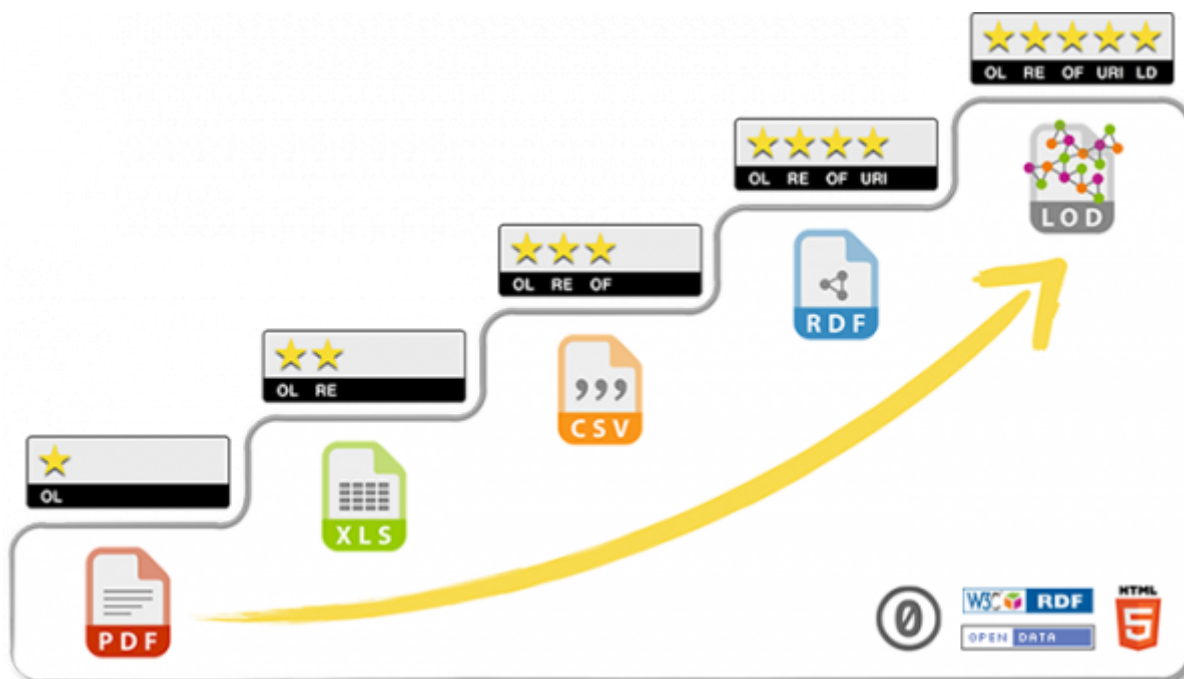


Stupně otevřenosti otevřených dat a česká legislativa

Míru otevřenosti dat je možné vyjádřit pomocí 5 stupňů otevřenosti, které jsou znázorněny na následujícím obrázku (stupeň je vyjádřen počtem hvězdiček).



Převzato z <https://5stardata.info/cs>

- stupeň 1 - datová je dostupná v síti WWW s vhodnými podmínkami užití otevřených dat (viz [Jak stanovit podmínky užití datových sad?](#)),
- stupeň 2 - datová sada je poskytována ve strojově čitelném formátu, který umožňuje automatizované strojové zpracování,
- stupeň 3 - datová sada je poskytována v otevřeném formátu, tj. ve formátu s volně dostupnou specifikací,
- stupeň 4 - na identifikaci entit v datové sadě se používají IRI,
- stupeň 5 - datová sada splňuje standard propojených dat.

Datová sada je tvořena údaji, které spolu souvisí. Při jejím zveřejňování v síti WWW musí být distribuována zájemcům v podobě souboru ke stažení, které nazýváme **distribuce datové sady**. Distribucí datové sady může být více. Všechny distribuce jedné datové sady ale musí mít stejný obsah. Mohou se lišit pouze ve formátu. Je tedy např. možné poskytnout distribuci datové sady ve formátu XLS a jinou distribuci ve formátu CSV. První formát je vhodný pro analytiku, kteří chtějí s údaji pracovat ve svém tabulkovém editoru. Druhý formát je vhodný pro programátory, kteří chtějí údaje automatizovaně převádět do svojí databáze, nad kterou budují svojí softwarovou aplikaci.

