

Technické standardy pro datové sady na stupni otevřenosti 3

Tato stránka byla nahrazena [Otevřenou formální normou pro Prostorová data](#).

Pro dosažení stupně otevřenosti datové sady 3 je nutno zajistit, aby kompletní obsah datové sady byl dostupný prostřednictvím 1 či více distribucí datové sady, které splňují následující standardy.

Obecným standardem datových sad na stupni 3 je, že každá distribuce datové sady musí být kódována v kódování UTF-8.

Standardy pro tabulkové datové sady

Distribuce datové sady, která reprezentuje tabulku, musí splňovat následující standardy:

- V jedné distribuci je reprezentována právě jedna tabulka.
- V distribuci neexistuje žádný jiný záznam než je řádek tabulky nebo její hlavička.
- Všechny záznamy v distribuci, tj. jednotlivé řádky tabulky, mají stejnou strukturu, která odpovídá hlavičce.
- Celý obsah datové sady je vyjádřen v distribucích v alespoň jednom z následujících formátů:
 - [Comma Separated Values \(CSV\)](#),
 - eXtensible Markup Language (XML, <http://www.w3.org/TR/xml/>), kde jednotlivé řádky tabulky jsou zapsány jako samostatné XML elementy strukturované pomocí zanořených XML elementů a XML atributů reflektujících hlavičku reprezentované tabulky, nebo
 - JavaScript Object Notation (JSON, <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>), kde jednotlivé řádky tabulky jsou reprezentovány jako samostatné JSON objekty strukturovaných do dle hlavičky reprezentované tabulky.

Standardy pro datové sady s hierarchickou strukturou

Distribuce datové sady, která obsahuje hierarchické strukturované objekty (tj. objekty složené z jiných objektů), které není vhodné reprezentovat v podobě tabulky dle předchozí podkapitoly, musí splňovat následující standardy:

- V jedné distribuci je reprezentována kolekce stejného typu objektů
- Objekty v kolekci mohou obsahovat další vnořené typy objektů.
- Celý obsah datové sady je vyjádřen v distribucích v alespoň jednom z následujících formátů:
 - eXtensible Markup Language (XML, <http://www.w3.org/TR/xml/>), jehož struktura vyjádřená v podobě XML elementů a XML atributů umožňuje výběr jednotlivých údajů reprezentovaných v datové sadě pomocí prostředků jazyka XPath či CSS selektorů bez nutnosti parsování získaných atomických údajů.
 - JavaScript Object Notation (JSON, ECMA-404, nebo

<http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>), jehož struktura vyjádřená v podobě JSON objektů a primitivních hodnot umožňuje výběr jednotlivých údajů reprezentovaných v datové sadě pomocí prostředků programovacích jazyků používaných pro tvorbu WWW aplikací (JavaScript, Java, .apod.) bez nutnosti parsování získaných atomických údajů.

Standardy pro datové sady v podobě kolekce textových dokumentů

Distribuce datových sad, které mají charakter textových dokumentů, musí splňovat následující pravidla a standardy:

- V jedné distribuci je reprezentována kolekce dokumentů, kde každý dokument je jednoznačně identifikovatelný v podobě souboru a lze jej z kolekce získat pomocí běžných programovacích prostředků.
- Všechny dokumenty v datové sadě jsou vyjádřeny alespoň v jednom z následujících formátů:
 - Formát odpovídající MIME typu text/plain obsahující pouze znaky textového dokumentu.
 - Libovolný jiný formát pro textové dokumenty, který je strojově čitelný, má otevřenou specifikaci a existuje volně dostupná knihovna, umožňující automatizovanou strojovou extrakci textového obsahu dokumentů v datové sadě (až do úrovně jednotlivých znaků).
 - Např. HTML, DOCX, ODT, DocBook či TeX

Standardy pro datové sady v podobě kolekce obrázků

Distribuce datových sad, které mají charakter obrázků, musí splňovat následující pravidla a standardy:

- V jedné distribuci je reprezentována kolekce obrázků, kde každý dokument je jednoznačně identifikovatelný v podobě souboru a lze jej z kolekce získat pomocí běžných programovacích prostředků.
- Každý obrázek je vyjádřen ve formátu, který je strojově čitelný a má otevřenou specifikaci.
 - V případě bitmapových obrázků musí formát umožnit přistupovat pomocí běžných programovacích prostředků k jednotlivým pixelům tvořícím obrázek.
 - V případě vektorových obrázků musí formát umožnit přistupovat pomocí běžných programovacích prostředků k jednotlivým vektorovým objektům tvořícím obrázek.

