

Slovenská technická univerzita v Bratislave  
Fakulta informatiky a informačných technológií  
študijný program: Informatika

---

Tomáš Mizera

# Prepájanie a čistenie verejných otvorených dát

Bakalársky projekt 1

Vedúci bakalárskeho projektu: Ing. Róber Móro, PhD.  
December 2018

## Anotácia

Verejné otvorené dáta (VOD) získavajú na popularite po celom svete. Vlády a štátne inštitúcie pracujú na adaptovaní princípov VOD do svojich legislatív a zverejňovaní dát pre verejnosť.

Proces zverejňovania však nie je proces automatický, vyžaduje si manuálne spracovanie predtým ako môžu byť dáta zverejnené. Takáto manuálna práca však často so sebou prináša problémy ako náhodné zavedenie chýb do dát. Problémom taktiež je, že i keď existuje portál, kde by sa mohli zverejňovať VOD ([data.gov.sk](http://data.gov.sk)), niektoré inštitúcie zverejňujú VOD na vlastných portáloch.

Cieľom nášho projektu je poskytnúť sprostredkovateľskú platformu, kde sa prepájajú znečistené VOD zverejňované na viacerých štátnych portáloch. Zároveň dúfame vo zvýšenie popularity o oblasť VOD a prispetiu k vytvoreniu transparentnejšiemu vzťahu medzi vládou a jej občanmi.

## **Annotation**

Open Government Data (OGD) gains more and more popularity around the world. Governments and state institutions work hard to adopt OGD principles into their legislature and disclose datasets to public.

However, the process of disclosing government data is not automated and needs a manual processing to be done. Manual processing often carries with it an errors spread across datasets. Also, although there is a centralized platform for raw datasets to be published on, some institutions tend to publish datasets on their own sites.

The goal of our project is to provide an intermediary platform, where messy OGD published in several government platforms can be viewed in human-friendly form. Moreover, it hopes to increase popularity of OGD and create more transparent relationship between government and its citizens.

# Obsah

<b>1 Úvod</b>	<b>1</b>
<b>2 Otvorené dáta</b>	<b>2</b>
2.1 Verejné otvorené dáta (VOD) . . . . .	3
2.1.1 Prínos VOD . . . . .	4
2.1.2 Zúčastnené strany vo VOD . . . . .	5
2.2 Merateľnosť kvality otvorených dát . . . . .	6
2.3 Prepájanie datasetov VOD . . . . .	8
<b>3 Verejné otvorené dáta vo svete a na Slovensku</b>	<b>10</b>
3.1 Verejné otvorené dáta na Slovensku . . . . .	11
3.1.1 Portál otvorených dát . . . . .	12
3.1.2 Centrálny register zmlúv . . . . .	12
3.1.3 Register a identifikátor právnických osôb, podnikateľov a orgánov verejnej moci . . . . .	14
3.1.4 Register partnerov verejného sektora . . . . .	15
3.1.5 Projekty vzniknuté na základe VOD na Slovensku . . . . .	17
3.2 Verejné otvorené dáta v Taliansku . . . . .	20
3.2.1 Taliansky portál otvorených dát . . . . .	20
3.2.2 OpenCoesione register projektov . . . . .	20
3.3 Diskusia . . . . .	22
<b>4 Dataset Ministerstva vnútra SR</b>	<b>25</b>
<b>5 Návrh spracovania získaných VOD</b>	<b>26</b>
<b>6 Spracovanie a zobrazenie VOD</b>	<b>26</b>
<b>7 Zhodnotenie a záver</b>	<b>26</b>
<b>8 Zdroje</b>	<b>26</b>
<b>Literatúra</b>	<b>26</b>
<b>9 Prílohy</b>	<b>27</b>
9.1 Časový harmonogram na prácu v druhom semestri . . . . .	28
<b>10 Hot ToDos</b>	<b>28</b>

# 1 Úvod

V úvode budem opisovať ako sa open data stali celosvetovým fenoménom, že sú myšlienkou, ktorá hýbe svetom a veľmi veľa krajín si ju adaptuje. Open data ponúka možnosť zapojiť sa bežným ľuďom do diania na verejnosti a byť tak súčasťou diania alebo ponúka možnosť slúžiť verejnosti aj v oblasti informatiky.

