

Formátová politika pro komplexní digitální objekty

Zbyšek Stodůlka

Archivy, knihovny, muzea v digitálním světě 2023



Národní archiv



Existuje nepřijatelný formát?

Pokračující diskuse v komunitě digitální archivace:

- 02/2023 – webinář OPF (moderace Sam Alloing)

Paul Wheatley, Valentijn Gilissen

- 09/2023 - iPRES 2023 (moderace Sam Alloing)

Valentijn Gilissen, Tyler Thorsted,

Leslie Johnston, Mickey Lindlar, Kate Murray



Existuje nepřijatelný formát? (stanoviska)

Zastarávání formátu nepředstavuje velké riziko. Plošné přístupy k předem promyšlené migraci velkého množství formátů bez identifikovaného rizika uchování nejsou jen o riziku poškození dat. Možná by bylo lepší nasměrovat naše omezené zdroje jinam.

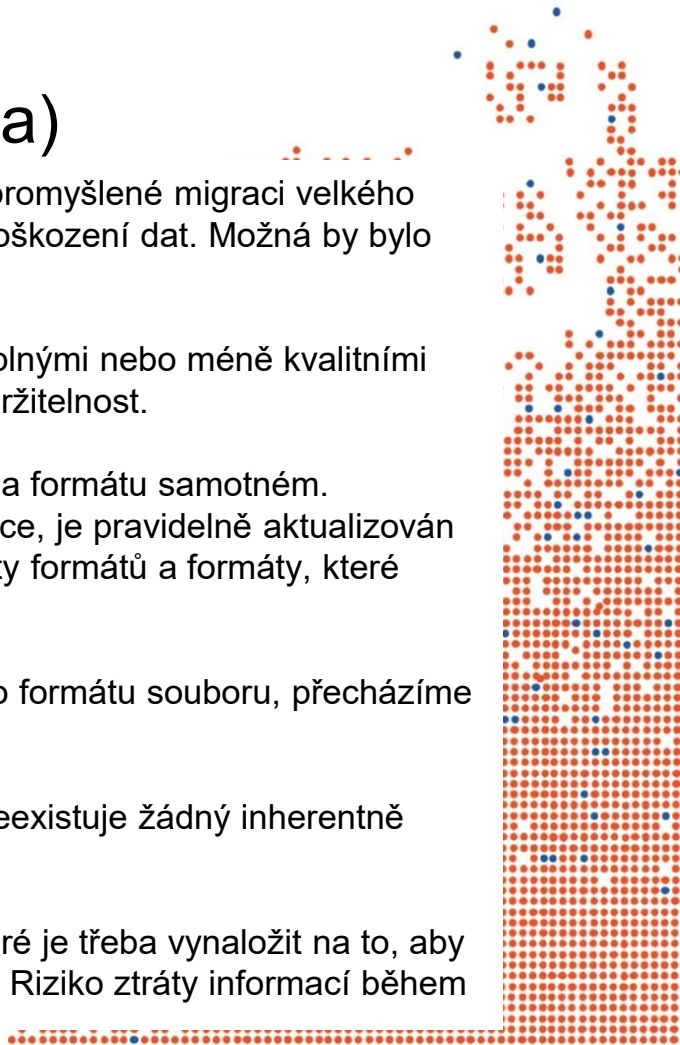
Politika s „nepřijatelnými“ formáty riskuje, že na konci bude disponovat neúplnými nebo méně kvalitními daty. Politika s „přijatelnými formáty“ riskuje nedostatek jakéhokoli úsilí o udržitelnost.

Uchovávání by mělo být založeno na vlastnostech formátu souboru, nikoli na formátu samotném. Máme seznam preferovaných a přijatelných formátů jako vodítko pro původce, je pravidelně aktualizován na základě jejich podnětů... ale není proskriptivní a archivy akceptují varianty formátů a formáty, které nejsou na jeho seznamu (převáděné dokumenty mohou být starší 25 let).

Upřednostňovaná formátová politika způsobila problémy s migrací špatného formátu souboru, přecházíme na úroveň znalostí (kolik informací o formátu známe).

