

Doporučení pro pořizování datových souborů při digitalizaci analogových originálů

Smyslem digitalizace analogových originálů je jejich rozšířená dostupnost (všechny druhy dokumentů), případně ochrana/záchrana (tato funkce je obecně pro digitalizaci uznávána pro záchranu zvukových a video nahrávek).

Z těchto cílů vyplývají i parametry pořizovaných digitálních datových souborů:

1. archivní soubor pořizovaný pro záchranu (úplnou náhradu) analogového dokumentu
2. soubor pořizovaný pro zpřístupnění dat, nepřímou ochranu nebo dokumentaci:
 - a. archivní soubor
 - b. uživatelské soubory

Archivní soubor jako úplná náhrada

Tyto soubory jsou pořizovány ve špičkových světových zvukových nebo video archivech. Jejich parametry odrážejí maximalistický přístup k zachování informací; jejich pořízení má smysl jen při špičkovém zvládnutí zachycení zdrojového analogového signálu, což v běžných podmínkách nepřichází v úvahu a tento materiál se těmito soubory nebude zabývat. Viz dále materiály IASA.

Soubor pro zpřístupnění, nepřímou ochranu nebo dokumentaci

Tento soubor musí zachytit a předat tu část analogových informací, jež jsou nezbytné ke splnění cíle komunikace s uživatelem souboru. Všechny informace zachytit nelze, protože při digitalizaci dochází z principu digitálního záznamu k redukci analogového signálu **vždy**. Z tohoto důvodu lze uvést jen doporučení, nicméně konkrétní rozhodnutí pořizovatelů digitálních záznamů je vždy individuální a musí zohlednit řadu faktorů, jako jsou charakter originálu (jeho poškození nebo míra detailů), kvalita pořizovacího zařízení (například kvalita optiky nebo čipu/zadní stěny) a jeho nastavení, jakož i nastavení podmínek při digitalizaci (kvalita osvětlení, fixace originálu) a plánované cíle digitalizace. Z těchto důvodů se někdy konkrétní parametry mohou lišit nebo být nižší než v tomto doporučení, aniž by to znamenalo snížení kvality výsledné komunikace.

Osoby rozhodující o digitalizaci by měly být seznámeny se základními parametry (rozlišení, barevná hloubka, datový tok; druhy a vlastnosti kompresních schémat a ztrátových kompresních technik) a vlastnostmi digitálního záznamu a jejich vlivem na kvalitu tohoto záznamu.

Obecně lze dále říci, že archivní soubor by měl být vždy ISO formát nebo formát celosvětově doporučený pro archivaci daného druhu digitálních dat, kdežto volba uživatelského souboru více záleží na vlastnostech cílené komunikace mezi aplikacemi, které zpřístupňují digitální data, a uživatelem, jemuž by měl být poskytnut takový stupeň komfortu, který odpovídá cílům programu/projektu digitalizace. Základní odpovědnost za volbu parametrů datových souborů

a jejich formátů nese vždy osoba odpovědná za projekt a osoba, která digitální záznamy pořizuje.

Míra znalostí, týkajících se problematiky datových souborů, schopnost uplatnit je v praxi, flexibilita při přijímání rozhodnutí a praktické zkušenosti jsou nejzávažnějšími faktory ovlivňujícími kvalitu zamýšlené komunikace. I velmi dobrá doporučení mohou přinést nedobré výsledky a naopak řada dobrých výsledků nevychází ze žádných obecně vyhlášených doporučení.

Doporučení pro jednotlivé typy dokumentů

Psané a tištěné dokumenty

a) barevné rukopisy a vzácné barevné tištěné dokumenty

Cílem je co nejvíce se přiblížit originálu, proto zde kromě nastavení běžných parametrů datových souborů velmi důležitou roli hraje **schopnost rekonstrukce věrného barevného podání** originálu. Z tohoto důvodu bude věnována pozornost kalibraci zařízení a uchování jejich ICC profilů, používání a reprodukce standardních (například GretacMacbeth) a specializovaných (například Memoria) kalibračních tabulek a jejich matematického vyjádření, jakožto i aplikace standardů pro metadatový popis digitálního obrazu (Z39.87, DIG35 a případně dalších).

Archivní soubor:

TIFF 24 bit nekomprimovaný

JPEG 24 bit, standardní kódování, ztrátová komprese max. 5 (Quality Factor) v Adobe Photoshop, resp. nastavená ztráta dat do 5 dílků nebo procent v jiných editorech, používajících cca. 100 stupňový průběh komprese
Obecně nejméně 200 dpi, lépe více

Uživatelský soubor

Parametry dle cílů komunikace, ale pro základní užití vždy 24 bit; doporučit lze:

- § JPEG s parametry nezbytnými pro splnění cílů komunikace, u vyšších kvalit viditelný vodotisk
- § nové formáty, užívající kompresi typu wavelet, především JPEG2000 nebo DjVu při vypnutí segmentace do vrstev (formát IW44) – tyto formáty však nejsou doporučeny WWW konsorciem

b) nebarevné rukopisy a nebarevné tištěné dokumenty zvláště moderní vydané po r. 1800

Pro tyto typy dokumentů (na zvážení u textových rukopisů a vzácnějších tisků; většinou u moderních tištěných dokumentů) lze pro jejich reprezentaci místo barevného 24 bitového obrazu volit 8 bitový obraz v 256 odstínech šedi, neboť cílem není věrná reprezentace

barevného podání. Vzhledem k tomu jsou bezpředmětná i další opatření mající za cíl věrnou reprodukci.

