolo možné zapísať existujúcimi štandardizovanými schémami.

Práve stanovenie signifikantných vlastností je dôležité pre uchovávanie digitálnych dokumentov v dlhodobej perspektíve, keďže digitálne objekty môžu byť v budúcnosti uchovávané v inej podobe, ako boli prijaté do dlhodobého úložiska v čase archivácie. Určenie signifikantných vlastností je teda vhodnou prípravou pre budúce migrácie a transformácie digitálnych dokumentov a pri ich stanovovaní bolo našim úmyslom zachytiť všetky vlastnosti, ktoré bolo možné zistiť z existujúcich charakterizačných nástrojov.

V uvedenom texte sme sa zaoberali schémou ndktech určenou pre elektronické publikácie ukladané vo formáte PDF. Zámerom do budúcnosti bude rozšíriť schému aj pre iné formáty, prípadne využívať aj ďalšie charakterizačné nástroje na zachytenie ďalších vlastností, ktoré by mohli byť potrebné k dlhodobej archivácii. Schéma ndktech sa vytvárala na Oddelení pro

Standards NK ČR od roku 2017 a odvtedy sa aktívne využíva pri digitalizácii v NDK.

Tento článok vznikol na základe inštitucionálnej podpory dlhodobého koncepčného rozvoja výskumnej organizácie Národní knihovna České republiky poskytovanej Ministerstvom kultúry ČR.

Zoznam použitých zdrojov a literatúry:

- AES Standard for audio metadata: Audio object structures for preservation and restoration, 2011 –. *Audio Engineering Society* [online]. New York: Audio Engineering Society [cit. 2022-10-10]. Dostupné na: <https://www.aes.org/publications/standards/se-arch.cfm?docID=84>
- ALTO Principles, 2016. *The Library of Congress* [online]. Washington, D. C.: *The Library of Congress* [cit. 2022-09-30]. Dostupné na: <https://www.loc.gov/standards/alto/description.html>
- ALTO: Technical Metadata for Layout and Text Objects 2014 –. *The Library of Congress* [online]. Washington, D.C.: The Library of Congress [cit. 2022-09-30]. Dostupné na: <https://www.loc.gov/standards/alto/>
- BEŇÁČKOVÁ, Miroslava, OSTRÁKOVÁ, Natalie a Richard MALLY, eds., 2019. *Definice metadatových formátů pro elektronické publikace. Monografie, dokument verze 2.3* [online]. Praha: Národní knihovna ČR [cit. 2022-10-10]. Dostupné na: https://standards.ndk.cz/ndk/standards-digitalizace/DMF_ebornmonografie_2.3.pdf
- BEŇÁČKOVÁ, Miroslava et al., 2020. Signifikantní vlastnosti: příspěvek ke kolektivnímu nevědomí. In: *ProInflow* [online]. Roč. 12, č.2, s. 21 – 44 [cit. 2022-10-10]. ISSN 1804–2406. Dostupné na: <https://doi.org/10.5817/ProIn2020-2-3>
- BRATKOVÁ, Eva. 2012. *Metadata a jejich schémata*. Praha: ÚISK, 60 s.
- BRYLAWSKI, Sam et al., 2015. *Guide to Audio Preservation* [online]. Council on Library and Information Resources [cit. 2022-10-10]. Dostupné na: <https://www.clir.org/wp-content/uploads/sites/6/pub164.pdf>
- CAPLAN, Priscilla, 2018. *PREMIS a jak mu porozumět*. Praha: Univerzita Karlova, 26 s.
- CHOU, Carol C. H. a Andrea GOETHALS, 2009. *Document Metadata: document technical metadata for digital preservation* [online]. Harvard Digital Library, Florida Digital Archive [cit. 2022-10-05]. Dostupné na: https://web.archive.org/web/20150907192814/http://claweb.fcla.edu/uploads/Lydia%20Motyka/FDA_documentation/documentMD.pdf
- CUBR, Ladislav, 2017. *Autenticita a digitální informace* [dizertačná práca]. Praha, 227 s. [cit. 2022-10-10]. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Dostupné na: <https://is.cuni.cz/webapps/zzp/detail/105596>
- CUBR, Ladislav, 2017 – 2018. *Analýza standardu Premis, věnující se signifikantním vlastnostem*. Praha: Národní knihovna ČR. Interní dokument Národní knihovny.
- CUBR, Ladislav, 2010. *Dlouhodobá ochrana digitálních dokumentů*. Praha: Národní knihovna České republiky, 154 s. ISBN 978-80-7050-588-5.