Pro hodnocení formátu je nezbytný kontext, který vychází z myšlenku, že neexistuje žádný inherentně dobrý nebo špatný formát. „Záleží“ je legitimní odpověď’.

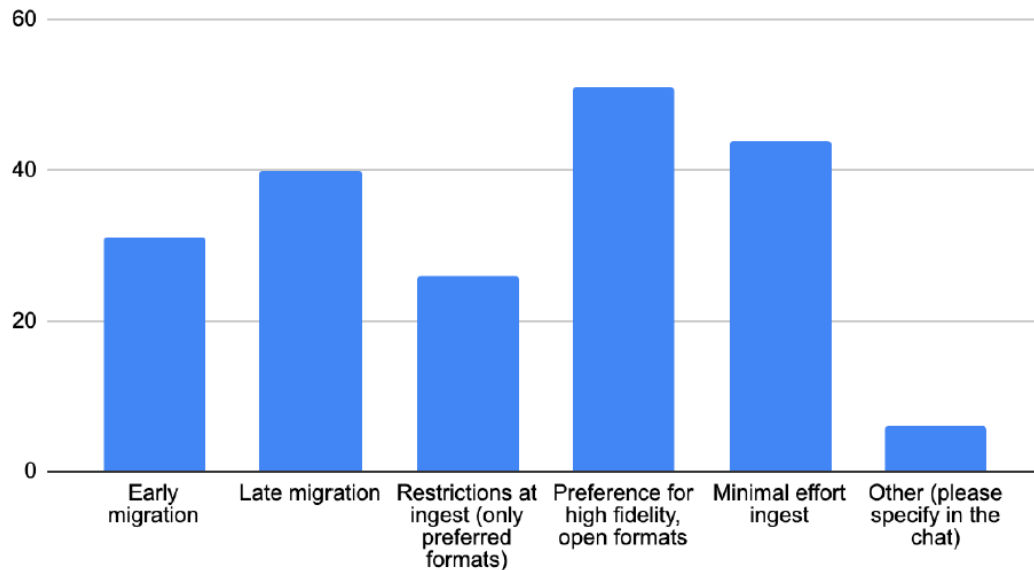
Zastarávání je skutečné, ale vždy subjektivní a relativní... popisuje úsilí, které je třeba vynaložit na to, aby byla data interpretovatelná a to různými kanály a pro různé scénáře použití. Riziko ztráty informací během migrace je příliš velké.



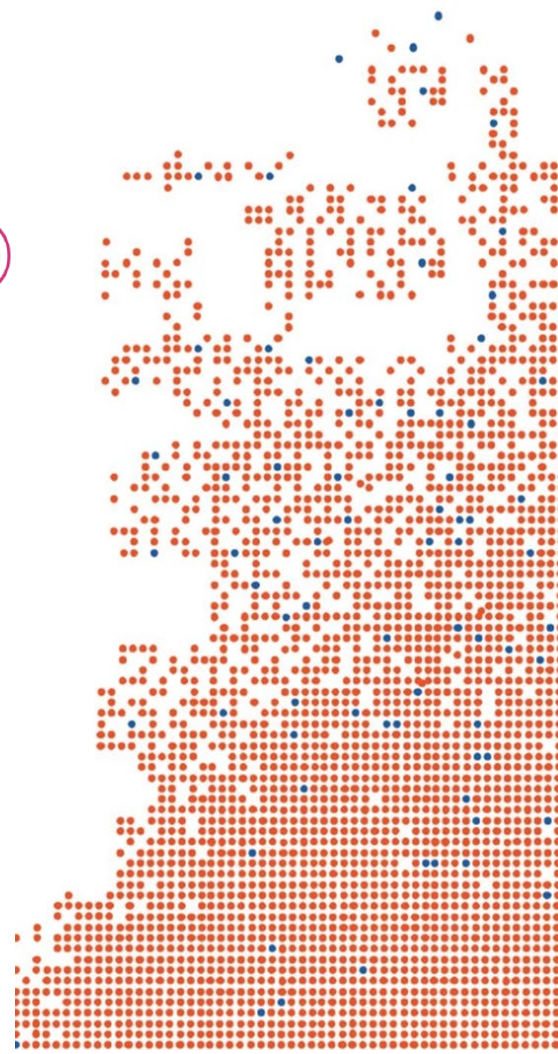
Existuje nepřijatelný formát? (závěry)