Standardy pro prostorové datové sady

Podle velikosti obcí a organizací - resp. přítomnosti oddělení GIS v rámci jejich organizační struktury - mohou být použity postupy a datové formáty dostupné v jejich IT infrastruktuře.

U malých obcí či organizací je celkem zbytečné budovat komplexní prostorovou infrastrukturu (Spatial Data Infrastructure - SDI). V případě, že obec či organizace nějakým způsobem spravuje data s prostorovou složkou (tj. geografická data, geodata), ať už bodová (jako např. seznam památných stromů nebo pomníků), liniová (lokální cestní síť) nebo plošná (přehled obecní zeleně) a chce tyto datové sady otevřít, měla by k tomu zvolit vhodný souborový formát:

- **Comma Separated Values** (CSV, IETF RFC4180) - jednotlivé záznamy atributové tabulky jsou zapsány jako řádky textového souboru a hodnoty záznamů jako řetězce oddělené čárkou (znak

UTF-8 s kódem U+002C), přičemž hodnota může být uzavřena v uvozovkách (znak UTF-8 s kódem U+0022), především pokud hodnota sama obsahuje jednu nebo více čárek. Položky s prostorovou informací mohou být vyjádřeny souřadnicemi, pokud se jedná o bodová data. U ostatních typů dat jako jsou liniová anebo polygonová data může být použit zápis WKT (viz *Poznámka k zápisu geometrie*). Dále by měl být tento zápis opatřen informací o souřadnicovém systému ve formě EPSG kódu (viz *Poznámka k souřadnicovým systémům*), např. v hlavičce souboru.

- [ESRI Shapefile](#) - vlastnosti jednotlivých prvků dodržují strukturu pro datovou sadu.
- [Geography Markup Language](#) (OGC GML) - jedná se formát XML pro prostorová data. Atributy prvků reflektují strukturu datové sady.
- [GeoJSON](#) - formát, kde jsou jednotlivé záznamy atributové tabulky reprezentovány jako samostatné JSON objekty strukturované dle hlavičky reprezentované tabulky a navíc obsahují prostorovou informaci.
- [OGC GeoPackage](#) - je nad souborovou databází [SQLite](#) postavený standardizovaný formát pro prostorová data. Lze do něj uložit stejnou strukturu jako do CSV a složitou geometrickou strukturu jako samostatný atribut.

Datové sady jsou publikovány tak, že všechny prvky (prostorové objekty) v datové sadě mají stejnou strukturu (stejně atributy).

Pokud jsou některá data organizací nebo obcí poskytována externím subjektem - např. firmou zpracovávající územní plán - měli by jednatelé trvat na odevzdání práce v *otevřeném formátu pro prostorová data* (viz seznam níže). Takovou datovou sadu lze pak bez dalšího úsilí otevřít. Podstatné je v této souvislosti se pokud možno vyhnout uzavřeným formátům (např. DWG, DGN, ESRI GeoDatabase, ...). Případně datové sady v těchto formátech převést před otevřením do výše doporučených otevřených formátů.

Je také vhodné udržovat systém metadat k datovým sadám (jejich slovní popis, označení původu, datum vzniku, licenci a tak dále) např. v samostatném souboru. Datové sady lze zveřejnit „tak jak jsou“ (pokud možno v otevřeném formátu, jak je zmíněno výše) na webových stránkách ve formě odkazů na soubory. K tomu není ani nutné pořizovat často nákladné webové mapové portály.

V případě publikace rozsáhlejší datové sady formou souborů je vhodné doplnit tyto soubory o [Atom feed](#) nebo je publikovat prostřednictvím webových služeb OGC. Tyto způsoby publikace umožňují jednodušší vyhledávání konkrétních dat v rámci datové sady.

U obcí s rozšířenou působností (ORP) a větších, je možné publikovat data v souladu se směrnicí INSPIRE (zákon [č. 380/2009 Sb.](#) <http://inspire.gov.cz/>) pomocí webových služeb [OGC](#):

- [OGC WFS](#) (Web Feature Service)
- [OGC WCS](#) (Web Coverage Service)

Data poskytovaná těmito službami lze použít pro další GIS analýzy, neboť se jedná o geodata. Pro vektorová data (passporty zeleně, silniční síť, mapu veřejných záchodků apod.) je vhodné použití OGC Web Feature Service, pro rastrová data (např. letecké snímky) potom OGC Web Coverage Service.

