



ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY

ÚRAD SPLNOMOCNENCA VLÁDY SLOVENSKEJ REPUBLIKY
PRE ROZVOJ OBČIANSKEJ SPOLOČNOSTI

Štandardy pre sprístupňovanie dátových zdrojov

Časť 1: datasety a zverejňovanie

V tomto dokumente sú uvedené požiadavky na štandardy sprístupňovania údajov v rámci portálu data.gov.sk. Popisuje štandardy týkajúce sa samotného sprístupňovania údajov.

Údaje sú vždy súčasťou určitého dátového zdroja (datasetu). Dataset je ucelená skupina údajov, vytvorená a udržiavaná za určitým účelom. Dátové zdroje verejnej správy sú vytvárané za účelom podpory služieb verejnej správy, služieb vo verejnom záujme alebo verejných služieb. Jednotlivé agendy súvisiace s výkonom verejnej správy môžu obsahovať viaceré dátové zdroje. Charakteristickým znakom dátového zdroja je jeho samostatná použiteľnosť (výpovedná hodnota) nezávisle od iných dátových zdrojov. Ďalšími znakmi popisujúcimi dátový zdroj sú názov, účel (zameranie), typy spracúvaných údajov a vzťahy medzi nimi, formát údajov a podobne – tzv. metadáta.

Sprístupnenie dátového zdroja zabezpečuje jeho správca na základe vlastného rozhodnutia a pomocou vlastných prostriedkov (t.j. na portáli otvorených dát budú uložené len katalógy, nie samotné údaje).

Tento dokument definuje minimálne štandardy, ktoré musia spĺňať datasety verejnej správy na portáli otvorených dát. Základná požiadavka na spôsob sprístupnenia údajov je, **aby údaje boli automatizovane strojovo spracovateľné**. Štandardy sú vytvorené na dosiahnutie tejto požiadavky. Pre minimálne štandardy je toto kritérium **základným merítkom** splnenia požiadavky na sprístupnenie údajov.

Zároveň sú správcovia dátových zdrojov povinní dodržiavať zákon č.275/2006 Z.z. o ISVS vznp. (ďalej "zákon") a príslušné štandardy pre ISVS vydané MF SR (§2 ods.3 písm. a) zákona). Na základe §3 ods.4 písm. d) „povinné osoby, ktoré sú správcami, sú povinné sprístupňovať verejnosti údaje z informačných systémov verejnej správy, ak osobitný predpis neustanovuje inak“ a e) „povinné osoby, ktoré sú správcami, sú povinné sprístupňovať alebo na požiadanie poskytnúť bezplatne iným povinným osobám údaje z informačných systémov verejnej správy potrebné na ich činnosť, ak osobitný predpis) neustanovuje inak“ zákona všetky ISVS už sú na sprístupňovanie údajov pripravené. To znamená, že iniciatíva data.gov.sk je skôr organizačnou záležitosťou. Štandardy pre ISVS aktuálne obsahuje Výnos MF SR č.312/2010 Z.z. (ďalej "výnos").

Všetky používané štandardy musia byť otvorené a technologicky neutrálne (§6 ods.1 zákona).



ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY

ÚRAD SPLNOMOCNENCA VLÁDY SLOVENSKEJ REPUBLIKY
PRE ROZVOJ OBČIANSKEJ SPOLOČNOSTI

Časť 1: formáty údajov

Rozlišujeme dva základné spôsoby sprístupnenia:

1. **prístup k súborom** s údajmi - ide o pasívny spôsob prístupu, kde server sprístupní "naraz" celý blok údajov (súbor), vyhľadávanie v údajoch je ponechané na používateľa
2. **prístup cez aplikačné rozhranie (API)** - ide o aktívny prístup, kde používateľ zadáva serveru dotazy na konkrétne požadované údaje, ktoré server po spracovaní dotazu vyhledá a odošle

Formáty súborov:

- ide spravidla o súbory obsahujúce tabuľky
- minimálne požiadavky na formáty:
 - [CSV](#), alebo
 - [XML](#) so schémou
- Odporúčané formáty: [Linked data](#) vo formátoch [RDF](#) ([Turtle](#), [RDF/XML](#), ...), [OWL](#), [SKOS](#), ...

