

# Národní standard formátů pro archivaci

Verze k 13. 6. 2022 – k připomínkám pro pracovní skupinu pro architekturu a řízení ICT

Národní archiv

Zpracovávání omezené množiny vybraných formátů v rámci životního cyklu dokumentu u původce nebo případně posléze archiváře v archivu – tzv. formátová normalizace – se ukázalo efektivním nástrojem řešení středně- a dlouhodobého uložení a zpřístupňování digitálních objektů dle mezinárodního standardu Otevřeného archivačního informačního systému (ČSN ISO 14721:2014).

Tato strategie se odráží v definici druhů dokumentů a stanovených formátů, prezentovaných v § 23 vyhlášky č. 259/2012 Sb. Primárně jsou určeny ke standardizovanému zajištění příjmu, vyřízení a uložení dokumentů v prostředí elektronického systému spisové služby. Mají zajistit jak občanům, tak úřadům, podání a doručení dokumentu ve formátu, který mohou zobrazit většinou běžnými prostředky výpočetní techniky bez nutnosti instalace specializovaných (proprietárních) programů. Současně jsou vhodné vzhledem ke své povaze i pro krátkodobé, střednědobé i dlouhodobé uložení v prostředí elektronických spisoven a digitálních archivů.

Zároveň však tato skupina formátů nemůže ve své úplnosti postihnout komplexní digitální objekty, které původci vytvářejí mimo elektronické systémy spisové služby, zejm. ve (specializovaných) informačních systémech, a to mnohdy i již historických s omezenými možnostmi exportu, či spravovaných nebo dokonce zveřejňovaných prostřednictvím různých dalších aplikací v podobě webových stránek/portálů, sociálních sítí atd. I tento informační obsah může mít a má trvalou hodnotu danou politickým, hospodářským, právním, historickým, kulturním, vědeckým nebo informačním významem. V neposlední řadě se v prostředí digitálních archivů uchovává řada digitálních reprodukcí archiválií vzniklých z analogové podoby. V řadě těchto případů mohou při volbě uchovávacího formátu převažovat na místo otevřenosti formátu kritéria hustoty záznamu a efektivity uložení komplexních objektů, kde se předpokládá jejich prezentace v tomtéž nebo obdobném specializovaném prostředí (např. rozsáhlé multimediální archivy, geodatové sbírky, databázové systémy atd.).

Podrobný popis jednotlivých formátů využívá údaje Kongresové knihovny ve Washingtonu ([https://www.loc.gov/preservation/digital/formats/fdd/browse\\_list.shtml](https://www.loc.gov/preservation/digital/formats/fdd/browse_list.shtml)).

Informaci, zda příslušný formát v elektronickém systému spisové služby je vhodný při exportu do digitálního archivu, obsahují Formátová pravidla jako součást kontroly vstupních datových balíčků (SIP) na adrese <https://validatorsip.nacr.cz>.

Legenda k následujícím tabulkám:

formát – název formátu souboru

verze – verze formátu souboru

PUID – identifikátor formátu v technickém registru PRONOM

(<https://www.nationalarchives.gov.uk/PRONOM/>)

kódování – vyžadované kódování

standard – zveřejněná specifikace formátu souboru

výstup eSSL – formáty/operace požadované pokud se jedná o výstup ze spisové služby

archivace – formát je vhodný pro dlouhodobé uchování

## Audio

Zahrnuje formáty používané ke kódování nahraného zvuku, k uložení informací se využívají ztrátové či bezztrátové metody kódování.

Obecné požadavky:

Bitrate min. 16 bitů, doporučeno 24 bitů, vzorkovací frekvence alespoň 44,1 kHz, doporučuje se vzorkování 96 kHz. Je však vhodnější ponechat nativní vzorkování než jej softwarově zvyšovat.

Výstup pro výběr dokumentů za archivácie a jejich uchovávání v digitálním archivu.

Formát	verze	PUID	kódování	standard	výstup eSSL	archivace
MPEG-1 Audio Layer II nebo MPEG-2 Audio Layer II (MP2)		fmt/198			výstupní formát	ano
MPEG-1 Audio Layer III nebo MPEG-2 Audio Layer III (MP3)		fmt/134		ISO/IEC-11172-3; ISO/IEC 13818-3	výstupní formát	ano
Waveform audio format (WAV)		-	modulace PCM	Multimedia Programming Interface and Data Specifications 1.0. IBM Corporation and Microsoft Corporation, August 1991. <a href="http://www.tactilemedia.com/info/MCI_Control_Info.html">http://www.tactilemedia.com/info/MCI_Control_Info.html</a>  Multimedia Data Standards Update April 15, 1994. <a href="http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/RIFFNEW.pdf">http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/RIFFNEW.pdf</a>	výstupní formát	ano
Free Lossless Audio Codec (FLAC)	1.21	fmt/279		FLAC Format Specification version 1.21. <a href="http://flac.sourceforge.net/format.html">http://flac.sourceforge.net/format.html</a>	§ 23 odst. 5	ano

## Binární soubory (data)

Zahrnují formáty strojově vytvořené zápisy bitů v binární soustavě. Jejich interpretace je zcela závislá na počítačovém programu. Většinou se doporučuje ponechat soubory v původní podobě.