The open government has been placed at the core of the political agenda at international level, following the national oriented directives such as, e.g., the US Transparency and Open Government Memorandum (Obama, 2009), further detailed in the Open Government Directive – OGD (Orszag, 2009) and the EU directive 2003/98/EC on the Re-use of Public Sector Information – PSI Directive (European Parliament and Council, 2003) [19].

In 2009, the American President Obama issued a Memorandum on Transparency and Open Government (McDermott, 2010; Obama, 2009). [9]

Ďalej tu budem (možno) hovoriť o tom ako by mohli otvorené dáta zmeniť situáciu na Slovensku, resp. o tom ako zmenili situáciu v Taliansku, na ich hackatone (**podľa článku Open government data ecosystem**).

Chcel by som tu taktiež spomenúť ako sme prišli k datasetu s ktorým robím, že vznikol počas #AllForJan Hackathonu, crawlovaním stránok ministerstva vnútra, je to dataset, ktorý obsahuje objednávky vykonané minv od roku XXXX, že nad ním bolo treba vykonať veľa čistiacich metód, následne sa integroval spoločne s datasetom faktúr a že bližšie je opísaný v časti *BLAH*.

Toto možno do záveru:: Ešte by som tu chcel povedať, že naša práca dúfa vo vytvorenie transparentnejšej spoločnosti na základe článku **TODO: Treba nájsť ktorý to bol**.

Ďalší hint: Open data is just one of a number of high-profile labels with the prefix "open". Open government, open access, open innovation, open education and open knowledge are some of the other initiatives and movements in this area [5].

Ešte by sa tu dalo napísať že každá spoločnosť sa môže rozhodnúť zverejňovať svoje dáta, vtedy hovoríme o otvorených dátach, viac v časti 2 Otvorené dáta.

p.p. by sa toto dalo dať ako podkapitola do analýzy:: Ešte by som chcel predstaviť aké postavy vystupujú v oblasti verejných otvorených dát, že to je vláda (vytvára datasety), občania (originálne ľudia, pre ktorých majú byť dáta určené, no často sú uložené v podobách, ktoré nie sú pre občanov, bez technického vzdelania, alebo proste niekto kto nemá čas sa hrabať v kope dát, ktoré sú divne uložené, zobrazené a pod., príp. pochopiteľné, preto treba tretiu osobu: sprostredkovateľa (intermediary), to je niekto kto berie dáta od štátu a posúva ich ďalej v peknej, čitateľnej podobe, alebo sprostredkúva ich zobrazenie v peknej podobe. **TODO: nájsť zdroj kde je toto definované**

### 2 Otvorené dáta

Pri snahe definovať otvorené dáta, je nápomocné sa najskôr pozrieť na všeobecnú definíciu otvorených znalostí. Open Knowledge International [2] hovorí, že o znalostiach môžeme hovoriť ako o otvorených, pokiaľ k nim vie ktokoľvek voľne pristupovať, používať ich, upravovať a šíriť ďalej, pričom vždy existuje snaha o maximálne zachovanie ich pôvodu a otvorenosti.

Z definície otvorených znalostí vidíme hneď niekoľko podmienok, ktoré musia byť dodržané, aby sa dali znalosti považovať za otvorené. Definícia otvorených znalostí však nie je úplne dostačujúca pre definovanie otvorených dát, preto [3] pokračuje a pre otvorené dáta pridáva podmienky, ktoré musia byť splnené, aby sa dáta dali prehlásiť za otvorené. Dáta preto musia byť:

- zverejnené pod otvorenou licenciou, ktorá zabezpečuje možnosti manipulovania s dátami podobne ako pri definícií otvorených znalostí (voľný prístup, možné šírenie,...) a akékoľvek ďalšie podmienky, ktoré dáta doprevádzajú, nesmú byť v rozpore s podmienkami licencie,
- zverejňované bezplatne, prípadne za cenu nie väčšiu ako cena ich produkcie,
- dostupné online,
- *strojovo čitateľné* a mať otvorený formát, kedy netreba špeciálny platený software pre ich zobrazenie a manipuláciu s nimi.

Otvorené dáta sa zverejňujú v rôznej podobe a forme, môže ísť o data-sety inštitúcií vo formátoch ako CSV, XLSS, RDF a pod., zároveň môže ísť o zverejnenie aplikačného rozhrania (API), vďaka ktorému sa vieme dopytovať do databázy vlastníka dát. Rôzne použité formáty zverejnených dát poukazujú na kvalitu zverejňovaných dát [18] !!!?. Kvalita otvorených dát je bližšie opísaná v časti 2.2 Merateľnosť kvality otvorených dát.

