

Štandardy a technologické odporúčania pre digitalizáciu

Štandardy a technologické odporúčania pre digitalizáciu

1.	Otvorené formáty a štandardy	3
2.	Snímanie a uchovanie textového obsahu	3
3.	Snímanie a uchovanie statických obrazov.....	4
4.	Snímanie a uchovanie zvukovo-vizuálnych záznamov	7
5.	Snímanie a uchovanie zvukových záznamov	7
6.	Dlhodobá archivácia.....	7
7.	Voľba médií.....	8
1.	Sprístupňovanie textového obsahu	9
2.	Sprístupňovanie statických obrazov	10
3.	Sprístupňovanie zvukovo-vizuálneho obsahu	10
4.	Sprístupňovanie zvukového obsahu	11
5.	Virtuálna realita:.....	11
6.	Web stránky:	11
7.	Bezpečnosť sprístupnenia digitálneho obsahu	12

Otvorené formáty a štandardy

Otvorené štandardy pomáhajú podpore interoperability a zaisťujú, že zdroje sú znovu použiteľné a môžu byť vytvorené a modifikované rozličnými aplikáciami. Taktiež pomáhajú vyhnúť sa závislosti na konkrétnom dodávateľovi.

Pri aplikáciách štandardov je potrebné brať ohľad na nasledovné východiská:

- snaha o implementáciu riešení založených na otvorených formátoch a štandardoch
- podpora a participácia na vývoji otvorených riešení
- zabezpečenie podpory existujúcich riešení pre komunikáciu a spoluprácu s otvorenými formátmi a štandardmi

Uchovávanie a manažment digitálneho obsahu

Snímanie a uchovanie textového obsahu

Textový dokument je v tomto prípade synonymický pojem pre textový obsah, resp. digitálny obsah založený na texte alebo, textovej forme.

Textový dokument **by mal byť** vytvorený a riadený v štruktúrovanom formáte, ktorý je vhodný pre generovanie HTML alebo XHTML dokumentov pre doručovanie.

Najvhodnejšia voľba pre uchovávanie textového dokumentu je vo forme založenej na definícii SGML zodpovedajúcej publikovanej DTD schéme alebo na definovanom XML dokumente. Textový obsah **môže byť** uchovaný aj v tzv. jednoduchých textových dokumentoch (*plain files*) alebo vo vnútri nejakého druhu databázy.

V niektorých prípadoch **môže byť** pre uchovávanie textových dokumentov použitý súkromný formát Adobe Portable Document Format (PDF). PDF súbory sú kompaktné a môžu byť zobrazované a tlačené za použitia voľne dostupného softvéru Adobe Acrobat Reader. Pri použití súkromného formátu vzniká potencionálne nebezpečenstvo prevzatia takéhoto dokumentu (napr. tret'ou stranou a pod.), a preto by malo byť zaistené preskúmanie migračnej stratégie, ktorá umožní v budúcnosti prechod na otvorené štandardy.

Štandardy:

ISO 8879:1986. Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML).

- HTML 4.01 HyperText Markup Language - <http://www.w3.org/TR/html401>
- XHTML 1.0 The Extensible HyperText Markup Language - <http://www.w3.org/TR/xhtml1/>
- Extensible Markup Language (XML) 1.0 – http://www.w3.org/TR/REC_xml/
- Adobe Portable Document Format (PDF) – www.adobe.com

Snímanie a uchovanie statických obrazov

Digitálne statické obrazy spadajú do dvoch kategórií:

- Rastrové digitálne obrazy (alebo bitmapové)
- Vektorové digitálne obrazy (objektovo - orientované)

Digitalizačný proces takmer vždy generuje rastrové obrazy; vektorové obrazy sú väčšinou vytvorené ako výstup grafických softvérov.

