

VÝBĚR VIDEO FORMÁTŮ V PAMĚŤOVÝCH INSTITUCÍCH

VIDEO FORMAT SELECTION IN MEMORY INSTITUTIONS

Vojtěch Kopský

Národní knihovna ČR

Abstrakt

Účel – Odborný článek pojednává o přístupu paměťových institucí k volbě formátu pro dlouhodobé a střednědobé uložení videa. Cílem je přiblížit čtenáři postup takového rozhodování a faktory, které ho ovlivňují.

Design/metodologie/přístup – Článek porovnává doporučení autoritních paměťových institucí, přístup menších institucí a institucí specializovaných na film.

Výsledky – Na základě studia formátových hodnocení, doporučení a formátových politik vybraných institucí jsme zjistili, že paměťové instituce obvykle ukládají digitalizáty analogového videa v nekomprimované nebo bezztrátově komprimované podobě. Nejčastěji používané formáty (kombinace kodeků a kontejnerů) jsou JPEG2000/MXF a FFV1/MKV, přičemž ten druhý se nám jeví jako výhodnější pro uložení výsledků digitalizace VHS kazet. Naopak obsah, který tyto instituce jako ztrátově komprimovaný získaly, případně v této podobě již vznikl, je v ní zatím ponecháván.

Originalita/hodnota – Článek přináší přehled informací o výběru vhodných formátů v paměťových institucích se zaměřením na archivaci videa.

Klíčová slova: dlouhodobé uložení kulturního dědictví, video, film, kodek, kontejner, komprese dat

Abstract

Purpose – The article deals with approaches of memory institutions to selection of the format for long term and midterm preservation of video. The goal is to present to reader the process of such decision making and factors that influence it.

Design/Methodology/Approach – The Article compares recommendation of authoritative memory institutions, approach of local institutions and institutions specializing in film.

Results – Following the analysis of format evaluations, recommendations and format policies of select institutions, we concluded that memory institutions preserve digitized analog video in uncompressed or lossless compressed form. Mostly used formats (combinations of codec and container) are JPEG2000/MXF and FFV1/MKV, while the later seems to us as more advantageous for preservation content from digitization of VHS cassettes. Contrary to that, content that those institutions received as lossy, or originated as such is preserved in native format.

Originality/Value – The Article presents summary of suitable format selection in memory institutions with respect to archiving of video.

Keywords: long term preservation of cultural heritage, video, movie, codec, container, compression of data

ÚVOD

Jedním z typů nosičů obsahu ve fondech českých knihoven jsou videokazety VHS, které, vzhledem k jejich omezené životnosti, bude třeba co nejdříve digitalizovat. Prvním krokem přípravy takového projektu je volba formátu vhodného pro dlouhodobé uložení obsahu. Cílem je zvolit formát, který i po dlouhé době budeme schopni přečíst, zobrazit jeho obsah. Postupujeme tak, že nejprve studujeme doporučení autoritních institucí, z nich vybereme formáty, které splňují naše technické požadavky a jsou kompatibilní s naším typem obsahu a o těch pak sháníme podrobnější informace související s jejich vhodností pro archivaci v hodnoceních autoritních institucí a dalších dostupných materiálech, například odborných publikacích, ale někdy i blozích, případně si přímo ověřujeme informace získané od autoritních institucí, například dostupnost specifikace či existenci potřebného software. Pro doplnění informací o rozšíření formátu v archivní komunitě studujeme formátové politiky lokálních institucí. A pro srovnání jsme v tomto konkrétním případě věnovali pozornost též doporučení institucí specializovaných na archivaci filmů.

Paměťové instituce, které se zabývají digitální ochranou kulturního dědictví, věnují velkou pozornost výběru správného formátu pro dlouhodobé uložení digitálního (vč. digitalizovaného) obsahu. Rovněž musí mít objektivní podklady pro rozhodnutí, které formáty přijímat od externích dodavatelů. Kromě firem, pracujících pro paměťové instituce na zakázku, (které by měly dodržovat specifikaci zadavatele), to ovšem mohou být též různé zájmové spolky, studentské projekty a podobně. Regionální instituce tak částečně vycházejí z doporučení autoritních institucí a částečně si vytvářejí vlastní formátové politiky, které zohledňují jejich potřeby.

Jako o autoritních institucích mluvíme o institucích, které vyvíjejí standardy a doporučení pro ostatní. V této práci jich bude zmíněno pět, které byly pro naše účely nejinformativnější. Chybí mezi nimi Britská knihovna, která má velmi podrobně zpracované formátové posudky (format assesment), ale bohužel se nevěnuje videoformátům. Jak uvidíme dále tyto instituce mají velmi rozdílné postupy, ale na druhou stranu mají poměrně značnou shodu na faktorech na jejichž základě formáty hodnotí.

V hodnocení formátů se digitální archiváři zaměřují především na jeho odolnost vůči zastarání, označovanou jako udržitelnost (sustainability). Sledují ovšem rovněž technické vlastnosti s ohledem na to, aby zvolený formát byl schopen nést význačné (signifikantní) vlastnosti uloženého obsahu.

V případě video formátů pak z těchto hledisek hodnotí některé instituce nezávisle kodek a kontejner, zatímco ostatní instituce hodnotí rovnou kombinace kodeku a kontejneru, jejichž využití zvažují. Pod pojmem **kodek** zde chápeme algoritmus pro kódování dat (nejčastěji video nebo audio) a jejich následné zpřístupnění, dekódování. Anglický název codec je zkratka z COder/DECoder. Kodeky mohou obsahovat i kompresi a šifrování. Příklady mohou být: H.264 (MPEG4), ProRes, JPEG2000 či FFV1. **Kontejner**, někdy též označovaný jako wrapper, je obálka, v níž jsou spolu s video obsahem uložena další data, především zvuková, případně titulky a různá metadata. Nejrozšířenější jsou AVI (Audio

Video Interleave), MOV (QuickTime), MXF (Material Exchange Format) a MKV (Matroska). Kontejner se vším svým obsahem odpovídá jednomu souboru, proto také přípona souboru je většinou odvozena od použitého kontejneru. Kodek ani kontejner nepovažujeme pro účely tohoto článku za formát, tím je až jejich kombinace. Nicméně, rizikové faktory a hodnotící postupy, které popisujeme níže, na ně lze aplikovat i samostatně.