- archivy by měly být otevřené, komunikovat a využívat know-how původců a expertních komunit
- zachovat původní digitální objekt (také pro transfer a uchování v archivu)
- formát pro transfer nemusí být formátem pro uchování
- při migraci by měla být prováděna validace formátu a testování jeho interpretovatelnosti (ve snaze vyhnout se lidské/strojové chybě)
- je nutný stroj času k nastavení správné politiky uchování pro původce / producenty dat

Průzkum



Which preservation strategies are you using?



Formátová politika (archivnictví)

2009/2012, 2015, 2019, 2022

Zákon č. 499/2004, o archivnictví a spisové službě
(+ vyhláška 259/2012, o podrobnostech výkonu spisové služby včetně předešlé)

- Statický textový a kombinovaný textový a obrazový dokument – PDF/A
 - Statický obraz (PNG, TIF/TIFF, JPEG)
 - Dynamický obraz (MPEG-1, MPEG-2, MPEG-4, GIF)
 - Audio (MP2, MP3, WAV [PCM])
 - Databáze/datové věty (XML, DTD)
 - Účetní záznamy (ISDOC 5.2+, jiný formát v souladu se syntaxí dle 2014/55/EU)
 - Metadata (eSSL) – XML (SIP)
- faktická nemožnost modernizace
- vedle podpory středně- a dlouhodobého uložení slouží k omezení množiny zpracovávaných formátů při příjmu dokumentů původci a současně poskytovaných původci občanům (přístupnost, otevřenost atd.)

https://archi.gov.cz/znalostni_baze:archivni_formaty

2023

Národní standard formátů pro archivaci

- schválen Radou vlády pro informační společnost 9. 6. 2023
- součást znalostní báze architektury e-Governmentu
- umožňuje flexibilitu digitálním archivům (risk management)
- nejen formáty, ale i typy balíčků – např. GeoSIP, SIARD aj.

Audio	Obrázek (rasterový)
Binární soubory (spustitelné)	Obrázek (vektorový)
Databáze a strukturovaná data	Prostorová data (GIS)
E-mailý	Sociální sítě (médiá)
Fonty	Technické výkresy a modely (CAD, BIM aj.)
Internet a intranet (web)	Text
Komprimace dat	Účetní záznamy
Kontejner	Audiovizuální a kinematografické dokumenty
Obraz disku	

Zahraniční inspirace formátových politik (Kongresová knihovna)

Sustainability of Digital Formats: Planning for Library of Congress Collections

[Introduction](#) | [Sustainability Factors](#) | [Content Categories](#) | [Format Descriptions](#) | [Contact](#)

Sustainability Factors

Table of Contents

- [Disclosure](#)
- [Adoption](#)
- [Transparency](#)
- [Self-documentation](#)
- [External dependencies](#)
- [Impact of patents](#)
- [Technical protection mechanisms](#)

Overview of factors

In considering the suitability of particular digital formats for the purpose of archiving, it is important to articulate important factors that affect choices. The seven sustainability factors influence the likely feasibility and cost of preserving the information content that institutions operate. They are significant whatever strategy is adopted for archiving on future computers, or a hybrid approach.

Additional factors will come into play relating to the ability to represent digital objects expected by future users. These factors will vary by genre or form of digital objects, whether digital or not, and not all digital formats for images are

site devoted to particular [Content Categories](#).

Disclosure

Disclosure refers to the degree to which complete specifications and tools for validating technical integrity exist and are available over the long term is not feasible without an understanding of how the information is represented (encoded) as bits.

