

Technické standardy pro datové sady na stupni otevřenosti 3

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



Část týkající se prostorových dat byla nahrazena [Otevřenou formální normou pro Prostorová data](#).

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: **2020/09/02 08:14**