Prípustné nie sú:

- tabuľky vyjadrené v textovom súbore
- proprietárne formáty tabuľkových súborov - napr. XLS
- súbory obsahujúce aktívne prvky tabuliek (napr. makrá, vzorce)
- iba obalenie nevhodného súboru do XML

API:

- ide o prístup ku aplikácii spravujúcej bázy údajov o datasete
- minimálne požiadavky:
 - [RESTful](#)
 - vo formáte [XML](#) alebo [JSON](#)
- odporúčané formáty: ako vyššie ale navyše aj vo formáte [RDF](#) ([Turtle](#), [RDF/XML](#), ...) cez [SPARQL endpoint](#)

Prípustné nie sú:

- nekonzistentné či neúplné dáta
- nekonzistentné či neúplné API (ak napríklad obsahuje funkciu "[getItem\(\)](#)" ale neobsahuje "[listItems\(\)](#)")



ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY

ÚRAD SPLNOMOCNENCA VLÁDY SLOVENSKEJ REPUBLIKY PRE ROZVOJ OBČIANSKEJ SPOLOČNOSTI

- nekonzistentná či neúplná dokumentácia k API
- nevysvetlené a neohlásené výpadky funkčnosti API

Časť 2: štruktúra údajov

Štruktúra údajov pre dátový zdroj:

- dátum, ku ktorému sú údaje platné, alebo informácia, že ide o aktuálne údaje v čase prístupu
- odporúčané: dátum najbližšej aktualizácie (pokiaľ nejde o sprístupnenie vždy aktuálnych údajov)
- sprístupnenie doplňujúcich informácií, ktoré majú napomôcť automatizovanému spracovaniu údajov dátového zdroja:
 - schémy údajov - pokiaľ schéma údajov (t.j. členenie dátového zdroja na typy údajov, konkrétne záznamy a vzťahy medzi týmito entitami) nie je triviálna
 - popis typov položiek - najmä v prípade, ak sú používané netypické dátové typy, číselníkové typy, skratky, zložené dátové typy a pod.
 - popis formátov v ktorých je dátový zdroj sprístupňovaný - napr. formáty súborov
 - popis možných nepravidielností v štruktúre

V prípade, že je správcovi údajov známe, že niektoré údaje sú neaktuálne, nesprávne, alebo neúplné, tieto údaje musia byť označené spôsobom umožňujúcim automatizovane ich odlíšiť od aktuálnych, správnych, alebo úplných údajov (o.i. to znamená, že prítomnosť takýchto údajov nie je sama osebe dôvodom na nesprístupnenie dátového zdroja).

Jednotlivé položky údajov ukladať spôsobom:

- umožňujúcim ich lokalizáciu (najmä odlíšenie od iných položiek) v rámci dátového zdroja
- čítanie automatizovaným spôsobom
- rovnakým spôsobom pre všetky dátové vety v určitom dátovom zdroji (dátová veta je množina súvisiacich položiek opisujúcich určitý objekt)

Prípustné nie sú:

- formáty určené na čítanie pre používateľa, ktoré neumožňujú automatizované spracúvanie údajov (napr. web aplikácia)
- nepravidielný formát dátovej vety
- nemožnosť izolovať z dátovej vety/súboru konkrétne položky



ÚRAD VLÁDY SLOVENSKEJ REPUBLIKY

ÚRAD SPLNOMOCNENCA VLÁDY SLOVENSKEJ REPUBLIKY
PRE ROZVOJ OBČIANSKEJ SPOLOČNOSTI

Časť 3: prístup k údajom

Údaje sa sprístupňujú prostredníctvom sieťovej infraštruktúry (§3 ods.2 písm. j) zákona), v sieti Internet.

Dátový zdroj musí mať určitú lokáciu, ktorá je stabilná:

- vyjadrená pomocou URL (resp. iného konkrétneho identifikátora, ak ide o prístup iným protokolom)
- zmena lokácie nastáva iba vo výnimočnom prípade, napr. pri zmene formátu údajov, nasadzovaní nového webového sídla

Prístupu sa venujú štandardy pre prepojenie, najmä §3, §4, §5 výnosu a štandardy pre prístup k elektronickým službám, najmä §9 výnosu a štandardy pre webové služby, §11 výnosu.

Server musí vyhodnocovať požiadavky na prístup bezstavovo, t.j. požiadavka je vyhodnotená bez ohľadu na spracovanie predchádzajúcich požiadaviek. Prípustné sú aj viaceré "pohľady" na údaje. Pod pohľadom rozumieme spoločne prezentovanú časť dátového zdroja, spravidla ide o na dátovom zdroji vykonanú reštrikciu (filter) určitých položiek, objektov, alebo vzťahov medzi objektmi - napr. ak z dôvodu ochrany osobných údajov nie je možné sprístupniť celý dátový zdroj, je prezentácia rôznych pohľadov žiadanou alternatívou.

Prípustné nie sú:

- "roztrúsené" umiestnenie údajov - napr. potreba preklikávania "stránok", parsovanie údajov zo stránok
- prvky určené na ovládanie používateľom - napr. tlačítka, grafické prvky
- požadovaná identifikácia, autentifikácia, či iná práca s používateľom