A spectrum of disclosure levels can be observed for digital formats. Non-proprietary, open standards are usually more fully disclosed. However, what is most significant for this sustainability factor is not approval by a recognized standards body, but the existence of tools from various sources is valuable in its own right and as evidence that specifications are adequate. The existence and adoption of a format, as indicated below. In the future, deposit of full documentation in escrow with a trusted archive will be a format for which documentation is not publicly available. Availability, or deposit in escrow, of source code for associated

[Back to top](#)

Adoption

Adoption refers to the degree to which the format is already used by the primary creators, disseminators, or users of information. Interchange between systems. If a format is widely adopted, it is less likely to become obsolete rapidly, and tools for migrating information between institutions.

Evidence of wide adoption of a digital format includes bundling of tools with personal computers, native support in Web browsers, the existence of many competing products for creation, manipulation, or rendering of digital objects in the format. In some license terms include royalties based on content usage. A format that has been reviewed by other archival institutions and

[Back to top](#)

Transparency

Transparency refers to the degree to which the digital representation is open to direct analysis with basic tools, including how the information is represented simply and directly will be easier to migrate to new formats and more susceptible to digital archaeology, development of a "universal virtual computer" concept proposed by Raymond Lorie will be simpler.¹

Transparency is enhanced if textual content (including metadata embedded in files for non-text content) is encoded in a standard order. For preserving software programs, source code is much more transparent than compiled code. For non-textual information, a format that has been reviewed by other archival institutions and

[Digital Preservation Home](#)

Search this site

 LIBRARY OF CONGRESS

[ASK A LIBRARIAN](#)

[DIGITAL COLLECTIONS](#)

[LIBRARY CATALOGS](#)

Search

[GO](#)

The Library of Congress > [Preservation](#) > [Resources](#) > [Recommended Formats Statement](#)

PRESERVATION



- [Preservation Home](#)
- [About](#)
- [Collections Care](#)
- [Conservation](#)
- [Digital Preservation](#)
- [Emergency Management](#)
- [En Español](#)
- [FAQ](#)
- [Preservation Science](#)
- [Resources](#)
- [Outreach & Training Opportunities](#)
- [Have a preservation question? Ask-a-Librarian](#)

Related Links

- [Donate](#)
- [Blog: Guardians of Memory. Preserving the National Collection](#)
- [Audio-Visual Preservation Board](#)
- [National Film Preservation Board](#)

Recommended Formats Statement

[Print](#) [Subscribe](#) [Share/Save](#) [Give Feedback](#)

[Resources Home](#) | [Educational](#) | [Research Reports](#) | [Supply Specifications](#) | [Recommended Formats Statement](#) | [Staff Publications](#)

[Main](#) | [Table of Contents](#) | [Introduction](#) | [Summary of Digital Format Preferences](#) | [Textual Works](#) | [Still Image Works](#) | [Moving Image Works](#) | [Audio Works](#) | [Musical Scores](#) | [Datasets](#) | [GIS, Geospatial and Non-GIS Cartographic](#) | [Design and 3D](#) | [Software and Video Games](#) | [Web Archives](#) | [Email](#)

Library of Congress Recommended Formats Statement 2023-2024

Recommended Formats Statement identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.

[2023-2024 Recommended Formats Statement](#) (HTML)

[2023-2024 Recommended Formats Statement](#) (PDF)

[2023-2024 Summary of Digital Format Preferences](#) (HTML)

[Change log from 2022-2023 to 2023-2024 version](#) (PDF)

Evaluation matrix template

This worksheet details how the Library of Congress determines if a format is "preferred" or "acceptable" based on both global/community criteria and local/institutional criteria. For more information, see [Introduction](#).

[Format evaluation template with sample data - 2023 with revisions](#) (XLSX)

Identifikace formátů (aktualizace registru PRONOM)

← → ↻ 🌐 nationalarchives.gov.uk/aboutapps/pronom/release-notes.xml

Updated Records

Portable Document Format

[fmt/95](#): **Acrobat PDF/A - Portable Document Format** (1a). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

[fmt/116](#): Windows Bitmap (Signature 3.0). Added the possibility of '10' at byte 28. Submitted by The National Library of Australia.