- DAPPERT, Angela, GUENTHER, Rebecca Squire a Sebastiane PEYRARD, eds., 2016. *Digital preservation metadata for practitioners: implementing PREMIS*. Cham: Springer. ISBN 978-3-319-43763-7.
- DILCIS Board, 2021. *E-ARK Archival Information Package (AIP). Specification for Archival Information Packages* [online]. Version: 2.0.4 [cit. 2022-10-10]. Dostupné na: <https://earkaip.dilcis.eu/pdf/eark-aip.pdf>
- GERRARD, David, MOONEY, James a Dave THOMPSON, 2017. Digital Preservation at Big Data Scales: proposing a step-change in preservation system architectures. In: *Library Hi Tech* [online]. Vol. 36, issue 3 [cit. 2022-10-10]. Dostupné na: https://ora.ox.ac.uk/objects/uuid:4e7801bd-f7cd-4a0f-b-287-9596e485e32d/download_file?file_format=pdf&safe_filename=DigitalPreservationAtBigDataScales_AuthorSubmittedVersion.pdf&type_of_work=Journal+article
- HUTAŘ, Jan, 2012. *Digitalizace, popis pomocí metadat a jejich formáty* [dizertačná práce]. Praha, 244 s. [cit. 2022-10-10]. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Dostupné na: <https://dspace.cuni.cz/bitstream/handle/20.500.11956/44181/140015545.pdf?sequence=1&isAllowed=y>
- KONEČNÝ, M., 2019. *Významné vlastnosti digitálních objektů a jejich určování*. Praha: Národní knihovna ČR. Interní studie pro Národní knihovnu.
- KOSEK, Jiří, 2014. *XML schémata* [online] [cit. 2022-09-30]. Dostupné na: <https://www.kosek.cz/xml/schema/index.html>
- LAZINGER, Susan S., 2001. *Digital preservation and metadata: history, theory, practice*. Englewood: Libraries Unlimited, s. 142 – 143. ISBN 1-56308-777-4.
- MILLER, Steven J., 2011. *Metadata for Digital Collections: a how to do it manual*. London: Facet, 343 s. ISBN 978-1-85604-771-5.
- Netzpublikationen: Lieferung von Metadaten im Format XMetaDissPlus (XMDP) an die Deutsche Nationalbibliothek. Metadattendokumentation für das Format XMetaDissPlus* [online], 2020. [Leipzig, Frankfurt am Main]: Deutsche Nationalbibliothek [cit. 2022-10-10]. Dostupné na: <https://d-nb.info/1217266461/34>
- NISO: Metadata for Images in XML Schema, 2017 –. *The Library of Congress* [online]. Washington, D.C.: The Library of Congress [cit. 2022-10-10]. Dostupné na: <https://www.loc.gov/standards/mix>
- OSTRÁKOVÁ, Natalie a Filip ŠÍR, 2017. Zvukové dokumenty ve fondech paměťových institucí v kontextu dlouhodobého uchování v ČR. Přípravná studie NK ČR k možnosti dlouhodobého uložení digitalizovaných dat. *Knihovna: knihovnická revue*. 28(1), 5 – 19. ISSN 1801-3252.
- OSTRÁKOVÁ, Natalie, Pavlína KOČIŠOVÁ a Miroslava BEŇAČKOVÁ. 2019. Vývoj standardu PREMIS a možnosti jeho dalšího využití ve standardech NDK. In: *ProInFlow* [online], roč. 11, č. 2, s. 72 – 85 [cit. 2022-10-10]. Dostupné na: <https://journals.phil.muni.cz/proinflow/article/view/2019-2-6/15505>
- OSTRÁKOVÁ, Natalie a Vojtěch KOPSKÝ, 2020. Posuzování souborových formátů z hlediska dlouhodobého uchování a návrh metodiky pro Národní knihovnu České republiky. In: *Knihovna: knihovnická revue*. Roč. 31, č. 2, s. 83 – 105. ISSN 1801-3252.
- OTTO, Jane Johnson, 2015. A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories. In: *Cataloging & Classification Quarterly* [online]. Rutgers University, vol. 48, Iss. 5, pp. 403 – 422 [cit. 2022-10-10]. Dostupné na: <https://doi.org/10.7282/T3WW7G11>
- POHANKA, Pavel, 2021. *Data-centric revoluce* [online] [cit. 2022-10-08]. Dostupné na: <https://pavelpohanka.cz/blog/data-centric-revoluce/>
- POMERANTZ, Jeffrey, 2015. *Metadata*. Cambridge. Massachusetts: The MIT Press, 252 s. ISBN 978-0-262-52851-1.
- PREMIS Editorial Committee, 2015. *PREMIS Data Dictionary for Preservation Metadata: version 3.0* [online]. Washington, D.C.: The Library of Congress [cit. 2022-10-05]. Dostupné na: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>
- Riksarkivet, 2012. *Premis_45_SNA_2015.xml* [online]. Stockholm: National Library of Sweden [cit. 2022-10-05]. Dostupné na: http://www.kb.se/namespace/mets/kbse_mets_profile_002.xml
- STEINKE, Tobias, ed., 2005. *LMER. Long-term preservation Metadata for Electronic Resources* [online]. Deutsche Bücherei Leipzig, Deutsche Bibliothek Frankfurt am Main, Deutsches Musikarchiv Berlin [cit. 2022-10-05]. Dostupné na: <https://d-nb.info/975115073/34>

■ Článek bol recenzovaný