HODNOTÍCÍ FAKTORY

Otázkou hodnocení formátů z hlediska vhodnosti pro archivaci se výzkumníci zabývají již druhou dekádu. Přehled hodnotících faktorů (v citované práci označovaných jako kritéria) přinesl ve své práci Adrian Brown z britských National Archives (Brown, 2008) a Judith Rogová a Caroline Wijkeová z Nizozemské Královské knihovny ve své studii navrhly kvantifikaci tohoto hodnocení pomocí rozdělení faktorů do dílčích složek (Rog a Wijk, 2008). Následující výčet vychází z přehledu zveřejněného Kongresovou Knihovnou (LOC, 2017). Případné odchylky v přístupu ostatních autoritních institucí jsou diskutovány u příslušných faktorů. Podrobnější srovnání těchto institucí nabízí naše starší práce (Ostráková a Kopský, 2020). V následujícím výčtu jsou nejdříve probrány faktory, na kterých panuje největší shoda, a současně ty, které se dají označit jako „faktory udržitelnosti“, a ke konci pak ty kontroverznější, a ne zcela jednoznačné z hlediska zařazení.

Otevřenost formátu (Disclosure) je pojem, který odkazuje na dostupnost dokumentace. Ta je nezbytná pro práci s formátem, vývoj software pro tvorbu, otevírání a validaci souborů. Výhodou je dokumentace standardizovaná ve spolupráci se standardizační institucí (ISO, IETF). Dokumentace často chybí u proprietárních formátů.

Rozšíření formátu (Adoption) je významným prediktorem jeho udržitelnosti. Vývojáři software mají zájem na implementaci rozšířených formátů do svých produktů a společně s uživateli mají zájem na jejich udržení. Je hodnoceno rozšíření obecné, rozšíření v paměťových institucích, a některé dokonce uvádí jako samostatné kritérium i výskyt formátu ve svém vlastním archivu.

Přehlednost (transparency) formátu je některými autory pojednávána z opačného pohledu jako „komplexita“ či „složitost“. Vyjadřuje přístupnost formátu pro analýzu pomocí základních nástrojů. U nejjednodušších formátů se očekává možnost „human readability“ tedy analýzy souboru prostřednictvím textového editoru. To samozřejmě není možné u kodeků využívajících kompresní algoritmy, k jejichž analýze jsou vždy potřeba speciální nástroje a považujeme je tedy za formáty s vysokou komplexitou. Kontejnery jsou většinou transparentní, byť jejich obsah není.

Vnější závislosti (External dependencies) jsou některými institucemi chápány jako závislost na externím obsahu (například používání odkazů na externí fonty v PDF), jinými pak jako závislost na specifickém hardwaru, operačním systému nebo programu. Ani zde pak není úplná shoda, například Kongresová knihovna uvádí, že kodek FFV1 externí závislosti nemá, zatímco knihovna Harvardovy univerzity (Goethals, 2016) označuje jeho závislost na jediném programu (FFmpeg) jako mírné riziko.

Autodokumentace (Self documentation) označuje metadata nezbytná pro správné zobrazení (rendering), například údaje o použitém barevném prostoru a údaje pro kontrolu neporušenosti souboru (fixity), což jsou většinou kontrolní součty (checksum neboli hash) – tuto schopnost může částečně nést kontejner, případně doprovodné (sidecar) soubory.

Mechanismy technické ochrany (Digital rights management) jsou považovány za rizikový faktor z hlediska dlouhodobé ochrany. Jde o možnost ukládat obsah v zakódované podobě, aby se tak zabránilo nelegálnímu kopírování obsahu. U video formátů, konkrétně u kontejnerů, je ovšem možnost jejich použití téměř univerzálně rozšířená, a z pohledu archivace je tedy nutné, aby byly zakázány ve specifikaci pro uložení. Tento faktor se tedy přinejmenším v případě videoformátů vztahuje spíše k formátové specifikaci než k formátu samotnému.

Patentové otázky (Impact of patents). Pokud je formát zatížen patenty, je jeho používání nákladnější. Zároveň ale existence patentů může blokovat vývoj formátu. U kodeků je navíc ten problém, že sestávají z řady komponent, algoritmů založených na různých principech, z nichž mnohé mají patentovou ochranu. Patenty k jednomu kodeku pak může držet více různých vlastníků. Licenci na jejich použití pak zajišťují tzv. patentové pooly, tedy konsorcia držitelů těchto patentů. Nejznámější je MPEG LA, který spravuje patenty H.26x až po H.264 (kodeky formátu kódování videa MPEG). Nicméně již pro následující H.265 (HEVC) se objevily další dva patentové pooly nárokuje si patenty s ním související. Jejich stupňující se finanční požadavky vedly nakonec k tomu, že firmy, které dosud tyto licence platily, (mezi které patří například Netflix nebo Google – YouTube), se sdružily do Aliance pro otevřený formát, která vyvinula pro účely streamování vlastní kodek AV1 (Ježek, 2018).

Implementace je faktor, který zahrnuje dostupnost software pro práci s daným formátem. Tím jsou myšleny programy na tvorbu souborů v daném formátu, jejich prohlížení, identifikaci a validaci, tedy testování, zda konkrétní soubor odpovídá specifikaci, kterou si archiv určil. Některé instituce řeší implementaci v rámci faktoru Rozšíření (Adoption), ale například KOST (níže) ho řeší samostatně. Přestože dostupnost software samozřejmě přispívá k udržitelnosti formátu, je zároveň nezbytná pro to, aby se s formátem vůbec dalo začít pracovat. Implementaci je tedy třeba ověřit ještě předtím, než začneme řešit ostatní faktory udržitelnosti.

Faktory kvality a funkcionality (Performance Factors). Kromě faktorů udržitelnosti jsou hodnoceny i faktory výkonnosti, související se schopností formátu nést význačné vlastnosti originálu, např. rozlišení, bitová hloubka, správa barev, možnost vícekanálového zvuku apod. Technické parametry ovšem jsou u řady institucí, včetně té naší, předmětem předvýběru, který předchází tomu, než je formát vůbec hodnocen z hlediska udržitelnosti. Požadavky na technické parametry se mohou lišit podle typu obsahu, a to někdy může některý formát diskvalifikovat, jako například formát FFV1/MKV, který není vhodný pro tzv. „broadcast material“ (viz. dále). Hodnotící kritéria ostatně nejsou univerzálně aplikovatelná, jak ukazuje specifický případ nativně digitálního obsahu, který je v současnosti nejlépe archivovat tak jak je, bez ohledu na to, jak je hodnocen z hlediska dlouhodobé udržitelnosti.