Pri vytváraní a uchovávaní rastrových obrazov je potrebné vziať do úvahy dva základné faktory:

- Formát súboru
- Parametre kvality digitálneho výstupu

Rastrové obrazy

Rastrové obrazy **by mali byť** vždy uchované v nekomprimovanej forme / formáte generovanom digitalizačným procesom bez akéhokoľvek následného zásahu aplikáciou alebo iným prostriedkom. Rastrové obrazy **MUSIA byť** vytvorené použitím jedného z nasledujúcich formátov:

- TIFF – Tagged Image File Format
- JPEG/SPIFF – JPEG Still Picture Interchange File Format
- JPEG 2000
- PNG – Portable Network Graphics
- GIFF – Graphical Interchange Format

Existujú dva základné parametre, ktoré je pri digitalizácii nutné vziať do úvahy:

- **Priestorové rozlíšenie** – frekvencia, v ktorej sú zachytávacím zariadením zobrazené vzorky originálu, vyjadrené ako množstvo vzoriek na palec (samples per inch – spi) alebo viac všeobecnejšie ako množstvo pixelov na palec (pixels per inch – ppi) vo výslednom digitálnom obraze
- **Farebné rozlíšenie (bitová hĺbka)** – množstvo farieb (alebo úrovní jasu) dostupných reprezentovať rôzne farby (alebo odtiene šedej) originálu, vyjadrené ako množstvo

bitov dostupných reprezentovať farebnú informáciu, napr. farebné rozlíšenie 8 bitov znamená, že je dostupných 256 rozdielnych farieb.

Vo všeobecnosti fotografické obrázky **by mali byť** vytvorené ako TIFF obrázky.

Digitálne obrázky **by mali byť** vždy vytvorené v najvyššom možnom rozlíšení a farebnej bitovej hĺbke vzhľadom na digitalizovaný obraz.. Digitalizačný proces **MUSÍ** identifikovať minimálnu úroveň kvality and informačnú hustotu.

Odporúčanie: Pre fotografické tlače **by malo byť** aplikované rozlíšenie 600 bodov na palec (dpi) s farebnou bitovou hĺbkou 24 bitov alebo 8 bitov pre odtiene šedej. Pre 35 mm statický film **by malo byť** aplikované rozlíšenie 2400 dpi.

Štandardy:

- Tagged Image File Format (TIFF)
<http://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf>
- Joint Photographic Expert Group (JPEG)
<http://www.w3.org/Graphics/JPEG/>
- JPEG Still Picture Interchange File Format
<http://www.jpeg.org/public/spiff.pdf>
- JPEG 2000
<http://www.jpeg.org/jpeg2000/>

Grafické nevektorové obrázky

Počítačovo-generované obrázky ako sú logá, ikony a kresby **by mali byť** vytvorené ako PNG alebo GIF obrázky s rozlíšením 72 dpi. S kreslenými obrazmi, ktoré pochádzajú z digitalizácie fyzických obrazov **by malo byť** nakladané tak ako je uvedené v časti „*Rastrové obrázky*“

Štandardy:

- Portable Network Graphics
<http://www.w3.org/TR/PNG>

Vektorové obrázky

Vektorové obrazové dokumenty **by mali byť** vždy vytvorené a uchované použitím otvoreného formátu ako je napríklad formát SVG a XML jazyka pre popisovanie takýchto dokumentov.

Môže byť použitý aj súkromný formát Macromedia Flash avšak **by mala byť** zabezpečené preskúmanie migračnej stratégie, ktorá umožní v budúcnosti prechod na otvorené štandardy. Použitie textu vo vnútri Flash formátu **by malo byť** vypustené, aby bolo možné vyvíjať viacjazyčné verzie.

Štandardy:

- Scalable Vector Graphics (SVG)
<http://www.w3.org/TR/SVG/>
- Macromedia Flash
<http://www.macromedia.com/>

Snímanie a uchovanie zvukovo-vizuálnych záznamov

Zvukovo-vizuálny záznam **by mal byť** uchovaný v nekomprimovanej forme bez akéhokoľvek postprodukčného zásahu (napr. redukcia šumu obrazu a pod.). Záznam **by mal byť** uchovaný použitím nekomprimovaného RAW AVI formátu bez použitia akéhokoľvek obrazového kodeku.

Odporúčanie č. 1: Zvukovo-vizuálny záznam **by malo byť** vytvorený v rozlíšení 720x576 pixelov, 25 fps - obrazových snímok za sekundu a 24-bitovej farebnej hĺbke a systém kódovania farieb PAL.

Pripúšťa, ale neodporúča sa, použiť aj formáty MPEG-1, MPEG-2, MPEG-4 alebo súkromné formáty WMV, ASF a Quicktime).