## Binární soubory (spustitelné)

Spustitelné binární soubory se doporučuje ponechat v původní podobě.

## Databáze a strukturovaná data

Kategorie zahrnuje databázové formáty, tabulky, statistická či vědecká data. Vzhledem ke komplexnosti rozsáhlých databázových systémů je vhodné upřesnit způsob exportu pro trvalé uložení v archivu s příslušným digitálním archivem/NDA již při návrhu systému. Export může mít podobu pravidelného výstupu datových vět před jejich smazáním z databáze, pravidelné předání snímku databáze po dohodnutém časovém úseku nebo export celé databáze včetně upřesnění nejčastějších databázových dotazů pro zpřístupnění databáze v archivu.

Formát	verze	PUID	kódování	standard	výstup eSSL	archivace
Extensible Markup Language (XML)/ XML Schema Definition (XSD)	1.0	fmt/101, x-fmt/280			výstupní formát/ponechat	ano
Document Type Definition (DTD)		x-fmt/315			výstupní formát	ano
SIARD	2.2	fmt/1196 (platí pro verzi 2.1)		DILCIS Board/Swiss Federal Archives <a href="https://dilcis.eu/content-types/siard">https://dilcis.eu/content-types/siard</a>	individuální posouzení digitálním archivem/NDA	individuální posouzení digitálním archivem/NDA
Comma Separated Values (CSV)		x-fmt/18			§ 23 odst. 6/individuální posouzení digitálním archivem/NDA	individuální posouzení digitálním archivem/NDA
JSON Data Interchange Format (JSON) vč. JSON-LD		fmt/817; fmt/880 (JSON-LD)			§ 23 odst. 6 + originál /individuální posouzení digitálním archivem/NDA	individuální posouzení digitálním archivem/NDA
Microsoft Excel (XLS)	97--	fmt/61, fmt 62 (dodat l verze			§ 23 odst. 2/ § 23 odst. 6 + originál	individuální posouzení digitálním archivem/NDA

		pro Apple?)				
Microsoft Excel (XLSX)	2007-	fmt/214			§ 23 odst. 2/ § 23 odst. 6 + originál	individuální posouzení digitálním archivem/NDA
OpenDocument Spreadsheet (ODS, OTS)	1.2	fmt/295			§ 23 odst. 2/ § 23 odst. 6 + originál	individuální posouzení digitálním archivem/NDA

## E-mail

Kategorie zahrnuje zprávy zasílané s využitím výpočetní techniky, zejména protokolu SMTP. Primárně se pro ukládání e-mailových zpráv využije formát PDF/A. Přílohy e-mailů je nutné převést dle požadavků na ostatní formáty. Původní elektronická podání se mohou zároveň předat navíc v původním formátu. Celé e-mailové účty se předávají ve formátu PST/MBOX (nutné zohlednit ochranu soukromí a u veřejnoprávních původců umožnit odstranění informací soukromé povahy).

Výstup pro výběr dokumentů za archiválie a jejich uchovávání v digitálním archivu.

Formát	verze	PUID	Kódování	standard	výstup eSSL	archivace
PDF/A	1-4	fmt/95; fmt/354; fmt/476; fmt/477; fmt/478; fmt/479; fmt/480; fmt/481		ISO 19005	výstupní formát (přílohy dle formátu)	ano
Internet Message Format (EML)		fmt/278; fmt/950		RFC 2822	§ 23 odst. 2 (přílohy dle formátu) + originál dle volby	§ 23 odst. 2 (přílohy dle formátu) + originál dle volby
Microsoft Outlook Item Message Format (MSG)	97-2003	x-fmt/430			§ 23 odst. 2 (přílohy dle formátu) + originál dle volby	§ 23 odst. 2 (přílohy dle formátu) + originál dle volby
Microsoft Personal Folders Format (PST)	2003-2007	x-fmt/249			pro celý účet ponechat	ano
MBOX Email Format (MBOX)		fmt/720		RFC 4155	pro celý účet ponechat	ano

## Fonty

Fonty se ponechávají v původní podobě, pokud není stanoveno jinak.

## Internetové a intranetové stránky (web)

Webové stránky mohou zahrnovat jak obsah přístupný v síti Internet, tak i obsah, který je prezentován v rámci lokální sítě intranet.