Kompresie je vnímána jako faktor udržitelnosti i výkonu. Některé instituce (KOST) ji dokonce hodnotí jako přínosnou pro udržitelnost díky velké úspoře úložného prostoru, ale převažující názor je, že ztrátová komprese je pro archivaci dlouhodobě neudržitelná. Obsah uložený do komprimovaného formátu totiž

sice může být vizuálně bezztrátový, ale do budoucna je třeba počítat s nutností jeho převodu do nového formátu a pak hrozí riziko hromadění ztrát informace (generation loss).

Z předpokladu, že všechny video formáty jsou z principu zastarávající, vychází George Blood (2011) z firmy **George Blood Audio and Video**, ve zprávě vypracované v rámci spolupráce s Kongresovou knihovnou. Společně se rozhodli pro řešení za použití bezztrátové komprese v kodeku JPEG2000, který zvolili v kombinaci s kontejnerem MXF za archivační master formát. K jeho očekávané dlouhodobé udržitelnosti pak odkazují pojmem „evergreen format“.

JPEG2000 je video kodek založený na rastrovém formátu JPEG2000, vyvinutém skupinou JPEG (Joint Photographic Experts Group). Pro kompresi využívá diskrétní vlnkovou transformaci a nepoužívá interframe kompresi, která je založena na sdílení informací mezi políčky filmu. Jeho ztrátově komprimovaná verze byla využívána pro streamování a filmovou distribuci. Byl standardizován jako ISO již v roce 2001 a v bezztrátové podobě ho používá řada paměťových institucí. Dalším bezztrátově komprimovaným kodekem, který se postupně rozšiřuje v archivní komunitě je FFV1, kterému se budeme podrobně věnovat na konci tohoto článku.

Na druhou stranu Blood ovšem počítá se střednědobou udržitelností ztrátově komprimovaných formátů, a proto doporučuje u souborů, které archivy v takových formátech obdrží, zatím obsah do bezztrátových formátů nepřevádět, ale pouze v případě nutnosti extrahovat data z fyzického nosiče (např. DV kazety) a převést do jiného kontejneru.

FORMÁTOVÁ HODNOCENÍ A DOPORUČENÍ – AUTORITNÍ INSTITUCE

Na základě výše popsaných faktorů hodnotí formáty autoritní instituce, které pak vyslovují doporučení ohledně vhodnosti hodnocených formátů pro archivaci. Hodnocení provádějí na větších či menších výběrech a může být slovní i bodové.

Koordinální centrum pro dlouhodobou archivaci elektronických dokumentů Švýcarského federálního archivu (Koordinationsstelle für die Dauerhafte Archivierung elektronischer Unterlagen KOST-CECO) nabízí na svých stránkách hodnocení několika vybraných nejpoužívanějších formátů (KOST, 2021a). Jedná se o relativně stručné a přehledné hodnocení slovní, uspořádané podle jednotlivých faktorů udržitelnosti doplněné bodovým hodnocením. JPEG2000 a FFV1 považují autoři za vhodné pro archivaci, nicméně upozorňují, že soubory s bezztrátovou kompresí jsou stále ještě velmi náročné na paměťový prostor. ProRes považují za nevhodný a u MPEG2 doporučují zatím počkat s převodem do nového formátu. Výsledky bodového hodnocení jsou též k dispozici v podobě srovnávací matice (KOST, 2021b).

Americký národní úřad pro archivaci a dokumentaci (National Archives and Records Administration – NARA) má systém hodnocení, kde pro jednotlivé faktory je určeno několik kritérií, formulovaných jako otázky, která jsou bodována různou vahou a rizikovost formátu je určena výsledným součtem. Výsledky jsou vyneseny v tzv. matici rizik (Risk matrix). (NARA, 2022a). Jedná se o formálně nejobektivnější přístup, ale v hodnocení bohužel chybí jak FFV1, tak JPEG2000 i většina ostatních kodeků a hodnoceny jsou především kontejnery. Jejich hodnocení v matici odpovídá i zařazení do kategorií v následující tabulce doporučení pro federální instituce (NARA, 2022b), kde ke kontejnerům jsou přiřazeny i kodeky. Preferovaný formát je zde ovšem uveden pouze pro kategorii „digital cinema“ a to DPX s nekomprimovaným obsahem a přijatelné (acceptable) formáty naopak pouze pro kategorii „digital video“, kde doporučuje formáty AVI a MOV s nekomprimovaným videem, WMV s kodekem VC-1, MPEG2 a MPEG4 (H.262 resp. H.264) a MXF s neztrátově komprimovaným JPEG2000.

Hodnotící matici používá též **Knihovna Harvardovy university (Harvard Library)**. V jednotlivých buňkách je zde použito hodnocení slovní, které je navíc zvýrazněno podbarvením příslušného políčka v tabulce, kdy přínosné vlastnosti jsou označeny zeleně, nevhodné červeně a neutrální žlutě. Na rozdíl od podobné matice, kterou Harvard vypracoval pro rastrové formáty, matice pro video formáty neobsahuje výsledné bodové hodnocení. Je nicméně užitečná jako přehled a srovnání shody formátů s požadavky pro archivaci. U JPEG2000 upozorňuje na vyšší náklady na správu, u FFV1 na závislost na jediném software (FFmpeg), u obou pak na omezené rozšíření (Goethals, 2016).

Iniciativa amerických federálních agentur pro zásady digitalizace (Federal Agencies Digital Guidelines Initiative – FADGI) přináší na svých stránkách srovnání pěti různých kodeků a pěti kontejnerů vhodných pro reformátování video obsahu z páskových médií, které se ve velkém rozsahu

zabývá technickými vlastnostmi (FADGI, 2014). U kodeků se jedná o dva nekomprimované (s bitovou hloubkou 8Bit a 10Bit a oba s barevným podvzorkováním 4:2:2), bezztrátově komprimované kodeky JPEG2000 a FFV1, a ztrátový kodek H.262, u kontejnerů pak o AVI, MOV, MKV, MXF a MPEG2. S výjimkou H.262 jsou všechny hodnoceny jako vhodné pro archivaci. V závěru studie jsou uvedeny vybrané příklady aplikací.