Odporúčanie č. 2: Pri digitalizácii zvukovo-vizuálneho záznamu s primárnym cieľom iba záchrany obsahu alebo archivácie a následného prevodu na nové analógové médium **SA MUSÍ** vytvoriť a uchovať digitálna kópia minimálne s parametrami definovanými v Odporúčaní č.1.

Snímanie a uchovanie zvukových záznamov

Zvukový záznam **by mal byť** uchovaný v nekomprimovanej forme bez akéhokoľvek postprodukčného zásahu (napr. redukcia šumu a pod.)

Záznam **by mal byť** vytvorený a uchovaný použitím nekomprimovaného formátu Microsoft WAV alebo Apple AIFF s 24 bitovým stereo zvukom na frekvencii 48/96 KHz pre digitálny originál.

Pripúšťa sa ale neodporúča sa použiť aj formáty ako MP3, WMA, RealAudio alebo SUN AU.

Dlhodobá archivácia

Pri dlhodobej archivácii je nutné sa snažiť zachovať tri základné ciele:

- technologické uchovanie
- technologickú emuláciu
- migráciu údajov

Prvé dve snahy zaisťujú možnosť prístupu k digitálnemu objektu použitím pôvodného softvéru a hardvéru alebo emuláciou pôvodného prostredia.

Pre dlhodobú archiváciu **by mala byť** zaistená stratégia založená na možnosti migrácie a zahrnutá do implementácie riešenia. V tomto prípade sú objekty/obsah migrovaný z jedného technologického prostredia do ďalšieho.

Kritické body migrácie pre dlhodobú archiváciu je Snímanie metaúdajov, ktoré sú potrebné pre podporu manažmentu obsahu a migračného procesu.

Archivácia dokumentov **by mala byť** zaistená na dva rôzne typy archivačného úložiska/média a na dvoch rôznych zemepisných miestach.

Voľba médií

Pre dlhodobú aj krátkodobú archiváciu je veľmi dôležité mať znalosť rôznych médií, pretože tie vyžadujú rôzne softvérové a hardvérové zariadenia pre prístup a majú rôzne podmienky archivácie a požiadavky na uchovávanie.

Média **by mali byť** obnovované v pravidelnom cykle doby životnosti média identifikovanom buď výrobcom alebo nezávislým zdrojom.

Ukážka príkladu životnosti média skladovaného pri rôznych klimatických podmienkach

Médium	25RH 10°C	30RH 15°C	40RH 20°C	50RH 25°C	50RH 28°C
D3 magnetická páska	50 rokov	25 rokov	15 rokov	3 rokov	1 rok
DLT magnetický páskový box (magnetic cartridge)	75 rokov	40 rokov	15 rokov	3 rokov	1 rok
CD/DVD	75 rokov	40 rokov	20 rokov	10 rokov	2 rokov
CD ROM	30 rokov	15 rokov	3 rokov	9 mesiacov	3 mesiacov

Typy médií

Média pre archiváciu sa v základnom koncepte rozdeľujú do troch základných skupín:

1. Mikrofilm
2. Magnetické média
3. Optické média

V súčasnosti sú magnetická páska a mikrofilm najrozšírenejšie média pre dlhodobú archiváciu ale optické média sa stávajú stále viac bežné a predovšetkým vhodné aj pre dlhodobú archiváciu.

Optické média sa vyvíjajú rýchlo a mali by eventuálne nahradiť zvyšné média pre dlhodobú archiváciu. V súčasnosti však magnetické disky ponúkajú rýchlejší čas prístupu, čo im dáva lepší predpoklad využitia za účelom hromadného uchovávaní.

Poznámka: Existuje množstvo diskusií vo vnútri ISO, ANSI a ICA ohľadne najvhodnejšieho uchovávaní elektronických záznamov. Nasledovné hodnoty ilustrujú príklady z praxe:

- Priemerná teplota: +18°C/-5°C
- Relatívna vlhkosť: +40%/-5%
- Interval medzi prepisom: 10 rokov

Sprístupňovanie digitálneho obsahu

Predpokladá sa, že prístup koncového používateľa ku elektronickým zdrojom bude primárne cez Internetové protokoly. Príprava pre publikovanie vyžaduje spracovanie digitálneho surového obsahu tak, aby sa generovali digitálne objekty vhodné pre použitie v kontexte Internetu, zvyčajne redukciou kvality, tak aby sa dosiahla veľkosť súborov vhodná pre transfer prostredníctvom dátových sietí a Internetu.