[fmt/208](#): Binary File. Added description, changed orientation to binary, added BIN file format and updated relationships. Submitted by The National Archives and Records Administration.

[fmt/286](#): HDF5 (1). Extension .nc added. Submitted by ETH-Bibliothek, ETH Zurich.

[fmt/287](#): HDF5 (2). Extension .nc added. Submitted by ETH-Bibliothek, ETH Zurich.

[fmt/354](#): **Acrobat PDF/A - Portable Document Format** (1b). Extension to byte gap between part and conformance tags added to both signatures. Submitted by The National Archives of the Czech Republic.

[fmt/372](#): Earth Resource Satellite Image Header Format. Added an offset of 3 bytes from EOF, at the request of the enquirer. Submitted by The National Library of Australia.

[fmt/468](#): ISO Disk Image. Broadened the BOF offset from 32769 to a range (32700- 37700). Submitted by The National Library of Australia.

[fmt/476](#): **Acrobat PDF/A - Portable Document Format** (2a). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

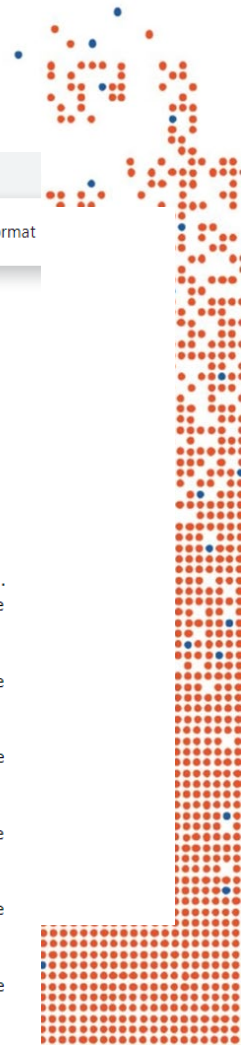
[fmt/477](#): **Acrobat PDF/A - Portable Document Format** (2b). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

[fmt/478](#): **Acrobat PDF/A - Portable Document Format** (2u). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

[fmt/479](#): **Acrobat PDF/A - Portable Document Format** (3a). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

[fmt/480](#): **Acrobat PDF/A - Portable Document Format** (3b). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.

[fmt/481](#): **Acrobat PDF/A - Portable Document Format** (3u). Additional signature added to allow for PDF conformance and PDF part tags to be alternated. Extension to BOF offset and 0-120 byte gap between part and conformance tags added to original signature. Submitted by The National Archives of the Czech Republic.



Prokazování totožnosti s využitím fyzických dokladů totožnosti

Práce s klientem veřejné správy na portálu

PŘÍSPĚVEK NA VÝKON STÁTNÍ SPRÁVY PRO OBCE

Předávání obsahu datového fondu

Příručka pro řízení organizačních změn roku 2020 (OCM)

Příručka plánování digitalizace služeb veřejné správy ČR

Přístup do Centrálního místa služeb pro podřízené/zřizované organizace

Přístup do Extranetu vytvořeného v CMS

RACI matice osob zodpovědných za vyplnění formuláře žádosti o stanovisko OHA

Rizika a jejich řízení

Seznam záměrů Digitálního Česka

Splnění zákonných povinností orgánů veřejné správy podle zákona o informačních systémech veřejné správy

Stavební blok

Stručná metodika efektivního modelování

Stručný návod na aktualizaci ohlášení využití veřejných údajů

Stručný návod na aktualizaci oprávnění na údaje

Stručný návod na ohlášení nové agentury

Stručný návod na přihlášení do AIS RPP Působnostní

Státní právnická osoba

Tabulky pro uvedení IK do souladu s IKČR

Technická specifikace platební brány

The Sharing and Reuse Framework for IT Solutions

Total cost of ownership (Celkové náklady vlastnictví)