Americká Kongresová knihovna (Library of Congress, dále LOC) se zabývá archivací většiny typů obsahu. Má na svých stránkách rozsáhlou databázi formátových popisů (Format Descriptions), které sice nemají charakter hodnocení, ale informace o formátech jsou zde uspořádány podle jejich relevance k jednotlivým faktorům udržitelnosti, takže je to přehledný zdroj informací o vhodnosti formátu pro dlouhodobé uložení (LOC, 2021a).

Dále pak LOC na svých stránkách zveřejňuje formátová doporučení (Recommended Format Statement) pro účely dlouhodobé ochrany (LOC, 2021b). Rozlišuje zde na formáty doporučené (preferred) a přijatelné (acceptable) a navíc podle různých druhů obsahu. Pro filmy na filmovém pásu i digitální doporučuje kontejner DCP (Digital cinema package) s nezakódovaným (unencrypted) obsahem buď nekomprimovaným, nebo v bezztrátové kompresi JPEG2000. Pro video doporučuje JPEG2000 v kontejneru MXF, který je považován za archivní master formát. Dále se v doporučení pro video vyskytují formáty ProRes (kontejner MOV) a MPEG-2 a XDCAM. U ProRes a MPEG2 lze po rozkliknutí odkazu dohledat, že obsah v těchto formátech obdržela LOC od externích producentů. U XDCAM lze předpokládat, že se rovněž jedná o obsah, který jako digitální již vznikl. Kodek FFV1 v kontejneru MKV řadí LOC do kategorie „přijatelné“ a protože má omezenou schopnost zaznamenat časový kód (timecode), je doporučován pouze pro videa bez titulků.

FORMÁTOVÉ POLITIKY LOKÁLNÍCH INSTITUCÍ

Zatímco výše uvedená formátová doporučení jsou jako návod určena také širší odborné veřejnosti, formátové politiky lokálních institucí (př. univerzitní knihovny), jsou informací o tom, jaké formáty používají oni sami, případně, jaké formáty přijímají od externích dodavatelů.

Formátové politiky těchto institucí jsme začali sledovat, abychom si doplnili informace o rozšíření (adoption) rastrových formátů, tedy statických obrázků. Původně jsme vycházeli ze seznamu institucí použitého ve studii (Rimkus at al, 2014) srovnávající důvěru ve formáty u členů severoamerického sdružení ARL (Association of Research Libraries) a průběžně jej doplňovali dalšími. Podařilo se nám dohledat dvacet pět institucí, které své formátové politiky prezentují na internetu. S výjimkou tří se ostatní z těchto institucí v té či oné míře věnují i video formátům. Situace zde je ovšem mnohem komplikovanější.

Instituce	rok vydání	Doporučené (Podporované, Preferované, Udržitelné, Level 1)		Přijatelné (Omezená podpora, Známé, Přijatelné, Level 2)	
		kodek	kontejner	kodek	kontejner
Alberta University Lib.	2018	JPEG2000			avi
Archivematica	2019	FFV1	mkv		
Canada Library and Archives	2015	JPEG2000, Uncompressed	avi, mov	H.264, JPEG2000, DV-NTSC, Apple ProRes, Avid DNxHD, VC-1	mp2, mp4, avi, mov, wmv
Connecticut Uni. Lib.	2018			RealVideo	ram, mov, mpg
Cornell University Lib.	2019	JPEG2000, Uncompressed	avi, mov	Theora, H.261/262, 263, 264	ogg, avi, mov
Deep Blue (Michigan Uni.)	2011				avi, mov, mp1, mp2, mp4
Duke uni Library	2022	FFV1	mkv, avi, dpx	H.264, H.262	mov, mpg, mpeg, mp4
Florida Uni. Libraries	2012	JPEG2000, Uncompressed	mj2, avi, mov,	Theora, H.261/262, 263, 264	ogg, avi, mov
Hawai'i Uni.	2019				mpg, mov
Houston University	2018	JPEG2000	mxr		avi, mov, mpg, mp4
Minnesota University	2014				avi, mov, mp1, mp2, mp4, wmv
North Carolina State Archives	2012	Uncompressed	avi	H.262, H.264, JPEG2000, Theora, WMV	avi, mov, mp2/4, mxr, ogg, asf
North Carolina State Uni. Lib.	2018				mpg, mov
Northwestern University	nezn.	Uncompressed	avi, mov	H.261, H.262, H.264	avi, mpg, mp4, mov, mkv
Purdue University Libraries	2012				avi, mpg, mp4, mov
Smithsonian Archives	2017	JPEG2000	mov, avi		mp4
Southern Illinois Uni.	2008			RealVideo (rm)	mov, mpg
Tasmanian Archives	2015	JPEG2000, Dirac, Raw bitstream		H.261, H.262, H.263, H.264	avi, mov, wmv
Texas Uni. Of Austin	2022	JPEG2000, Uncompressed	avi, mj2, mov, mkv		
Texas A&M Uni.	2014		avi	H.264	mp4, mov
UK Data service	2014	JPEG2000	mp4, ogg, mj2		avhcd
Uni. of Washington Libraries	2014	FFV1, Unc., MJPEG, H.264	mkv, avi, mxr, mov, mp4	Theora, ProRes, JPEG2000, H.261, H.262	ogg, mov, mj2, mp1, mp2

Tabulka 1. Formátové politiky lokálních institucí.
V závorkách záhlaví jsou alternativní názvy kategorií používané některými institucemi

V porovnání s rastrovými formáty (Ostráková & Kopský, 2022), kde se všechny formátové politiky v kategorii „doporučené“ shodují na nekomprimovaném TIFF (doplněném u některých o jeden bezztrátově komprimovaný – JPEG2000) u video formátů (Tabulka 1.) jsou v kategorii „doporučené“ nekomprimované formáty pouze hruba u třetiny z nich. Bezztrátově komprimované kodeky jsou zde zastoupeny více, nejčastěji JPEG2000, méně FFV1 a v jednom případě (Tasmanian Archives) bezztrátově komprimovaný kodek Dirac, který původně pro své potřeby vyvinula BBC. U rastrových obrázků se v kategorii doporučené často vyskytovaly i ztrátově komprimované formáty. U video formátů jsme zaznamenali jen jeden takový případ (resp. jednu instituci, která má v doporučených dva ztrátové kodeky – MJPEG a H.264). U třetiny institucí pak v kategorii doporučených nejsou video

formáty vůbec, a uvedeny jsou pouze formáty „přijatelné“. Tyto instituce tedy zatím nemají svůj master formát a ve své formátové politice pouze oznamují, které formáty přijímají pro uložení. V této kategorii je u většiny institucí řada ztrátově komprimovaných kodeků, což je ale pochopitelné, protože se často jedná o obsah, který ve ztrátovém formátu byl danou institucí přijat, případně v něm už vznikl (např. studentské projekty, orální historie apod.). Metodiky tvorby formátových politik jsou značně nejednotné, což je vidět již z toho, že názvy kategorií přijetí formátu se u nich liší, viz záhlaví tabulky.