Pre získanie veľkého počtu používateľov a jeho zvyšovanie **by malo byť** zaistené sprístupnenie digitálneho obsahu v rôznych veľkostiach alebo formátoch alebo rozlíšeníach a farebných bitových hĺbkach.

Zvukový a zvukovo-vizuálny obsah **by mal byť** prístupný na stiahnutie (downloading) ako aj pre streaming.

Dôraz **musí byť** kladený na skutočnosť, že variácie existujú v

- typoch hardvérových zariadení a softvéru na strane klienta použitého používateľom
- úrovni obmedzenia šírky pásma, s ktorou používateľ disponuje

Sprístupňovanie textového obsahu

Použitá znaková sada v texte **by mala byť** prenášaná v HTTP záhlaví a prípadne zaznamenaná aj vo vnútri dokumentu. ISO 10646, UTF8

Textový obsah **MUSÍ BYŤ** sprístupnený v formáte XHTML 1.0 alebo HTML 4 (alebo v následnej verzii).

V niektorých prípadoch sa pripúšťa ale neodporúča použiť ako náhradu súkromné formáty PDF, RTF alebo MS WORD.

Štandardy:

- HTML 4.01 HyperText Markup Language - <http://www.w3.org/TR/html401>
- XHTML 1.0 The Extensible HyperText Markup Language - <http://www.w3.org/TR/xhtml1/>
- Portable Document Format (PDF) <http://www.adobe.com/>

1. Sprístupňovanie statických obrazov

Fotografické obrázky

Fotografické obrázky **MUSIA BYŤ** sprístupnené na Webe vo formáte JPEG/SPIFF. Mala by byť zaistená prístupnosť obrazov v rôznej kvalite a veľkosti tak aby bola zaistená čitateľnosť primerane ku kontextu použitia.

Náhľady **by mali byť** sprístupnené v rozlíšení 72 dpi a 24 bitovej alebo 8 bitovej farebnej hĺbke s použitím maximálne 100-200 pixelov pre najdlhší rozmer.

Obrázky pre celo-obrazovú prezentáciu **by mali byť** sprístupnené v minimálne rozlíšení 150 dpi a 24 bitovej alebo 8 bitovej farebnej hĺbke s použitím 600 pixelov pre najdlhší rozmer.

Grafické nevektorové obrázky

Obrázky **by mali byť** sprístupnené použitím formátu Graphical Interchange Format (GIF) alebo Portable Network Graphics (PNG).

Vektorové obrázky

Obrázky **by mali byť** sprístupnené použitím formátu Scalable Vector Graphics (SVG).

Sprístupňovanie zvukovo-vizuálneho obsahu

Pri sprístupnení zvukovo-vizuálneho obsahu **by malo byť** prihliadnuté k možnostiam používateľského prístupu a jeho obmedzenej šírke pásma a **mala by byť** poskytnutá používateľovi možnosť prístupu k rôznemu rozsahu súborov a streamov rozdielnej kvality.

Stahovanie (downloading)

Zvukovo-vizuálny obsah na stiahnutie **by mal byť** sprístupnený použitím formátu MPEG-1 alebo súkromným formátmi Microsoft Audio Video Interleave (AVI), Windows Media Video (WMV) alebo Apple Quicktime.

Streaming

Zvukovo-vizuálny obsah pre streaming **by mal byť** sprístupnený použitím formátov Microsoft Advanced Streaming Format (ASF), Windows Media Video (WMV) alebo Apple Quicktime.

Sprístupňovanie zvukového obsahu

Pri sprístupnení zvukového obsahu **by sa malo** prihliadnuť na možnosti používateľského prístupu a jeho obmedzenej šírke pásma a **mala by byť** poskytnutá používateľovi možnosť prístupu k rôznemu rozsahu súborov a streamov rozdielnej kvality.

Sťahovanie (downloading)

Zvukový obsah **by mal byť** sprístupnený v zhustenej forme použitím formátu MPEG Layer 3 (MP3) alebo súkromných formátov RealAudio (RA) alebo Microsoft Windows Media Audio (WMA). Dátový tok **by mal byť** 256 kbps pri zachovaní CD zvukovej kvality alebo 160 kbps pri zachovaní postačujúcej zvukovej kvality.

Nekomprimovaný zvukový záznam **môže byť** poskytnutý vo formáte Microsoft WAV/AIFF alebo Sun AU formáte.

