
Popis

Cíl: Cílem této metodiky je doporučení vhodných formátů archivních kopií dokumentů (Master Copy) a metadat pro výstupy z digitalizace z knihovných UK. Respektování vybraných formátů a metadat usnadní harmonizaci produkovaných digitalizátů a dlouhodobou ochranu na celouniverzitní úrovni ¹.

1. Metadata (MD)

Rozlišujeme tyto základní typy metadat:

- **Popisná** (někdy též nazývána bibliografická)
 - U popisných metadat platí "čím více, tím lépe". Pro popis jakýchkoliv typů digitalizovaných dokumentů je žádoucí využívat standardní popisná MD. Na UK jsou nejčastěji využívány [MARC21](#), [Dublin Core](#) a [MODS](#). Jediným "povinným" prvkem je v popisných metadatech je identifikátor. Může jít např. o systémové číslo z [CKIS](#) (Centrální katalog Univerzity Karlovy).
- **Strukturální**
 - Definují vnitřní organizaci či strukturu digitálního objektu. V případě zpřístupnění jsou nezbytná pro správné zobrazení a navigaci digitálního objektu. Svě významné postavení mají i v oblasti dlouhodobé ochrany/archivace. **Pro strukturální metadata je žádoucí využít dílčí část (<structMap>) metadatového standardu [METS](#) ².**
- **Administrativní**
 - Slouží především ke správě a řízení digitálních objektů. Poskytují časové informace o vzniku a úpravách digitálního objektu, specifikují okolnosti digitalizace. Často do této kategorie spadají ochranná a technická metadata.
- **Ochranná** (nebo též archivační)
 - Podporují a dokumentují proces dlouhodobé ochrany. Někdy jsou považovány za podsložku technických a administrativních MD. Zahrnují například informace o dokumentech a způsobu jejich využití; změny, které byly na souborech provedeny, odkazuje na provenienci; udržuje seznam zodpovědných osob za dlouhodobou ochranu apod. Za minimum v tomto ohledu lze považovat informaci o provenienci (kdy, kdo a jakým nástrojem výsledné digitalizáty) vytvořil.
- **Technická**
 - Technická metadata reprezentují technické charakteristiky počítačových souborů výstupů digitalizace (např. údaj o datovém formátu zdroje, velikosti souboru, kontrolním součtu aj.). Měla by respektovat XML schéma [NISO Metadata for Images in XML](#) (NISO MIX). Toto schéma nabízí formát pro výměnu a/nebo uchování dat specifikovaných v Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006).

2. Dlouhodobá ochrana - formáty

Kritéria: rozšířenost, podpora, ochranné mechanismy (DRM apod.), zveřejnění, kvalita dokumentace, stabilita, snadná identifikace a validace, práva duševního vlastnictví, podpora metadat, komplexnost, interoperabilita, životaschopnost, vícenásobná použitelnost (re-usability) ³.

Je žádoucí volit formáty bez komprese. Vodotisky pro archivní kopie řešit pomocí autentizačního bitového vzorku (tzv. skrytý vodotisk), nikoliv pomocí viditelného vodotisku.

1. Preferované formáty

Název	Přípona	MIME Type	PUID	Další informace
Tagged Image File Format (TIFF)	.tiff .tif	image/tiff image/tiff-fx		ISO 12234-2:2001, Photography – Electronic still picture imaging –

¹ Předmětem tohoto doporučení nejsou uživatelské kopie (User Copy), které jsou generovány z archivních digitalizátů.

² METS představuje standard pro kódování popisných, administrativních a strukturálních metadat k objektům, vyjádřený pomocí jazyka XML.