FORMÁTOVÁ DOPORUČENÍ INSTITUCÍ SPECIALIZOVANÝCH NA ARCHIVACI FILMŮ

Ve výstupech institucí specializovaných na archivaci filmových materiálů převažují informace o ochraně analogových originálů, a technické stránce jejich restaurování a digitalizace. Komentáře na téma formátů jsou u většiny z nich poměrně stručné.

Přehledovou publikaci na toto téma přináší Mezinárodní federace filmových archivů (Fédération Internationale des Archives du Film – FIAF), která má na svých stránkách dokument věnovaný digitalizaci filmového pásu. V závěrečné části dokumentu (Ruivo & Gant, 2019) je průzkum užití formátů u filmových archivů, z nichž většina ukládá filmy do DPX. Čtyři archivy navíc jako druhý formát využívají TIFF, který díky možnosti uložit více obrázků v jednom souboru (kontejneru) může sloužit i jako video formát. Bezztrátovou kompresi FFV1 používají dva, a JPEG2000 používá pro uložení jeden. Kodek JPEG2000 v kontejneru DCP ovšem využívá řada z nich pro distribuci, nicméně zde je použita ztrátová komprese. Nejčastěji používaná rozlišení jsou 2K a 4K, mezi méně obvyklá patřilo 3K (nativní rozlišení scanneru Arriscan používaného v Cinematica Brasileira), 6K a 8K

Studie FIAF sleduje i použitou bitovou hloubku. Minimální je 10 bitů na vzorek, což odpovídá i upozornění George Blooda v článku, citovaném v úvodu, který upozorňuje, že použití hloubky 8bitů na vzorek může způsobovat páskování (banding), kdy u pozvolných barevných přechodů vznikají při nízkém vzorkování ostré přechody mezi odstíny. Ve studii FIAF používá u formátů DPX 10 bitů na vzorek deset archivů, zbytek pak 16 bitů na vzorek, kromě jednoho, který používá 24 bitů na vzorek. U formátu TIFF používají všichni 16 bitů na vzorek, což je dáno tím, že u tohoto formátu není jiná možnost než 8 nebo 16.

Jednou z institucí sledovaných ve srovnání FIAF je Britský filmový institut (British Film Institute – BFI), který ve své vlastní formátové politice řeší nejen skeny filmového pásu, ale i video formáty. Doporučuje nekomprimovaný či bezztrátově komprimovaný formát (4:2:2 10bit) pro video s nízkým rozlišením, kompresní formáty (AVC Intra 100, Apple ProRes 422 (HQ) a HDCAM-SR.) pro video ve vysokém rozlišení (BFI, 2011).

Mezinárodní asociace pro zvukové a audiovizuální archivy (International association of sounds and audiovisual archives (IASA) – publikovala Instrukce k digitalizaci videozáznamů, v jejichž druhé (B) části se věnuje formátům (IASA, 2019). Pro uložení digitalizovaného videa doporučuje buď nekomprimovaný V210 v kontejnerech (AVI, MOV nebo MXF), nebo bezztrátové FFV1 (MKV, AVI, QuickTime) a JPEG2000 (MXF). Autoři upozorňují, že toto doporučení několika různých možností je nezbytné s ohledem na to, že v této oblasti zatím není konsensus.

V České republice se digitalizaci filmů věnuje Národní filmový archiv (NFA), jehož metodika (Zahradníček, 2014) současně řeší i jejich restaurování, nicméně je zaměřena pouze na obsah na

filmovém pásu. Pro archivaci skenů zdrojových materiálů volí kodek JPEG2000 v kontejneru MXF. V současnosti NFA spolupracuje s Národním archivem na přípravě „Národního standardu formátů pro archivaci“, který by měl být hotov koncem tohoto roku.

FFV1

Při rešerši další dostupné literatury jsme se nejčastěji setkávali s publikacemi o archivním využití kodeku FFV1, včetně digitalizace VHS kazet. Vzhledem k tomu, že tento kodek je uveden v řadě formátových doporučení i formátových politik, a splňuje podmínky pro archivační master formát, uvažovali jsme o něm jako o jednom z možných kandidátů pro ukládání obsahu z VHS kazet a zaměřili se na něj blíže. FFV1 se ukázal být kodekem, který má v archivnictví poměrně specifické postavení, kdy se část archivní komunity podílí na jeho vývoji, včetně specifikace a nástrojů pro práci s ním.

FFV1 je rychlý bezztrátový kodek, který je pro archivní komunitu atraktivní díky velmi detailnímu způsobu zabezpečení dat. Jednotlivá políčka videa jsou pro účely komprese rozdělena do 4, 6, 9, 12, 16, 24 nebo 30 segmentů (slices) a program vytváří kontrolní součty jak pro každé políčko filmu, tak pro každý jeho segment zvlášť. To znamená, že v případě poškození dat lze mnohem přesněji dohledat, kde k němu došlo. Reto Kromer (2018) to ve své analýze vhodnosti FFV1 pro archivaci označuje za klíčový faktor pro dosažení profesionální ochrany audiovizuálních souborů.

Dalším podstatným plusem je pro archiváře fakt, že tento kodek je otevřený, což jim přináší možnost ovlivnit jeho vývoj. Byl vytvořen v rámci projektu FFmpeg, který je zaměřen na vývoj open source software pro multimediální aplikace. Obsahuje většinu známých video kodeků a od roku 2013 (FFmpeg 2.1) i stabilní FFV1.3. Na vývoji verze 1.3 spolupracoval Peter Bubestinger z Rakouské mediátéky a Dave Rice z firmy MediaArea, která je přímo zapojená do archivní aplikace FFV1. Vývoj byl též Rakouskou mediátkou financován.