Pokud je obsah datové sady příliš velký, je možné jej rozdělit do více datových sad. Každá pak bude mít svoje distribuce, které se liší pouze ve formátu. U datových sad, kde je důležité přesně informovat o provedených změnách, je doporučováno zveřejnit datovou s iniciálním obsahem a poté zveřejňovat datovou sadu se seznamem provedených změn (tj. jaké záznamy byly smazány a jaké byly vytvořeny či aktualizovány a jak). Dělení na menší datové sady ale není možné provádět tak, že jako jednotlivé datové sady poskytujeme jednotlivé záznamy. Při určování toho, co je datová sada, se vždy řídíme

pravidlem, že datová sada je taková sada údajů, kterou chtějí naši uživatelé získat jako jeden celek, tj. získat v jednom souboru ke stažení. Tento celek je možné z důvodů velikosti rozdělit, ale vždy na co nejmenší nutný počet částí.

Specifikace stupňů otevřenosti datových sad

Stupeň otevřenosti 1

Stupeň otevřenosti 1 vyžaduje, aby byly distribuce datové sady dostupné on-line a s jasným vymezením podmínek užití. Neklade však žádné požadavky na datové formáty, ve kterých jsou distribuce datových sad zveřejňovány. Proto tento stupeň není považován za dostatečný stupeň otevřenosti.

V případě publikace prostorových dat sem spadají mimo jiné i často používané webové služby [OGC WMS](#) (Web Mapping Service) a [OGC WMTS](#) (Web Mapping Tile Service). Služby nezpřístupňují vlastní data, ale pouze obrázky generované z těchto dat - ty tedy v žádném případě nelze považovat za strojově zpracovatelné (viz stupeň 2).

Stupeň otevřenosti 2

Stupeň otevřenosti 2 je charakteristický tím, že distribuce datové sady jsou poskytovány ve strojově čitelném formátu. Pojem strojově čitelnosti je v oblasti IS dobře známý. Distribuce datové sady je vždy vytvářena za účelem zaznamenání určité množiny údajů. Strojová čitelnost distribuce datové sady pak znamená, že je distribuce vyjádřena v takovém formátu, který umožňuje co nejsnazší přístup k jednotlivým zaznamenaným údajům pomocí běžných a volně dostupných programovacích prostředků (programovacích jazyků a knihoven).

Příkladem je datová sada, která má charakter tabulky. Distribuce musí být vyjádřena v takovém formátu, který umožňuje pomocí běžných programovacích prostředků získávat z tabulky jednotlivé řádky a obsah buněk v řádcích. Jiným příkladem je datová sada tvořená kolekcí textových dokumentů. Musí být zaznamenána ve formátu, který umožňuje pomocí běžných programovacích prostředků získávat jednotlivé dokumenty, jejich jednotlivé znaky a značky označující sémantiku jejich částí jako jsou nadpisy (různých úrovní), tabulky, seznamy, obrázky apod.

Co nejsnazší přístup k údajům znamená, že údaje reprezentované v distribuci datové sady je možné přímo číst pomocí programovacích instrukcí bez nutnosti jakéhokoliv předzpracování distribuce, které nesouvisí s čtením jednotlivých údajů. Předzpracováním distribuce, která nesouvisí se čtením jednotlivých údajů, může být např. nutnost:

- zpracovat distribuci pomocí OCR nástrojů:
 - např. kolekci dokumentů v podobě naskenovaných stránek není možné považovat za distribuci datové sady tvořenou kolekcí dokumentů na stupni otevřenosti 2,
 - podobně např. tabulku reprezentovanou v podobě obrázku není možné považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2 (ať se jedná o samostatný obrázek nebo je obrázek vložen do jiného souboru, např. do XLS či DOC souboru);
- vyhledat tabulku zanořenou v textu či mezi jinými objekty:

- např. dokument ve formátu HTML obsahující tabulku vyjádřenou v podobě HTML elementu `<table>` vnořenou v textu a mezi jinými HTML elementy nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2;
- vyhledat tabulku mezi jinými tabulkami:
 - např. tabulku v souboru ve formátu XLS či XLSX, který obsahuje více tabulek nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2,
- rekonstruovat obsah tabulky ze struktur, které nejsou primárně určeny pro reprezentaci tabulkových dat, ale pro vyjádření vizuálních objektů ve tvaru tabulek,
 - např. dokument ve formátu DOC či DOCX obsahující tabulku nelze považovat za distribuci datové sady s tabulkou na stupni otevřenosti 2.

