

Dlouhodobá ochrana digitálního audiovizuálního obsahu

Audiovizuální (AV) záznamy z 20. století jsou ohroženy a jejich digitalizace je řešením, které s sebou přináší nový problém: jak dlouhodobě chránit digitální AV dokumenty. Současné technologie totiž nemají adekvátní řešení pro ochranu souborů s AV obsahem. Můžeme se sice inspirovat příklady dobré praxe v oblasti dlouhodobé ochrany digitálních dat, je však třeba si uvědomit tři věci: 1) AV sbírky by měly využívat stávající technologie pro digitální knihovny a ochranu digitálních dokumentů, 2) technologie podporující multimedia je třeba dále vylepšovat, 3) technologie skladování dat a informační technologie obecně by měly časem lépe podporovat specifické požadavky AV souborů.

Problémy ochrany digitálních audiovizuálních dokumentů

V evropských AV sbírkách (v archivech nebo jiných institucích) je podle odhadů kolem 50 milionů hodin zvukových dokumentů, filmů a videa, v převážné většině v analogových formátech. Kvůli zastarávání, zhoršování kvality médií a zastarávání formátů je dnes ohroženo asi 70 procent těchto dokumentů a do 30 let bude ohroženo všechno.

Velké digitalizační programy již začaly. Odhaduje se, že se v posledním desetiletí v Evropě digitalizovalo 10 milionů hodin AV materiálu. A zatímco se AV sbírky snažily rychle vyměnit pásky a gramofonové desky za počítačové soubory, aby zachránily jejich obsah, okolní svět si začal uvědomovat, že vyvstává nový problém: dlouhodobá ochrana digitálních dat.

Technické problémy spojené se správou a provozem rozsáhlých sbírek digitálních souborů řeší technologie digitálních knihoven. O soubory se musíme starat, je třeba je pojmenovat, přesouvat do nových úložišť (a to často!), kopírovat pro zpřístupňování, zakódovat podle měnících se podmínek zveřejňování a kontrolovat. Digitální objekty musí mít metadata. Ta se vytvářejí katalogizací nebo automatickým sklizením (metadata jsou potřebná k zajištění globální dostupnosti digitálních objektů ve standardizovaných systémech). Jakmile sbírky dosáhnou určité velikosti, je prakticky nemožné je udržovat manuálně. Pro automatické vytváření, údržbu a správu přístupových oprávnění velkých sbírek digitálních objektů se nabízejí nástroje technologie digitálních knihoven, které jsou popsány v řadě manuálů.

Dva světy: Technologie digitálních knihoven se nejvíce rozšířily v akademickém světě. AV sbírky se však z velké části nalézají mimo tuto sféru. Největšími držiteli AV obsahu jsou televizní a rozhlasové vysílatelé. Další velké sbírky se nalézají v muzeích filmu a v jiných kulturních institucích (jednou z největších filmových sbírek ve Spojeném království je „Imperial War Museum“). Vysílatelé jsou různorodí, avšak jejich počítačová a technická experti, ani management rozhodující o technologických otázkách a zajišťující finance zpravidla neví vůbec nic o akademických knihovnách a technologiích digitálních knihoven.

Při dlouhodobé ochraně AV souborů je třeba především znát, rozumět, financovat a používat dostupné nástroje pro správu digitálních knihoven, které dokáží soubory strukturovaně organizovat.

Za druhé je třeba si uvědomit, že digitální knihovny jsou nástroje pro správu digitálních objektů (tj. zajišťují, že soubory jsou dostupné a neztratí se), ale neřeší problém dlouhodobé ochrany těchto objektů. Zastaráváním digitálních souborů se zabývá technologie dlouhodobé ochrany digitálních dokumentů. Soubory v zastaralém formátu lze převést podle nových standardů a do nových formátů, funkčnost zastaralých softwarových a hardwarových technologií lze prodloužit emulací.

Jsou vypracována kritéria hodnocení spolehlivosti digitálních repozitářů i obecná metodika dlouhodobé ochrany digitálních dat (model OAIS). V IT odděleních správců AV sbírek těžko najdeme někoho, kdo kdy slyšel o modelu OAIS, což ztěžuje financování a implementaci ochrany digitálních AV dokumentů. Naštěstí je tu evropský projekt MEMORIES, který přizpůsobuje model OAIS a podobné techniky speciálně pro prostředí AV sbírek.

Za třetí je třeba si uvědomit, že specifické požadavky na práci s AV soubory nejsou plně podporovány technologiemi digitálních knihoven a ochrany digitálních dokumentů, jak budeme probráno dále v textu.

Specifické problémy digitálních audiovizuálních dat a souborů

Vzhledem k zmiňované existenci „dvou světů“ nejsou profesionální vysílací formáty (zvláště MXF) obvykle podporovány nástroji digitálních knihoven a ochrany digitálních dokumentů. Ostatní „standardní“ formáty mají podporu lepší, ale mnohé z nich (např. AVI, WMV) jsou proprietárními formáty, což samo o sobě představuje problém pro jejich uchování.