Z definície otvorených dát je zrejmé, že existuje subjekt, ktorý svoje dáta zverejňuje, ako aj subjekt, ktorý tieto dáta využíva podľa vlastnej potreby tak, ako to umožňuje licencia, pod ktorou sú dáta zverejňované. Viaceré zdroje [6, 17, 14, 9] však spoločne identifikujú aj tretí subjekt, sprostredkovateľa, ktorý vystupuje ako prostredník medzi zverejňovateľom dát a ich prijímateľom. Na obrázku č.1 môžeme vidieť vzťah týchto troch subjektov, ktoré Ubaldi [17] definuje nasledovne:

*Poskytovateľom* dát sa rozumie subjekt, ktorý sprístupňuje svoje dáta verejnosti. Poskytovateľ dát môže byť podľa obrázka č.1 buď subjekt z verejného sektora (ministerstvo, štátna inštitúcia,...) alebo subjekt z privátneho sektora (firmy, živnostníci). Pri zverejňovaní dát subjektom z verejného sektora, hovoríme už o verejných otvorených dátach, bližšie definovaných v časti 2.1 Verejné otvorené dáta (VOD).

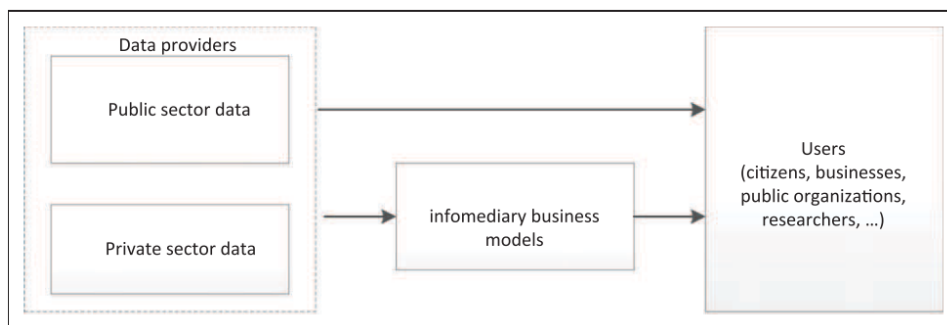
## 2. OTVORENÉ DÁTA

---

*Používateľom* sa rozumie subjekt, ktorý získava informácie od poskytovateľa dát a manipuluje s nimi podľa vlastného uváženia práve tak, ako to dovoľuje otvorená licencia. Môže ísť o občanov štátu, neziskové organizácie, vedcov, firmy, a ďalšie osoby.

Medzi poskytovateľom dát a používateľom však môže nastať technologická priepasť, kedy forma akou poskytovateľ zverejňuje svoje dáta je pre používateľa technologicky neprístupná (napríklad ak poskytovateľ zverejňuje dáta prostredníctvom otvoreného API, bežný používateľ bez technického vzdelania nedokáže z takto otvorených dát nič vyčítať) [6]. Túto priepasť však práve premostňuje *sprostredkovateľ*, ktorý spája poskytovateľa dát a ich prijímateľov tým, že dáta od poskytovateľa poskytuje používateľom v prijateľnejšej pochopiteľnejšej podobe.

V oblasti verejných otvorených dát vystupujú podobné subjekty ako v otvorených dátach, ktoré však už nadobúdajú podobu reálnych osôb zo sveta, tieto sú bližšie opísané v časti 2.1.2 Zúčastnené strany vo VOD.



Obr. 1: Zúčastnené strany v otvorených dátach [9]

### 2.1 Verejné otvorené dáta (VOD)

Podobne ako každá inštitúcia, môže aj vláda (vládne inštitúcie, ministerstvá) pre občanov zverejňovať svoje dáta. V takomto prípade sa vláda stavia do pozície poskytovateľa dát a občania sa stávajú prijímateľmi vládných dát. Popri výmene dát medzi vládou a občanmi získavajú otvorené dáta prívlastok *verejné otvorené dáta* (ďalej iba VOD) **!ZDROJ** z anglického Open Government Data.