Archivní soubor

TIFF 8 bit v 256 odstínech šedi, nekomprimovaný

JPEG 8 bit v 256 odstínech šedi, standardní kódování, ztrátová komprese max. 5 (Quality Factor) v Adobe Photoshop, resp. nastavená ztráta dat do 5 dílků nebo procent v jiných editorech, používajících cca. 100 stupňový průběh komprese

Obecně nejméně 200 dpi, lépe více

Uživatelský soubor

Parametry dle cílů komunikace:

- § JPEG s parametry nezbytnými pro splnění cílů komunikace
- § nové formáty, užívající kompresi typu wavelet jako JPEG2000
- § nejvhodnější jsou však technologie Mixed Raster Content, tzn. plná segmentace ve formátech DjVu nebo LuraDocument LDF resp. JPM; v případě problémů se zobrazením jen kódování wavelet
- § lze užít i formát PDF
- § pokud nejsou cílovými formáty výše uvedené, lze snížit barevnou hloubku na 4 (16 odstínů šedi) resp. 2 bity (4 odstíny šedi) a použít pro reprezentaci formáty GIF nebo PNG
- § lze zvážit aplikaci černobílého zobrazení (může se týkat i dokumentů v bodu a)); k tomu je třeba vhodná volba systému redukce barev s ohledem na výslednou funkci černobílého obrazu (čtení OCR, podání polotónů atp.) a příp. před převodem vyrovnat nerovnoměrnost zón v obraze; použít pro prezentaci formát TIFF/CCITT Fax Gr. 4, PNG nebo GIF; případně použít kompresi typu JBIG2 (DjVu JB2 nebo LuraDocument JBIG2).

c) moderní barevné dokumenty, například periodika

Pro digitalizaci a zpřístupnění moderních barevných dokumentů se využívá barevné hloubky 24 bit – jinak ***archivní soubor*** obdobně jako bod b).

Pro ***zpřístupnění*** dop. využít technologii Mixed Raster Content, zejm. formátů DjVu příp. LDF s plnou segmentací obrazu.

d) kartografické materiály

Kartografické materiály, plány atp. se snímají s takovým rozlišením, aby bylo možné rozpoznat nezbytné detaily. Volená barevná hloubka je 24 příp. 8 bit dle barevnosti originálu. Vznikají tak obvykle velmi velké datové soubory.

Pro ***archivaci*** dop. užívat TIFF nekomprimovaný nebo JPEG, kompresní poměr pro archivaci obdobně jako výše.

Pro ***zpřístupnění*** dop. užívat technologii MrSID.

Fotografie

Při digitalizaci fotografií dop. využít velkého rozlišení – orientačně 600 dpi.

Archivní formát

TIFF 24 nebo 8 bit nekomprimovaný dle požadavků na záznam barev.

JPEG 24 nebo 8 bit dle požadavků na záznam barev; komprese viz doporučení pro archivní JPEG výše.

Uživatelský soubor

JPEG s kvalitativními parametry dle očekávání/požadavků uživatele. Z dalších formátů dop. PNG (účinná bezztrátová komprese), příp. další formáty s kompresí wavelet, zejm. JPEG2000.

Zvukové dokumenty

Vznikají převodem klasických zvukových dokumentů, jako jsou voskové válečky, černé šelakové a vinylové desky, magnetické pásky, rozhlasové vysílání apod.

Archivní formát

WAVE, nejméně 16 bit (dop. 24 bit), 48 kHz

Uživatelský soubor

PCM CDDA (audio CD); MP3 (příp. WMA) - datový tok dle smyslu komunikace (MP3 dop. 160 kbps pro případný pozdější zápis CD audio).

Video a film

Vznikají převodem klasických video nosičů, jako jsou kazety VHS a SVHS, Video 8, Video 8 záznam XR, Video Hi 8 atp. resp. klasického filmového materiálu či televizního vysílání. Kvalitativní parametry by měly odpovídat kvalitě digitalizovaného záznamu.

Archivní formát

MPEG-2 DVD kompatibilní, video PAL 720 x 576 bodů, VBR kódování, nejméně 4 mbps (pro kazety VHS a Video /Hi/8 vyšší hodnoty nemají smysl); zvuk 48 kHz nejméně 224 kbps (MPEG-1, L2); kvalitnější snímky /TV vysílání/ dop. 8 mbps video

Uživatelský soubor

Pokud je cílem přehrání ve stolním přehrávači:

- § Cílové médium video DVD, pak hodnoty odpovídající archivnímu formátu, případná redukce do 3 mbps (video) a na 192 kbps audio (MPEG-1, L2) možná vč. změny kódování zvuku na LPCM nebo Dolby Digital.

§ Cílové médium Video CD (VCD), pak legální nastavení pro VCD, tj. video MPEG-1 PAL 352 x 288 bodů, 1,15 mbps, audio 224 kbps (MPEG-1, L2).

§ DivX (profil Home Theater) formátu AVI nebo DIVX (*.avi, *.divx)

Pro další způsoby sdílení videa lze doporučit formáty jim vlastní, tj. kromě výše uvedených formátů a nastavení i WMV resp. MPEG-4 ve formátu AVI, kodek Xvid, nebo MP4 Nero Digital (Home Theater). Jiná nastavení a další formáty jsou možné pro jiné způsoby komunikace.

Obecné poznámky

Národní programy digitalizace užívají tyto formáty:

a) *Memoria* – archivace JPEG (*.jpg); zpřístupnění sady obrazů v JPEG, resp. GIF (*.gif - malý náhled a černobílá reprezentace); pro zpřístupnění map je používán formát MrSID (*.sid).

b) *Kramerius* – archivace JPEG (*.jpg); zpřístupnění DjVu (*.djv, *.djvu)

Doporučené bitmapové formáty WWW konsorcia jsou GIF, JPEG a PNG.

Další informace: <http://digit.nkp.cz>, Technické standardy, Data

Adolf Knoll dne 24. října 2005
adolf.knoll@nkp.cz