Rakouská mediátka byla jednou z prvních institucí, která se rozhodla již v roce 2011 pro využití formátu FFV1/MKV pro archivaci videa z formátů VHS a DigiBeta. Stalo se tak po intenzivním testování několika bezztrátových kodeků (Bubestinger, 2016). Autor v této práci též upozorňuje na počínající zájem subdodavatelských firem o kodek FFV1, zejména těch, které se digitalizací videa zabývají komerčně a často tuto práci vykonávají na zakázku pro paměťové instituce.

Další dobře zdokumentovaný příklad je z Indiana University, kde působí Media Digitization and Preservation Initiative (MDPI). Ta z nekomprimovaného formátu V210 přešla na FFV1, poté co její ředitel technických operací Mike Casey (2017) otestoval, zda obsah do něj převedený lze bezztrátově převést zpět na V210. V citovaném článku se vyjadřuje i k volbě kontejneru MKV (Matroska), u kterého oceňuje především podporu tvorby kontrolních součtů a rozsáhlých strukturovaných metadat.

Vlámský institut pro archivaci Meemoo se v roce 2020 rozhodl využít datové migrace z páskových jednotek LTO-6 na LTO-8 k současné formátové migraci z JPEG2000/MXF do FFV1/MKV (Meemoo, 2020). Pro JP2 se rozhodli v roce 2013, kdy FFV1 byl teprve ve vývoji. Na MXF jim nevyhovovalo, že je spravován uzavřeným konsorciem, nebyl publikován jako otevřený formát, a dovoluje větší množství

variant, což vede k nadměrné komplexitě. Původně do FFV1/MKV převáděli soubory ve formátech DV, DVCAM and DVCPRO, a jelikož se osvědčil, rozhodli se do něj konvertovat i JP2/MXF.

Významnou roli při vývoji formátu FFV1/MKV hraje konsorcium PREFORMA (PREservation FORMAts for culture information/e-archives), projekt EU, který zastřešuje několik iniciativ zabývajících se standardizací archivačních formátů, například vývoj validátoru formátu PDF-A, VeraPDF (v rámci konsorcia OPF, OPC a PDF Association) nebo vývoj formátu TI-A, standardizované verze TIFFu. V oblasti video formátů spolupracuje s firmou MediaArea vyvíjející s institucionální podporou otevřený software, která vyvinula validátor MediaConch (Conformance Checker) souborů s kodekem FFV1 v kontejneru MKV spolu se zvukem v LPCM – Linear Pulse Code Modulation (MediaArea, n.d.)

MediaArea rovněž připravila plán pro standardizaci tohoto kodeku u Komise pro technickou stránku internetu (Internet Engineering Task Force – IETF). Tato standardizační organizace byla zvolena proto, že jako jediná ze tří zvažovaných (IETF, SMPTE a ISO) zpřístupňuje specifikace formátů zdarma, což je v souladu s open source přístupem konsorcia PREFORMA. Kodek byl komisí standardizován v loňském roce jako RFC9043 (IETF, 2021).

Pracovní skupina při IETF s názvem Cellar (Codec Encoding for LossLess Archiving and Real-Time Transmission), pracuje na standardizaci FFV1/MKV se zvukem v kodeku FLAC (IETF, 2020). Skupina rovněž předložila návrh specifikace FFV1 verze 4 (IETF, 2022). E-mailová konference (mailing list) skupiny Cellar je spolu s konferencí FFmpeg hlavním diskusním fórem uživatelů FFV1.

NO TIME TO WAIT

Symposium No Time To Wait je setkání pracovníků zaměřených na standardizaci v práci s archivními formáty, a především uživatelů a zájemců o využití kodeku FFV1. Poprvé se konalo v roce 2016 v Berlíně souběžně s meetingem IETF (Blewer, 2016). Akce byla předznamenána rok předtím na meetingu IETF v roce 2015 v Praze, kde byli přítomni zástupci MediaArea i PREFORMA a společně s pracovníky IETF se zde zabývali koordinací standardizace FFV1/MKV.

Na symposiu archiváři představují své zkušenosti s implementací FFV1, a zazněly zde i kritické příspěvky. V následující přednášce například Kieran O'Leary popisuje problémy, které vyvstaly při pokusu transkódovat do kodeku FFV1 videa z kodeku ProRes, kdy docházelo k nesprávnému přenesení poměru stran (O'Leary, 2019). O'Leary uzavírá, že v tuto chvíli migraci z ProRes do FFV1 nedoporučuje. Kieran O'Leary ovšem předtím úspěšně použil FFV1/MKV jako cílový formát pro převod z nekomprimovaných formátů DPX a TIFF (O'Leary, 2016).

Ve stejném roce přednesl svou přednášku i Derek Buitenhuis, který prezentoval vývoj dekodéru pro FFV1 na základě jeho IETF specifikace (Buitenhuis, 2019). Velkým přínosem jeho práce byla zpětná vazba pro tvůrce specifikace, s nimiž konzultoval různé nejasnosti, což vedlo k řadě upřesnění. Ostatně význam zkratky RFC v názvu specifikace je „request for comments“ (žádost o komentáře), což bývá uváděno jako připomínka toho, že zpětná vazba je v komunitě vývojářů a uživatelů formátu FFV1/MKV vždy vítána.

DISKUZE

Vycházeli jsme ze studia doporučení autoritních institucí a dalších dostupných zdrojů, tedy většinou publikací a jako doplňujícího zdroje i formátových politik, které sice vzhledem ke svým nejednotným hodnotícím kritériím mají nižší reprezentativnost, ale pro orientační představu o rozšíření sledovaných formátů v paměťových institucích mají smysl je sledovat. Dospěli jsme k závěru, že pro dlouhodobé uložení obsahu digitalizovaných VHS kazet bude vhodné použít bezztrátovou kompresi. Zjistili jsme, že nejpoužívanějšími kodeky jsou JPEG2000 a FFV1 (v kontejnerech MXF, resp. MKV).