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 systémů. 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 (XML Schema Definition/XSD by mělo obsahovat popis elementů)
Document Type Definition (DTD)		x-fmt/315			výstupní formát	ano
SIARD	2.2	fmt/1777		DILCIS Board/Swiss Federal Archives https://dilcis.eu/content-types/siard	individuální posouzení digitálním archivem/NDA	individuální posouzení digitálním archivem/NDA
Comma Separated Values (CSV)		x-fmt/18			§ 23 odst. 6 + originál /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			§ 23 odst. 2/ § 23 odst. 6 + originál	individuální posouzení digitálním archivem/NDA
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



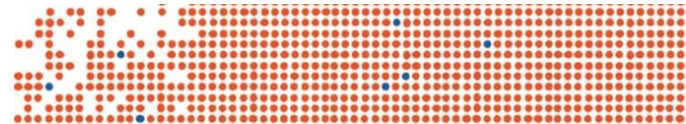


Technické výkresy a modely (CAD, BIM atd.)

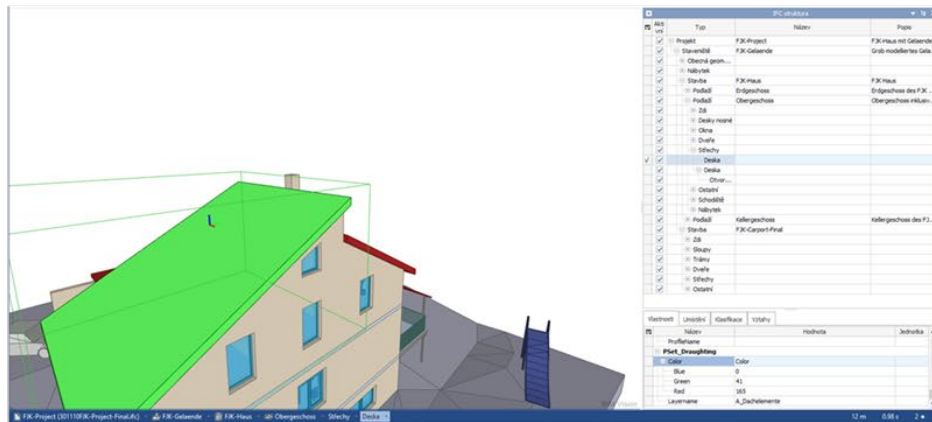
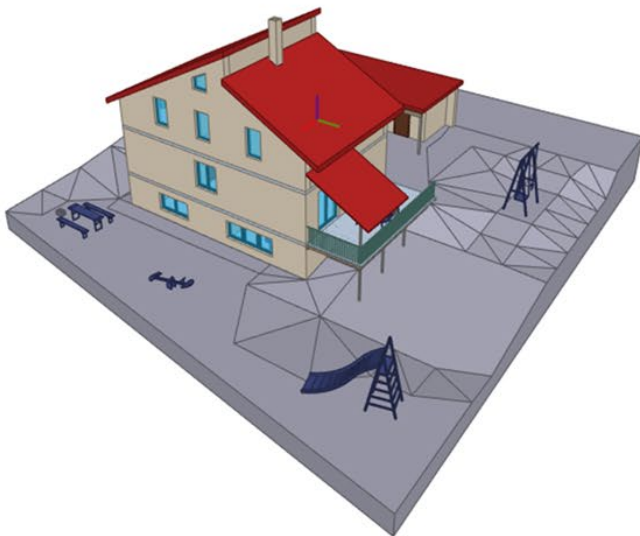
Zahrnuje formáty 2D a 3D zobrazení využívané při tvorbě a výměně dat v oblasti grafických, geometrických, matematických a inženýrských aplikací.