Dále lze doporučit i další OGC webové služby [WMS](#) (Web Map Service) a [WMTS](#) (Web Map Tile Service). *Nejedná se však o otevřený způsob distribuce dat*, neboť tyto služby neposkytují geodata, ale pouze jejich otisk ve formě obrázku. Takto poskytovaná data nelze dále analyzovat prostřednictvím GIS nástrojů. Služby OGC WMS a WMTS jsou vhodné pro tvorbu webových mapových aplikací a pro využití desktopovými GIS programy jako podkladové mapy.

Při použití webových služeb OGC je třeba dodržet vhodné nastavení serveru s ohledem na podporované souřadnicové systémy, na jeho očekávanou vytíženost či bezpečnostní rizika apod.

Větší organizace veřejné správy s pokročilou infrastrukturou pro prostorová data (SDI) publikující metadata k prostorovým datovým sadám v souladu se standardem [OGC CSW](#) propojí tato metadata do portálu otevřených dat.

Poznámka k proprietárním službám

Některé nástroje GIS obsahují proprietární služby pro publikaci dat. Data publikovaná prostřednictvím těchto služeb nelze považovat za dostatečně otevřená. Nebývají podporována existujícími softwary třetích stran, licenční podmínky jejich využití nejsou často transparentní. Jedná se například o tzv. REST API v softwarech firmy ESRI.

Poznámka k souřadnicovým systémům

EPSG - je uznávaný registr mimo jiné pro popis souřadnicových systémů, podporovaný všemi programy pro práci s prostorovými daty (GIS). Relevantní hodnoty pro Českou republiku jsou

- 5514 - Systém jednotné trigonometrické sítě katastrální (S-JTSK) - používaný např. v katastru nemovitostí. Pro tento systém je mimo jiné v GIS typické, že souřadnice nabývají negativních hodnot (jsou uvozeny znaménkem minus).
- 4326 - World Geodetic System 1984 (WGS84), někdy zvaný též "souřadnice GPS". Souřadnice bývají uvedeny ve stupních s desetinným číslem. Některé publikační formáty (např. GPX nebo KML) počítají přímo s tím, že data jsou uložena v tomto souřadnicovém systému.
- 3857 - tzv. Web (nebo též Google) Mercator - je systém používaný na většině globálních webových aplikací. Jeho nevýhodou je mimo jiné zkreslení významné velikosti ploch směrem k pólům.

Další lze dohledat např. na stránce <http://epsg.io>

Poznámka k zápisu geometrie

WKT - Well Known Text - je standardní formát zápisu vektorové geometrie. Popis je mimo jiné k nalezení ve specifikaci OGC [Simple Features](#).

Typ	Příklad
Bod	POINT (30 10)
Linie	LINestring (30 10, 10 30, 40 40)
Polygon	POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))

Pomocí WKT lze samozřejmě uložit i komplikovanější datové typy, jako jsou MULTIPOINTS či CompoundCurve a další.

Pokud uvažujete o použití WKT pro popis atributu geometrie u objektů, zvažte publikaci otevřené datové sady v některém k tomu určeném otevřeném formátu (ESRI Shapefile, KML, OGC GeoPackage nebo GeoJSON).

Problematikou otevírání prostorových dat se podrobně zabývá publikace [Otevírání prostorových dat - případová studie](#).

Standardy pro ostatní typy datových sad

Jiný formát či jazyk může poskytovatel dat použít pouze v případě datové sady, jejíž charakter neumožňuje použití některého z formátů uvedených v předchozích podkapitolách nebo datová sada spadá do oblasti, ve které jsou běžně používány jiné formáty (které splňují podmínku strojové čitelnosti a otevřenosti specifikace).

Ať je zvolen jakýkoliv formát, musí být zajištěno, že celý obsah datové sady je vyjádřen v distribucích v alespoň jednom formátu, který splňuje podmínku strojové čitelnosti a otevřenosti specifikace a je v dané oblasti běžně používán.

From:

<https://opendata.gov.cz/> - **Otevřená data**

Permanent link:

<https://opendata.gov.cz/standardy:technicke-standardy-pro-datove-sady-na-stupni-3> 

Last update: **2019/10/04 12:53**