³ <https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>
https://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf

				Removable memory – Part 2: TIFF/EP Image data format (Tag Image File Format / Electronic Photography) [TIFF/EP] TIFF [Tagged Image File Format], revision 6.0, final, June 3, 1992. Available from: < http://partners.adobe.com/public/developer/tiff/ >
JPEG 2000 part 1 (JP2)	.jp2	image/jp2	x-fmt/392	ISO/IEC 15444-1:2004, Information technology – JPEG 2000 image coding system: Core coding system (Also issued as ITU-T.800) ISO/IEC 15444-4:2004, Information technology – JPEG 2000 image coding system: Conformance testing

2. Akceptovatelné formáty

Název	Přípona	MIME Type	PUID	Další informace
JPEG 2000 part 2 (JPX)	.jpx .jpf	image/jpx	fmt/151	
Portable Network Graphics (PNG)	.png	image/png	fmt/13 (verze 1.2)	
Graphics Interchange Format (GIF)	.gif	image/gif	fmt/4 (verze 89a)	

3. Technická specifikace digitalizace

- **kontrolní součty** - pro kontrolu neporušenosti vytvářených souborů je nutné generovat alespoň jeden ze tří nejčastěji využívaných kontrolních součtů - MD5, SHA nebo CRC)
- **názvová konvence souborů** - pro potřeby případného univerzitního LTP systému je vhodné dodržovat jednotné názvosloví výsledných souborů digitalizačních výstupů. V názvech souborů se nedoporučuje používat diakritiku, mezery a symboly
 - Návrh názvu souboru: identifikátor_fakulta/soucast_zkraceny-nazev_dalsi-volitelne-popisky.formát. Identifikátor může být např. SYSNO z Alephu nebo jiný jednoznačný identifikátor, který je evidován (např. 123456789 ETF Evangelia skrin12 police7.tif)
- **evidence průběhu digitalizace** plánovaných a již digitalizovaných dokumentů, včetně přiděleného identifikátoru/názvu souboru, který identifikátor obsahuje (např. pomocí některého z tabulkových procesorů)
- **strukturování a seskupování logických celků** - pomocí METS nebo např. při vkládání dat do DigiToolu formou CSV
- **namátková kontrola** naskenovaných výstupů, aby bylo možné včas doladit případné nedostatky
- **DPI/PPI** - minimální počet pixelů na palec pro archivní dokumenty (Master Copy) je 300 dpi. U původních předloh většího formátu a kde to umožňuje velikost zálohovacích zařízení, je možné pro obrazové dokumenty využít i vyšší DPI/PPI.

4. Zpřístupnění pro obraz

Podrobnější popis repozitářů UK pro všechny typy dokumentů lze najít na tomto webu, v sekci [Repozitáře](#).

Název	URL	Typ dat	Kontakt
Digitální univerzitní repozitář DigiTool	http://digitool.is.cuni.cz		

	aliases: http://repozitar.cuni.cz http://repositor.cuni.cz http://repository.cuni.cz	Podrobněji na stránkách Ústřední knihovny	digitool (at) cuni.cz
Kramerius FSV	http://kramerius.fsv.cuni.cz (plné texty pouze v budově knihovny FSV)	Podrobněji na stránkách knihovny FSV	knihovna (at) fsv.cuni.cz

5. Návrh workflow (fáze 1-7)

Fáze	Popis
1. výběr a příprava dokumentů	Výběr dokumentů k digitalizaci, kontrola záznamu v CKIS, kontrola v Registru digitalizace (příp. nahlášení záměru), apod. Konzultace s Ústřední knihovnou, volba repozitáře pro zpřístupnění digitalizátů, zajištění vhodného úložiště (viz fáze 5).
2. příprava digitalizace	Ustanovení odpovědných osob, evidence průběhu digitalizace apod.
3. pilotní provoz a testování	Pilotní testování workflow, a to včetně pilotní předávky dat do repozitáře. Zhodnocení kvality zpřístupnění.
4. proces digitalizace	Proces skenování (respektování vnitřní metodiky), ukládání do vhodných archivačních formátů, vytvoření archivní kopie (Master copy) a uživatelské kopie (User copy), průběžná namátková kontrola, apod.
5. úprava digitalizačních výstupů	Postprodukce - kontrola kvality, ořezy, sestavení finálních digitálních objektů apod., přiřazení MD (technických, popisných apod.).
6. uložení, dlouhodobá ochrana	Zápis na vhodná úložiště, zabezpečení min. 2 kopií (online, offline).
7. zpřístupnění	Zpřístupnění uživatelských kopií do vybraného repozitáře nebo digitální knihovny.

6. Terminologie

Vysvětlení některých výše uvedených termínů a akronymů lze najít v abecedně řazeném [seznamu](#) .