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

formát	verze	PUID	kódování	standard	výstup eSSL	archivace
Portable Document Format – PDF/A (PDF)	1-4	fmt/95, fmt/354, fmt/476, fmt/477, fmt/478, fmt/479, fmt/480, fmt/481		ISO 19005	výstupní formát + originál	ano
AutoCAD Drawing (DWG)	2018/2019/2020/2021	fmt/1395			ponechat	ano
Industry Foundation Classes (IFC)	2x2/2x3/ 4	fmt/659, fmt/699, fmt/700		 ISO 16739-1	vizualizace ve VDF+ originál	vizualizace ve VDF + originál
Extensible 3D (X3D)	3.0-3.3	fmt/579, fmt/580, fmt/581, fmt/582		ISO/IEC 19775, ISO/IEC 19776, ISO/IEC 1977	vizualizace ve VDF + originál	vizualizace ve VDF + originál



Přístup (IFC - zobrazení rozdílným software, vhodná doprovodná vizualizace)



```
#104707=IFCSLAB('3d1SXOdInFBAQvPV0xKzV5',#16,$,$,$,#104690,#104706,$,.ROOF,);
```

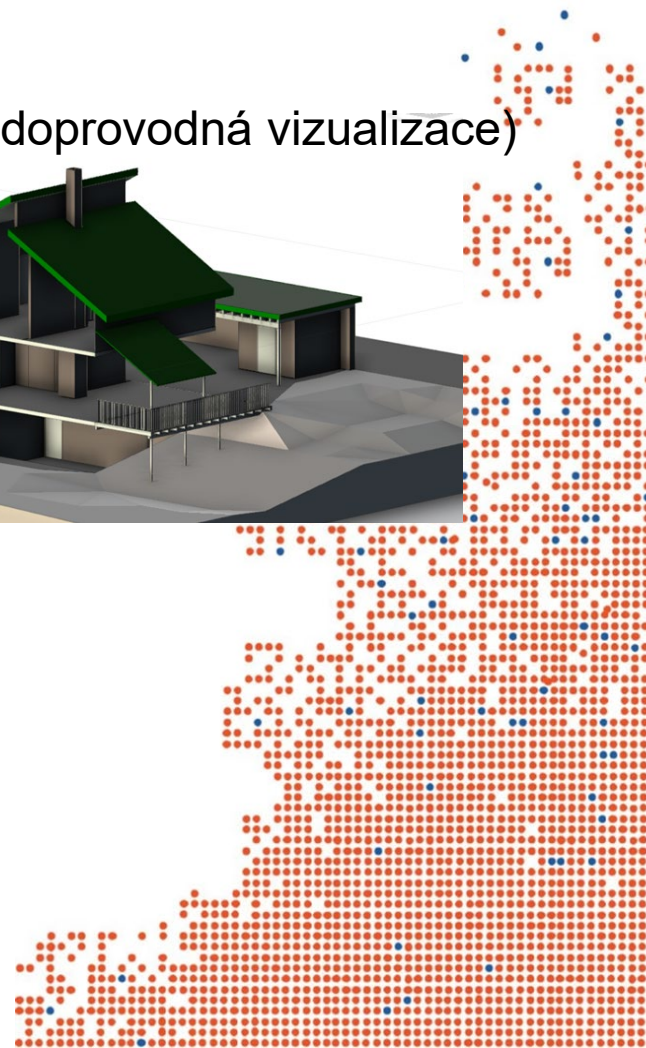
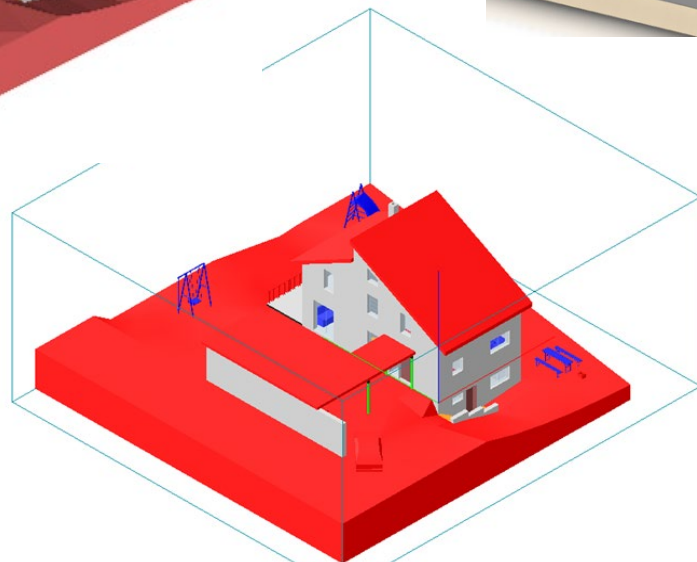
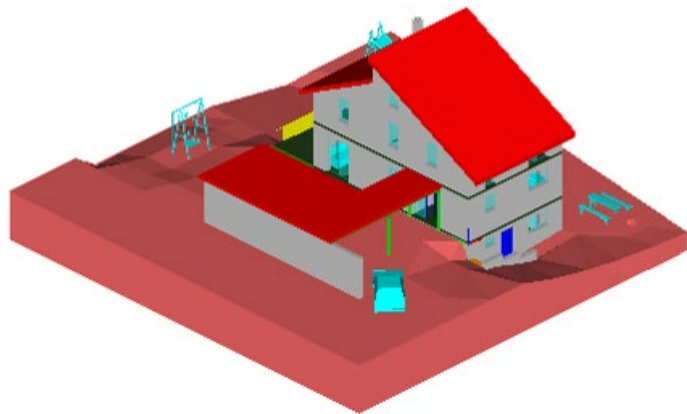
```
#104711=IFCPROPERTYSINGLEVALUE('Layername',$,IFCLABEL('A_Dachelemente'),$);
```

```
#104712=IFCPROPERTYSINGLEVALUE('Red',$,IFCINTEGER(165),$);
```

```
#104713=IFCPROPERTYSINGLEVALUE('Green',$,IFCINTEGER(41),$);
```

```
#104714=IFCPROPERTYSINGLEVALUE('Blue',$,IFCINTEGER(0),$);
```