Již při designu webových stránek je vhodné dodržet zásady přístupnosti – viz např. Metodický pokyn k zákonu č. 99/2019 Sb., o přístupnosti internetových stránek a mobilních aplikací a o změně zákona č. 365/2000 Sb., o informačních systémech veřejné správy a o změně některých dalších zákonů, ve znění pozdějších předpisů (včetně požadavků normy EN 301 549 V3.2.1), což usnadní i realizaci exportu pro trvalé uložení.

Primárním způsobem uložení webových stránek v prostředí Internet je harvesting, kdy rozsah a frekvenci je nutné dopředu domluvit s příslušným archivem. Tento způsob však nedokáže zahrnout export databází, autorizovaného obsahu, komplexních JAVA skriptů nebo třeba streamovaného videa, které musí být při exportu předány způsobem odpovídajícím charakteru těchto dat (dle kategorizace výše).

Lokálně vytvořený export by měl zahrnout obsah celé domény včetně vnitřně odkazovaných URL.

Výstup pro výběr dokumentů za archiválie a jejich uchovávání v digitálním archivu.

Formát	verze	PUID	kódování	standard	výstup eSSL	archivace
PDF/A (pouze pro export jednotlivých stránek)	1-4	fmt/95; fmt/354; fmt/476; fmt/477; fmt/478; fmt/479; fmt/480; fmt/481		ISO 19005	výstupní formát	ano
HTML/HTML aj. vzniklé exportem (např. pomocí HTTrack)		-			individuální posouzení digitálním archivem/NDA	ano
WARC (pro harvesting archivem)	1.1	fmt/1281		<a href="https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/">https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/</a>	/individuální posouzení digitálním archivem/NDA, rozbalit + originál	ano

## Komprimace dat

U souborů pro komprimaci dat se vyžaduje jejich dekomprimace a uplatnění požadovaných formátových převodů na obsažené soubory. V původní podobě se předává pouze ve vybraných případech (kontejnery ASiC).

## Kontejner

Kategorie zahrnuje elektronické dokumenty využívající elektronické autentizační prvky (např. FO/ZFO, ASiC). U většiny se vyžaduje převod obsažených dokumentů do výstupního datového formátu a zároveň předání originálu.

## Obraz disku

U obrazů disku se vyžaduje jejich dekomprimace a uplatnění požadovaných formátových převodů na obsažené soubory. Zároveň se předává obraz disku v původní podobě.

## Obrázek (rastrový)

Výstup pro výběr dokumentů za archiválie a jejich uchování v digitálním archivu.

formát	verze	PUID	kódování	standard	výstup eSSL	archivace
Tagged Image File Format (TIFF)	6	fmt/353		TIFF Revision 6.0 Final — June 3, 1992 Adobe Systems Incorporated	výstupní formát	ano
Portable Network Graphics (PNG)	1.2	fmt/13		ISO/IEC 15948	výstupní formát	ano
Joint Photographic Experts Group File Interchange Format (JPEG/JFIF)	1.02-	fmt/44; x- fmt/398; x- fmt/390; x- fmt/391; fmt/645; fmt/1507		ISO/IEC 10918	výstupní formát	ano
Jpeg2000 (JP2)	Part 1 (JP2)	x- fmt/392		ISO/IEC 15444-1	§ 23 odst. 3/individuální posouzení digitálním archivem/NDA	individuální posouzení digitálním archivem/NDA

## Obrázek (vektorový)

U grafiky využívající vektorové algoritmy se vyžaduje převod do PDF/A v podobě náhledu a zároveň předání souboru v původní podobě.

## Prezentace

Uplatňuje se postup dle požadavků pro Textové dokumenty.

## Prostorová data (GIS)

Kategorie zahrnuje formáty pro zaznamenání prostorových, tj. geografických a mapových dat.

Vzhledem ke své komplexnosti může být řešením exportu z informačního systému spravujícího prostorová data datový balíček tzv. GeoSIP dle metodiky Transformace digitálních prostorových dat pro účely trvalého uložení v digitálním archivu (<https://www.mvcr.cz/soubor/metodicky-navod-c-3-2022-narodniho-archivu-transformace-digitalnich-prostorovych-dat-pro-ucely-trvaleho-ulozeni-v-digitalnim-archivu.aspx>).