Při srovnání kodeků JPEG2000 a FFV1, zjišťujeme, že jejich stoupenci v podstatě tvoří dvě myšlenkové školy. První sama sebe označuje jako zaměřenou na „standard s velkým S“ (Murray, 2014), protože si zvolila kodek JPEG2000, který v té době již byl standardizován. Druhá, označovaná v článku Murrayové jako zaměřená na „open source“, dala přednost otevřenému kodeku FFV1, který standardizovala dodatečně.

Pro JPEG2000 sice rovněž existují open source programy, ale instituce, které se pro něj rozhodly, používají komerční software, a tak vývojáři tohoto open source, na rozdíl od FFV1, nemohou mít zpětnou vazbu z archivní komunity. Peter Bubestinger (2016) před rozhodnutím pro FFV1 testoval i programy pro JPEG2000. U souborů vytvořených komerčním nástrojem *Morgan „M-JPEG2000“ codec package* nesouhlasily kontrolní součty (framemd5 validace) a implementace JPEG2000 v FFmpeg neumožňovala použití bezztrátové komprese. Nakonec testoval implementaci „libopenjpeg“, která nepodporovala multithreading (současné využití všech jader procesoru), což se odrazilo na rychlosti. Kromě toho, že se pro JPEG2000 doporučuje spíše komerční software, je zpoplatněn i jeho standard, na rozdíl od FFV1.

Na druhou stranu u FFV1/MKV jsou již zmíněné problémy se záznamem časového kódu, kvůli kterým ho LOC zařadila pouze do přijatelných (acceptable) formátů. Z tohoto pohledu dává užití JPEG2000/MXF jako archivního masteru pro video smysl tam, kde sbírky obsahují i tzv. „broadcast material“, který obsahuje několik timecodů a titulků. Rovněž jsme viděli, že FFV1 nelze doporučit jako cílový kodek pro převod obsahu z kodeku ProRes, nicméně migrace ztrátových formátů se v tuto chvíli obecně nedoporučuje.

ZÁVĚR

Porovnáme-li přístupy k ukládání digitalizovaného analogového videa s těmi pro uložení naskenovaného filmového pásu, uvidíme zde jistý překryv. I u filmových scanů se setkáme s jejich uložením do bezztrátově komprimovaného formátu, ale nekomprimované formáty převažují. Filmy jsou také skenovány ve vyšším rozlišení než video a s použitím vyšší bitové hloubky. Naproti tomu u digitalizátů analogového videa se sice můžeme setkat i s uložením do nekomprimovaného formátu, ale bezztrátově komprimované kodeky zde převažují, přičemž se většinou jedná o JPEG2000 a FFV1.

FFV1/MKV v tuto chvíli není vhodný pro úplně všechny účely, ale pro uložení digitalizovaného obsahu z VHS kazet plně vyhovuje. Je standardizován pro účely archivace a má ověřenou specifikaci. Existuje pro něj software nejen na tvorbu, ale i pro jeho validaci. Rozšíření formátu FFV1/MKV sice není velké, ale to je do značné míry kompenzováno tím, že se jedná o rozšíření v archivní komunitě, které zaručuje, že jsou zkušenosti s jeho užitím pro dlouhodobé uložení. Pokud by v budoucnu bylo potřeba převést FFV1/MKV do nového archivačního master formátu, pak to bude možné při úplném zachování obsahu.

DEDIKACE

Článek vznikl na základě institucionální podpory dlouhodobého koncepčního rozvoje výzkumné organizace Národní knihovna České republiky poskytované Ministerstvem kultury ČR.

LITERATURA

BFI. (2011, listopad). *BFI Collection Policy*. British Film Institute.

<https://www2.bfi.org.uk/sites/bfi.org.uk/files/downloads/bfi-collection-policy-2014-07-02.pdf>

Blewer, A. (2016) *No Time to Wait: Standardizing FFV1 and Matroska for Preservation*. MediaArea.

<https://mediaarea.net/blog/2016/07/26/No-Time-To-Wait-Preservation-FFV1-Matroska-Symposium>

Blood, G. (2011) *Refining Conversion Contract Specifications: Determining Suitable Digital Video Formats for Medium-term Storage*. FADGI. [https://www.digitizationguidelines.gov/audio-](https://www.digitizationguidelines.gov/audio-visual/documents/IntrmMastVidFormatRecs_20111001.pdf)

[visual/documents/IntrmMastVidFormatRecs_20111001.pdf](https://www.digitizationguidelines.gov/audio-visual/documents/IntrmMastVidFormatRecs_20111001.pdf)

BROWN, Adrian. (2008) I Digital preservation guidance note 1. The National Archives (UK).

<http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

Bubestinger, P. (2016) Review and Comparison of FFV1 versus Other Lossless Video Codecs for Long-Term Preservation. Sustainable Audiovisual Collections Through Collaboration. Indiana University Press.

<https://publish.iupress.indiana.edu/read/sustainable-audiovisual-collections/section/1c4a2741-70ae-499e-b9ff-7b2b545232ec>

Buitenhuis, D. (2019) *Wrote an FFV1 Decoder in Go for Fun*. No Time To Wait - YouTube.

<https://www.youtube.com/watch?v=4HB7v7dItWE>

Casey, M. (2017, březen 27) *White Paper: Encoding and Wrapper Decisions and Implementation for Video Preservation Master Files*. Indiana University Media Digitization and Preservation Initiative (MDPI).

<https://mdpi.iu.edu/doc/MDPIwhitepaperrev.pdf>

FADGI. (2014). *Digital File Formats for Videotape Reformatting*. Federal Agencies Digital Guidelines Initiative.

https://www.digitizationguidelines.gov/guidelines/video_reformatting_compare.html?loclr=blogsig

Goethals, A. (2016, 16 dubna) *Video Format Matrix*. Harvard Library Digital Preservation Services Wiki.