Je-li to vhodné a účelné, lze distribuce na stupni otevřenosti 2 poskytovat v komprimovaném tvaru (např. jako ZIP archiv apod.). To připadá v úvahu především v případě velkých distribucí nebo v případě kolekcí textových dokumentů.

Protože stupeň otevřenosti 2 neklade žádné další požadavky na datové formáty, ve kterých jsou distribuce datových sad zveřejňovány, a umožňuje i využití různých proprietárních či komerčních formátů, není považován za dostatečný stupeň otevřenosti.

Stupeň otevřenosti 3

Stupeň otevřenosti 3 oproti stupni 2 navíc vyžaduje, aby specifikace formátu, ve kterém je distribuce datové sady vyjádřena, byla otevřená. Tzn., aby byla vyhledatelná a zdarma dostupná v síti WWW a aby existovaly volně dostupné programovací nástroje pro jejich zpracování (tj. knihovny apod.). Příkladem mohou být např. formáty definované v podobě RFC (např. formát CSV, který je definovaný v RFC 4180) nebo formáty definované konsorciem WWW (např. formát XML, který je definovaný v W3C Recommendation Extensible Markup Language (XML) 1.0 (Fifth Edition)).

V distribuci datové sady mohou být vyjádřeny pouze údaje tvořící datovou sadu a konstrukce vyjadřující jejich typ či sémantiku (sémantiku vyjadřuje např. hlavička tabulky v CSV souboru nebo XML tagy a atributy v XML elementu ohraničující údaj s určitým významem). Do distribuce datové sady nepatří konstrukce vyjadřující formátování (barvy nebo styl písma, v tabulkových datových sadách pak navíc ohraničení či slučování buněk, atd.), konstrukce s grafickými prvky, které nejsou součástí údajů v datové sadě, apod.

Do stupně otevřenosti 3 nespadá formát PDF. Taktéž sem nespadají formáty kancelářských aplikací bez otevřené specifikace, jako jsou např. formáty Microsoft Office ve verzích do roku 2003 (tj. DOC, XLS apod.).

Do stupně otevřenosti 3 spadají formáty kancelářských aplikací s otevřenou specifikací (např. formáty Office Open XML či OpenDocument). Distribuce datových sad vyjádřené v těchto formátech však mohou obsahovat pouze údaje tvořící datovou sadu, nikoliv instrukce, které definují jejich formátování při zobrazování na obrazovce či tisku.

Pro prostorová data je vhodné zvolit některý z otevřených formátů OGC (Open Geospatial Consortium), jako je [GML](#), [GeoJSON](#) nebo [GeoPackage](#). Pro reprezentaci prostorové složky dat je vhodné využít prostorových objektů v některém ze základních formátů, například GML nebo WKT.

Za otevřený a pro některé typy dat i vhodný formát můžeme považovat proprietární ESRI Shapefile (firma ESRI neklade žádná licenční omezení na jeho použití) nebo komunitní GeoJSON (ten zatím nebyl žádným oficiálním orgánem standardizován). Oba tyto formáty mají širokou podporu v dalších

softwarech, nejsou ale vhodné na všechny typy datových sad.



Technické standardy pro datové sady na stupni otevřenosti 3

Stupeň otevřenosti 4

Stupeň otevřenosti 4 zavádí povinnost v distribuci datové sady identifikovat entity, kterých se týkají údaje obsažené v datové sadě. Identifikátory musí mít tvar Internationalized Resource Identifier (IRI). Např. v distribuci datové sady o knihách musí mít všechny entity jako knihy, autoři či vydavatelé přiřazen identifikátor ve tvaru IRI.

Stupeň otevřenosti 5

Stupeň otevřenosti 5 je nejvyšším stupněm otevřenosti. Dále rozšiřuje stupeň 4. Vyžaduje, aby distribuce splňovala standardy propojených dat (angl. Linked Data). Standardy propojených dat jsou postaveny nad dobře známými standardy sítě WWW a umožňují vyjadřovat souvislosti mezi různými datovými sadami v podobě strojově zpracovatelných odkazů, které se podobají hypertextovým odkazům v síti WWW.