Přístup (IFC - zobrazení rozdílným software, vhodná doprovodná vizualizace)





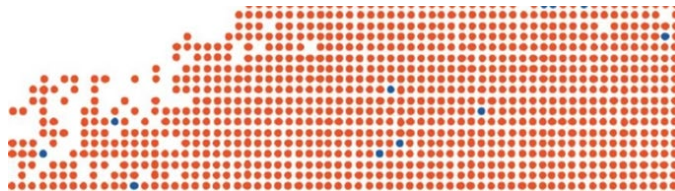
Prostorová data (GIS)

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-trvalého-ulozeni-v-digitalnim-archivu.aspx>).

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

formát	verze	PUID	kódování	standard	výstup eSSL	archlvac
Geographic Tagged Image File Format GeoTIFF (TIF, TIFF)	1.0	fmt/155		http://geotiff.maptools.org/spec/geotiffhome.html	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/1700		OGC standard http://www.opengis.net/doc/IS/geopackage/1.1	GML + originál/ponechat (vhodné připojit náhled ve VDF)	ano
GeoJSON (GEOJSON)	1.0	fmt/1367		RFC 7946 https://tools.ietf.org/html/rfc7946	ponechat (vhodné připojit náhled ve VDF)	ano
LiDAR Point Cloud (LAS)	1.2 -1.4	fmt/370 (pro 1.2)		LAS Specification 1.4, http://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf	ponechat (vhodné připojit náhled ve VDF)	ano



Emulace – např. EaaS (Emulation as a Service Infrastructure)

EaaS Demo UI

localhost/admin/#/admin/create

Create New Machine

Choose System

Windows Linux Apple Other

General

Machine Name

01VMWIN98

System Settings

Select operating system / System

search or choose a system...

- Windows XP (32bit)
- Windows XP (64bit)
- Windows 95
- Windows 98**
- Windows 2000
- Windows Server 2008
- Windows Vista
- Windows Server 2003

Cancel Save

Build: 57407D816F
UI-Build: 8610097D81

Závěr

- standard pomůže dále kultivovat prostředí (původci, dodavatelé, digitální archivy)
- standard zahrnuje nejen formáty, ale i variantní datové struktury pro předání (např. GeoSIP, SIARD atd.)
- pravidelná aktualizace - změny v PRONOM, nové typy datových balíčků (např. v přípravě E-ARK SIP pro 3D objekty atd.)



Díky za pozornost
zbysek.stodulka@nacr.cz