Reference a další informace

Audiovisual Status Surveys:PRESTO: <http://presto.joanneum.ac.at>PrestoSPACE: <http://www.prestospace.eu>TAPE: <http://www.tape-online.net/survey.html>
http://www.tapeonline.net/docs/audiovisual_research_collections.pdf**General Guides to Preservation:**<http://www.bbcarchive.org.uk/><http://digitalpreservation.ssl.co.uk/>**A general list of digital curation tools:**<http://twiki.dcc.rl.ac.uk/bin/view/Main/DevelopmentToolList>JHOVE: <http://hul.harvard.edu/jhove/>**DROID:**<http://droid.sourceforge.net/wiki/index.php/>**Introduction PRONOM:**<http://www.nationalarchives.gov.uk/pronom>**National Library of New Zealand Library metadata extractor:**<http://meta-extractor.sourceforge.net/>OAI: <http://www.openarchives.org/OAI/openarchivesprotocol.html>**OAIS:**http://www.dpconline.org/docs/lavoie_OAIS.pdf**Migration:**<http://www.library.cornell.edu/iris/migration/>**Emulation:**<http://www.dlib.org/dlib/october00/granger/10granger.html>**Repository evaluation criteria reference:**<http://journals.tdl.org/jodi/article/view/199/180>**Projects and initiatives:****MEMORIES:**<http://www.memories-project.eu/>**European Digital Library:**<http://www.europeana.eu/>**Formats:****MXF:**<http://www.digitalpreservation.gov/formats/fdd/fdd000013.shtml>**WAV format specification:**<http://www.digitalpreservation.gov/formats/fdd/fdd000001.shtml>**Presentations:**

Video Formatting and Preservation, Carl Fleischhauer, NDIIPP DLF Forum, Philadelphia, 6 November 2007

<http://www.diglib.org/forums/fall2007/presentations/Fleischhauer.pdf>

Další problémy se týkají konkrétního obsahu souborů.

- Většina AV souborů je komprimována. To, co se ztratilo z „původní kvality“ během komprese, zůstane ztraceno již navždy. Procesy ochrany by měly dbát na zachování co největší kvality, což je vlastnost, kterou je potřeba definovat a implementovat do současných technologií.
- Obsah vázaný na časové rozložení potřebuje nástroje s časovou dimenzí (katalogizace, navigace, editace).
- AV soubory jsou komplexní soubory. Koncept „obalu“ (wrapper) byl vyvinut právě pro rozpoznání komplexity typického AV souboru: mnohonásobné signály, složitější typy metadat - včetně časoprostorových (titulky) a číselných (časový kód).
- Uchovávání AV dat se týká mnoha typů souborů: bezeztrátově a ztrátově kódovaných, různých forem proxy přístupů (podpora přístupu v různých formátech, např. Real, Windows Media, MPEG, AVI, QuickTime, Flash), souborů v různých fázích editace a nových kombinací. Týká se také celé škály právních aspektů: mnoho zainteresovaných stran i správců sbírek, nejednotnost norem v různých zemích. Je tedy nutné uchovávat komplexní celek obsahující AV signály reprezentující informace, metadata a související právní data.

Zpřístupňování

Knihovny mají tradici jednotného přístupu. Souborné katalogy založené na standardizovaných metadatech zpřístupňují „jakoukoliv knihu odkudkoliv“.

Mnoho AV sbírek je tradičně uzavřených, nebo přístupných pouze profesionálům nebo pro obchodní účely.

Digitální knihovny pokračují v tradici rozšířeného a jednotného přístupu, často v národním nebo mezinárodním měřítku jako v případě Evropské digitální knihovny (EDL). Aby mohly být AV sbírky přístupné prostřednictvím velkých projektů jako EDL, potřebují technologie digitálních knihoven. Na druhou stranu tyto digitální knihovny musí vynaložit více úsilí k pochopení výše uvedených problémů digitálních audiovizuálních dat a souborů. Digitální knihovny potřebují zejména nástroje na časově strukturovaný přístup jak k AV signálu, tak k metadatům (např. rozsah práv v rámci jediného AV souboru se může každou minutou měnit). Většina AV obsahu je v držení institucí, které v minulosti nespolupracovaly s žádnými knihovnami a které mohou upřednostňovat omezování přístupu ke „svému“ obsahu. Marketing, ochrana značky a právní otázky brání vytvoření „Evropského audiovizuálního portálu“. EDL možná nebude nikdy obsahovat záznam vysílání BBC.

Co dělat

Přes zmíněné problémy lze učinit několik základních doporučení pro uchovávání AV obsahu

- **Uchovávejte artefakt:** Uchovávejte „originál“, i když je v komprimované podobě. Uchovávejte bity, ať se děje, co se děje. AV obsah má jednu výhodu, je ho hodně v relativně malém počtu formátů. Mohou existovat techniky pro přehrávání bitů.
- **Dekódujte do nekomprimované formy** a uchovávejte v nekomprimované formě (zároveň s originálem). V případě videa je to náročný požadavek (100 GB za hodinu TV obraz o 625 řádcích), ale úložná zařízení jsou nyní velmi levná.
- **Vylepšete metadata.** Nestačí pouze zaznamenat příponu souborů (např. wav, avi). Existuje např. více než 50 registrovaných variant kódování v rámci definice formátu wav. Koncovku mpg používají formáty MPEG-1 i MPEG-2. V ideálním případě bude k dispozici nástroj pro extrakci metadat. V opačném případě je potřeba manuální testování a dokumentace.
- **Nejste sami:** používejte registry formátů souborů, softwarové repozitáře, platformy pro emulace a směrnice pro uchovávání uvedené v citovaných odkazech.