Otevřená formální norma pro Propojená data.

Stupně otevřenosti v kontextu veřejné správy České republiky

V kontextu veřejné správy České republiky je nejnižším stupněm otevřenosti stupeň 3. Datové sady publikované na stupních otevřenosti 0, 1 a 2 nejsou považovány za otevřené.

Jednotlivé orgány veřejné správy mohou publikovat své datové sady na nižších stupních otevřenosti, ale tyto datové sady pak nejsou otevřené. Neznamená to však, že otevřená datová sada nesmí být publikována na stupni otevřenosti 2. Je vždy nutno zajistit publikaci jejího kompletního obsahu na stupni otevřenosti alespoň 3 a vedle toho lze její obsah publikovat i na stupni 2 (příp. 1, pokud to je smysluplné a účelné).

V praxi to znamená, že uživatelům dat je dostupná jak varianta na úrovni 2 tak varianta na úrovni 3. V případě datové sady s tabulkou je např. možné zveřejnit distribuci datové sady v podobě XLSX souboru, který obsahuje řadu grafických prvků a formátování pro přehledné zobrazení tabulky na obrazovce uživatele - čtenáře (stupeň 2) a vedle toho další distribuci této datové sady v podobě jednoduchého CSV souboru, který žádné grafické prvky a formátování neobsahuje a umožňuje uživatelům - programátorům aplikací a datovým analytikům co nejsnazší přístup k údajům v tabulce

(stupeň 3).

Dále je možné, aby poskytovatel provozoval nad svými otevřenými daty nějakou webovou, mobilní či jinou softwarovou aplikaci, která data zpřístupňuje široké veřejnosti (kterou často nezajímají strojově čitelná data, ale právě aplikace nad nimi vytvořené). Existence nebo neexistence takových aplikací však nijak neovlivňuje otevřenost datové sady. Tj. pokud poskytovatel dat provozuje pouze takovou aplikaci (např. HTML formulář pro vyhledávání záznamů a jejich následné zobrazení) a neposkytuje kompletní obsah datové sady na stupni otevřenosti 3, neposkytuje otevřená data.

Za otevřená data se nepovažuje webová služba (nebo obecně API), která umožňuje získávat jednotlivé záznamy tvořící datovou sadu. Především proto, že účelem otevřených dat je poskytovat kompletní obsah a ne jednotlivé záznamy. Poskytovatel dat ve veřejné správě typicky nedisponuje infrastrukturou s takovou kapacitou, aby umožnil všem zájemcům o data získat kompletní obsah datové sady prostřednictvím takové webové služby. Je tak nucen omezovat počet přístupů k webové službě v čase, což je v přímém rozporu s podstatou otevřených dat. Poskytování otevřených dat není v rozporu s poskytováním dat prostřednictvím webových služeb, tyto přístupy se mohou vhodně doplňovat a mohou vedle sebe koexistovat. Není ale možné v žádném případě označovat takovou webovou službu jako otevřená data.

Z hlediska prostorových dat je zásadní zákon 123/1998 Sb. o právu na informace o životním prostředí, který zajišťuje transpozici evropské Směrnice INSPIRE do české legislativy. Směrnice INSPIRE se zabývá standardizací prostorových dat v Evropě, určuje jejich strukturu, definuje způsob jejich publikace a usnadňuje jejich vyhledávání prostřednictvím vyhledávací služby nad katalogem metadatových záznamů. Směrnice INSPIRE neukládá povinnost všechny datové sady otevírat, určuje způsob, jak se mají prostorová data publikovat. Publikace prostorových dat je dle Směrnice INSPIRE zakotvena mezi úrovněmi 3 (otevřené formáty a webové služby pro poskytování dat) a 4 (jednotlivé datové sady a služby jsou vyhledatelné na úrovni metadatových záznamů). Směrnice INSPIRE nepřináší nové standardizační postupy. Využívá stávajících standardů a norem a na jejich základě vytváří komplexní návod pro standardizaci a publikaci.

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

Permanent link:
<https://opendata.gov.cz/informace:stupn%C4%9B-otev%C5%99enosti-datov%C3%BDch-sad>

Last update: **2020/06/03 09:37**