Výstup pro výběr dokumentů za archivace a jejich uchovávání v digitálním archivu.

formát	verze	PUID	kódování	Standard	výstup eSSL	archivace
Geographic Tagged Image File Format GeoTIFF (TIF, TIFF)	1.0	fmt/155		<a href="http://geotiff.maptools.org/spec/geotiffhome.html">http://geotiff.maptools.org/spec/geotiffhome.html</a>	ponechat	ano
Geography Markup Language (GML)	3.2.1.-	x-fmt/227		ISO 19136:2007	ponechat (vhodné připojit náhled ve VDF)	ano
GeoPackage (GPKG)	1.0-1.31	fmt/729 zalias SQLite Database File Format. musí se upřesnit)		OGC standard <a href="http://www.opengis.net/doc/S/geopackage/1.1">http://www.opengis.net/doc/S/geopackage/1.1</a>	ponechat (vhodné připojit náhled ve VDF)	ano
GeoJSON (GEOJSON)	1.0	fmt/1367		RFC 7946 <a href="https://tools.ietf.org/html/rfc7946">https://tools.ietf.org/html/rfc7946</a>	ponechat (vhodné)	ano

					připojit náhled ve VDF)	
LiDAR Point Cloud (LAS)	1.4	fmt/59 – (alias AutoCAD Last Saved Layer State, musí se upřesnit asi novým PUID)		LAS Specification 1.4, <a href="http://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf">http://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf</a>	Ponechat (vhodné připojit náhled ve VDF)	Ano

### Sociální sítě (médiá)

Obsah prezentovaný skrze sociální média (např. Facebook, Twitter, Slack, LinkedIn, Instagram atd.).

Primárním způsobem uložení webových stránek v prostředí Internet je harvesting, kdy rozsah a frekvenci je nutné dopředu domluvit s příslušným archivem.

Při lokálním exportu z účtu příslušného sociálního média je vhodné vybrat možnost exportu v HTML (+ CSV).

Postup dále dle kategorie Webové stránky.

### Technické výkresy a modely (CAD, BIM atd.)

Zahrnuje formáty 2D a 3D zobrazení využívané při tvorbě a výměně dat v oblasti grafických, geometrických, matematických a inženýrských aplikací.

Výstup pro výběr dokumentů za archivácie a jejich uchovávání v digitálním archivu.

formát	verze	PUID	kódování	standard	výstup eSSL	archive
PDF/A	1-4	fmt/95; fmt/354; fmt/476; fmt/477; fmt/478; fmt/479; fmt/480; fmt/481		ISO 19005	výstupní formát	Ano
AutoCAD Drawing (DWG)	2018/2019/2020/2021	fmt/1395			ponechat	Ano



Industry Foundation Classes (IFC)			2x2	ISO 16739-1	vizualizace ve VDF+ originál	vizualizace ve VDF + originál
Extensible 3D (X3D)				ISO/IEC 19775-1	vizualizace ve VDF + originál	vizualizace ve VDF + originál

### Textové dokumenty

Kategorie zahrnuje formáty reprezentující soubory textové a kombinované textové a obrazové dokumenty. Jedná se zpravidla o vyjádření ve strukturované nebo nestrukturované podobě, ve značkovacím jazyce nebo např. reprezentace zdrojových kódů. U posledních dvou je vhodné vedle převedu předat i původní soubor.

Výstup pro výběr dokumentů za archiválie a jejich uchování v digitálním archivu.

formát	verze	PUID	kódování	standard	výstup eSSL	archivace
PDF/A	1-4	fmt/95; fmt/354; fmt/476; fmt/477; fmt/478; fmt/479; fmt/480; fmt/481		If/aSO 19005	výstupní formát (přílohy dle formátu)	Ano
ASCII (TXT)		x- fmt/111			ponechat	Ano

### Účetní záznamy

Kategorie zahrnuje účetní záznamy v elektronické podobě.

Výstup pro výběr dokumentů za archiválie a jejich uchování v digitálním archivu.

formát	verze	PUID	kódování	standard	výstup eSSL	archivace
ISDOC Information System Document (ISDOC)	5.2-	fmt/1587, fmt/1570			výstupní formát	Ano

### Video

Formáty pro zaznamenávání, kopírování, přehrávání, vysílání a zobrazování pohyblivého obrazu.

Výstup pro výběr dokumentů za archiválie a jejich uchování v digitálním archivu.

formát	verze	PUID	kódování	standard	výstup eSSL	archivace
Graphics Interchange Format (GIF)	89a	fmt/4			výstupní formát	Ano
MPEG 1 Video (MPEG, MPG)		x- fmt/385		ISO/IEC 11172	výstupní formát	Ano
MPEG 2 Video (MPEG, MPG)		x- fmt/386		ISO/IEC 13818	výstupní formát	Ano
MPEG 4 Video (MPEG, MPG)		fmt/199		ISO/IEC 14496	výstupní formát	Ano
Audio Video Interleaved Format (AVI)		fmt/5	nekomprimovaný 4:2:2	Multimedia Programming Interface and Data Specifications 1.0	§ 23 odst. 4	Ano
Material Exchange Format (MXF)	1-3	fmt/200; fmt/783; fmt/784; fmt/785; fmt/786; fmt/787; fmt/788; fmt/789; fmt/790	MPEG-2 nebo J2K beztrátově komprimovaný –	ST 377-1:2011	§ 23 odst. 4	Ano