Streaming

Zvukový obsah pre streaming **by mal byť** sprístupnený použitím formátu MPEG Layer 3 (MP3) alebo súkromných formátov RealAudio (RA) alebo Microsoft Windows Media Audio (WMA).

Virtuálna realita:

Formáty pre VR **MUSIA BYŤ** prístupné prostredníctvom rozhrania schopného prezentovať objekty na štandardne dostupných technologických zariadeniach a prostrediach s ohľadom na technologický čas.

Formát Apple Quick Time VR (QTVR) nie je skutočným formátom VR ale ponúka niektoré užitočné možnosti v tejto oblasti pre simuláciu virtuálnej reality.

Štandardy:

- Web3D Consortium
<http://www.web3d.org>
- QuickTime VR
<http://www.apple.com>

Web stránky:

Stránka **MUSÍ BYŤ** dostupná prostredníctvom internetového prehliadača. Toto je normálna zaistene použitím protokolov HTML, XHTML a http 1.1. Ak sú použité iné protokoly (napr. Z39.50) prístupnosť **MUSIA** zaistiť komunikačné brány pre prístup k údajom prostredníctvom web prehliadača.

Stránka **MUSÍ BYŤ** prístupná cez rôzne prehliadače, hardvérové riešenie, softvérové platformy a pre rôznych koncových používateľov.

Stránka **NESMIE BYŤ** prehliadateľná len prostredníctvom jednej softvérovej alebo hardvérovej platformy.

Stránka **MUSÍ BYŤ** prístupná pre prehliadače ktoré plne uspokojujú odporúčania organizácie W3C ako sú štandardy HTML/XHTML, kaskádové štýly CSS a DOM – Document Object Model.

Stránka **MUSÍ MAŤ** implementované riešenie W3C Web Accessibility Initiative (WAI), aby dosahovali vysoký stupeň prístupnosti aj pre znevýhodnených používateľov.

Štandardy:

- Cascading Style Sheets (CSS), Level 2
<http://www.w3.org/TR/REC-CSS2/>
- Web Content Accessibility Guidelines (WCAG) 1.0
<http://www.w3.org/TR/WCAG10/>
- HTML 4.01 HyperText Markup Language
<http://www.w3.org/TR/html401>
- XHTML 1.0 The Extensible HyperText Markup Language
<http://www.w3.org/TR/xhtml1/>
- Web Accessibility Initiative (WAI)
<http://www.w3.org/WAI/>

Bezpečnosť sprístupnenia digitálneho obsahu

Zariadenia použité pre doručovanie digitálneho obsahu musia pracovať s maximálne možným zabezpečeným spôsobom. **MUSIA BYŤ** dodržané všetky odporúčania v manuáloch použitých operačných systémov. Všetky známe bezpečnostné záplaty **MUSIA BYŤ** včas aplikované.

Zariadenie **MUSÍ BYŤ** konfigurované pre nutný minimálny počet sieťových služieb. Ak je to možné, zariadenie by sa malo nachádzať za firewallom s prístupom do internetu len na tých portoch, ktoré sú nevyhnutné pre zabezpečenie sprístupnenia digitálneho obsahu.

Správa a používanie osobných informácií **MUSÍ** vyhovovať legislatíve. Ak sú informácie s osobnými údajmi prenášané z klienta na server alebo späť cez internet **MUSÍ BYŤ** použitý protokol Secure Sockets Layer (SSL) pre kryptovaný prenos údajov.

Štandardy:

- Secure Sockets Layer (SSL) 3.0
<http://wp.netscape.com/eng/ssl3/>

Zdroje:

Minerva - Technical Guidelines for Digital Cultural Content Creation Programmes,
Version 1.0: Revised 08 April 2004,

Guidelines on best practices for using electronic information, How to deal with machine-readable data and electronic documents, Luxembourg: Office for Official Publications of the European Communities, 1997 : ISBN 92-828-2285-0

Technical Advisory Service for Images Advice Paper, Jún 2006,
<http://www.tasi.ac.uk/advice/creating/image.html>,

NOF-Digitise Technical Advisory Service, The Digitisation Process, An Information Paper from the NOF Technical Advisory Service, Jún 2006
http://www.ukoln.ac.uk/nof/support/help/papers/digitisation_process/