<https://docs.google.com/spreadsheets/d/1rR7HNoQswcOrl66yeRRI2qMGDKzYQxitrOmD7nfVFGQ/edit>

IASA. (2019). *Guidelines for the Preservation of Video Recordings - Part B. Video Signal, Preservation Concepts, and Target Formats. Technical Committee Standards, Recommended Practices, and Strategies*, IASA-TC 06, Edition 1. [https://www.iasa-](https://www.iasa-web.org/sites/default/files/publications/IASA-TC_06-B_v2019.pdf)

[web.org/sites/default/files/publications/IASA-TC_06-B_v2019.pdf](https://www.iasa-web.org/sites/default/files/publications/IASA-TC_06-B_v2019.pdf)

IETF (2020) *Codec Encoding for LossLess Archiving and Realtime transmission (cellar) - Charter for Working Group*. Internet Engineering Task Force. <https://datatracker.ietf.org/wg/cellar/charter/>

IETF. (2021). *FFV1 Video Coding Format Versions 0, 1, and 3*. Internet Engineering Task Force. <https://datatracker.ietf.org/doc/html/rfc9043>

IETF. (2022). *FFV1 Video Coding Format Version 4*. Internet Engineering Task Force. <https://datatracker.ietf.org/doc/draft-ietf-cellar-ffv1-v4/>.

Ježek, P. (2018). *Změní AV1 a AVIF svět?* Root. <https://www.root.cz/clanky/zmeni-av1-a-avif-svet/>

KOST - Koordinationsstelle für die Dauerhafte Archivierung elektronischer Unterlagen (2021a) *Video Formats: Introduction*. https://kost-ceco.ch/cms/kad_video_de.html

KOST - Koordinationsstelle für die Dauerhafte Archivierung elektronischer Unterlagen (2021b) *Evaluation matrix*. <https://kost-ceco.ch/cms/Bewertung.html>

Kromer, R. (2018, 26. dubna). *Matroska and FFV1: One File Format for Film and Video Archiving?* https://retokromer.ch/publications/JFP_96.html

LOC. (2017) *Sustainability factors*. Library of Congress. <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml#technical>

LOC. (2021a, 30 listopadu). *Format Descriptions for Moving Images*. Library of Congress. <https://www.loc.gov/preservation/digital/formats/fdd/vid> MediaArea. (n.d.). [MediaConch https://mediaarea.net/MediaConcho_fdd.shtml](https://mediaarea.net/MediaConcho_fdd.shtml)

LOC. (2021b) *Recommended Formats Statement*. Library of Congress. <https://www.loc.gov/preservation/resources/rfs/RFS%202021-2022.pdf>

MediaArea. (n.d.). *MediaConch*. <https://mediaarea.net/MediaConch>

Meemoo. (2020, 2. září). *From MXF/JPEG2000 to MKV/FFV1: why we're switching to a new file format*. Meemoo Flemish Institute for Archives. <https://meemoo.be/en/news/from-mxfjpeg2000-to-mkvffv1-why-were-switching-to-a-new-file-format>

Murray, K. (2014, 3. prosince) *Comparing Formats for Video Digitization*. The Signal <https://blogs.loc.gov/thesignal/2014/12/co> MediaArea. (n.d.). [MediaConch https://mediaarea.net/MediaConchcomparing-formats-for-video-digitization/](https://mediaarea.net/MediaConchcomparing-formats-for-video-digitization/)

NARA. (2022a – 31. března) *NARA File Format Risk Matrix*. National Archives and Records Administration. <https://github.com/usnationalarchives/digital->

[preservation/blob/master/Digital_Preservation_Risk_Matrix/NARA_File_Format_Risk_Matrix_2020_331.xlsx](#)

NARA. (2022b – 4. květen) *Transfer Guidance Tables*. National Archives and Records Administration. <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html#digitalmovingimages>

O'Leary, K. (2016) *Introduction to FFV1 and Matroska for Film Scans*. <https://kieranjol.wordpress.com/2016/10/07/introduction-to-ffv1-and-matroska-for-film-scans/>

O'Leary, K. (2019) *Is ProRes/MOV to FFV1/MKV a Good Idea? No Time To Wait - YouTube*. [No Time To Wait - S04E04 - Kieran O'Leary - Is ProRes/MOV to FFV1/MKV a Good Idea? - YouTube](#)

Ostráková, N. & Kopský, V. (2022) Archival file format selection and design of file format sustainability matrix for the National library of the Czech Republic. Review of the NCD. <http://www.ncd.matf.bg.ac.rs/issues/40/2.pdf>

Ostráková, N. & Kopský, V. (2020) *Posuzování souborových formátů z hlediska dlouhodobého uchování a návrh metodiky pro Národní knihovnu České republiky*. *Knihovna – knihovnická revue* <https://knihovnavue.nkp.cz/archiv/2020-2/recenzovane-prispevky/posuzovani-souborovych-formatu-z-hlediska-dlouhodobeho-uchovavani-a-navrh-metodiky-pro-narodni-knihovnu-ceske-republiky>

Rimkus, K., Padila, T. Popp, T. and Greer, M. (2014) Digital Preservation File Format Policies of ARL Member Libraries: An Analysis. *D-Lib Magazine*, 20 (3/4), doi:10.1045/march2014-rimkus <https://www.dlib.org/dlib/march14/rimkus/03rimkus.html>

ROG, J. & WIJK, C. (2008). *Evaluating File Formats for Long-term Preservation*. National Library of the Netherlands. <https://wiki.dnb.de/download/attachments/83791059/Evaluating+File+Formats+for+Long-term+Preservation+unpubl.pdf?version=1&modificationDate=1382441353000>

Ruivo, C. & Gant, A. (2019) *Survey on Long-term Digital Storage and Preservation*. FIAF Technical Commission. https://www.fiafnet.org/images/tinyUpload/2019/04/Preservation_Digital_Statement_Final.pdf

Zahradníček, J. (2014) *Koncepce digitalizace, digitálního restaurování a digitální archivace v NFA v letech 2014–2020*. Národní filmový archiv. <https://nfa.cz/wp-content/uploads/2014/08/Koncepce-digitalizace-digit%a1n%adho-restaurov%a1n%ad-a-digit%a1n%ad-archivace-v-NFA-v-letech-2014-2020.pdf>

POZNÁMKA O AUTOROVI

Vojtěch Kopský

Autor této analýzy je formátový specialista Oddělení pro standardy Národní knihovny ČR. Toto oddělení analyzuje, zavádí a rozšiřuje standardy potřebné pro dlouhodobou ochranu digitálních dokumentů. Oddělení pro standardy také zastřešuje standardizaci v rámci projektu Národní digitální knihovny (NDK) a poskytuje konzultace ostatním digitalizujícím knihovnám a institucím.

E-mail: vojtech.kopsky@nkp.cz