Vláda počas svojej bežnej prevádzky produkuje a zhromažďuje množstvo dát, ktoré môžu byť pre verejnosť zaujímavé a nápomocné v rôznych oblastiach [17], ako napríklad:

- verejné dáta ako zdravotnícke, kriminálne dáta, dáta z oblasti školstva, dopravy, meteorológie,

## 2. OTVORENÉ DÁTA

---

- rôzne štatistické údaje, napríklad údaje o počte hlasov vo voľbách, počet obyvateľov obcí,
- údaje o štátnom hospodárení, vládnych rozpočtoch, výdavkoch,
- údaje o právnických osobách a živnostníkoch, ktorými vláda disponuje a neobsahujú citlivé dáta,
- mnoho ďalších dát ako geografické, kultúrne, agrikultúrne dáta, dáta vyšších územných celkov, miest a obcí.

Zverejňovanie dát, ktoré sú svojím obsahom zaujímavé pre verejnosť dokáže viesť k väčšiemu používaniu VOD a tým aj k väčšiemu dopytu po nových dátach [12]. Táto paralela pozitívne prispieva na fakt, že zverejňovanie vládnych dát hrá zároveň významnú rolu pri vytváraní väčšej transparentnosti vlády [15]. *!TODO:Možno lepšie by bolo opísať obrázkom, takým cyklom.*

V najbližších kapitolách (2.1.1-2.3) sú postupne opisované jednotlivé časti VOD ako ich prínos, strany vystupujúce špeciálne vo VOD, kvalita VOD a možnosť a spôsoby ich prepájania.

### 2.1.1 Prínos VOD

Táto kapitola sa pozerá na prínos VOD z dvoch pohľadov: najskôr sa pozerá na pozitíva, ktoré so sebou VOD prinášajú do spoločnosti a neskôr aj na negatíva, resp. okolnosti, s ktorými treba rátať pri zverejňovaní VOD. Nasleduje niekoľko výhod, ktoré so sebou VOD prinášajú.

- *Proaktívne odhaľovanie dát.* Vládne dáta sú klasicky zverejňované na požiadavku, kedy musí žiadateľ o prístupnenie dát prísť na konkrétny úrad a žiadosť podať, toto sa nazýva reaktívne odhaľovanie dát. VOD však so sebou prinášajú proaktívne odhaľovanie, a teda dáta sú najskôr zverejnené a uložené na jednotnom portáli ešte predtým ako vznikne požiadavka na ich zobrazenie [16].
- *Transparentnosť.* Zverejňovaním stále väčšieho množstva zaujímavých dát, napríklad z oblasti štátneho rozpočtu, dokážu VOD verejnosti lepšie komunikovať aké má vláda práve stanovené ciele a ako dobre sa jej darí tieto ciele naplňať [17], je to teda schopnosť verejnosti vidieť čo sa deje vo vláde [8]. *!Pripísať: čo všetko vieme získať, čas, čísla faktúr, a mrte veľa ďalších vecí, ktoré sú uvedené v časti XXXX, kde je príklad*
- *Zodpovednosť vlády.* Vytvorenie väčšej transparentnosti dokáže vytvárať tlak na vládu, aby zodpovednejšie pristupovala k vykonávaniu svojich povinností.
- *Väčšia zapojenosť občanov.* VOD sú zároveň označované ako efektívny nástroj k zapájaniu občanov do verejného diania. Občania môžu zo



## 2. OTVORENÉ DÁTA

---

zverejnených dát získavať nové informácie a na ich základe robiť napríklad informované rozhodnutia, no taktiež môžu surové zverejnené dáta ďalej spracúvať a vystupovať tak ako sprostredkovatelia dát. [6, 10]

- *Zníženie korupcie.* Ak sú k dispozícii správne dáta, občania dokážu jednoducho sledovať správnosť vládnych krokov a rýchlo zareagovať, poukázať na chybu, ak narazia na záznamy, ktoré sú podozrivé [11].

Okrem uvedených výhod prinášajú VOD so sebou ešte mnoho ďalších. Janssen, Charalabidis a Zuiderwijk [10] vo svojom článku vypracovali prehľadovú tabuľku v ktorej uvádzajú ďalších 30 výhod VOD, ktoré rozdelili do troch kategórií (Politickej a sociálnej, ekonomickej a technickej).

*Pridať tabuľku z Janssena:)*

Cenou týchto výhod je však veľké úsilie, ktoré musí byť vynaložené zo strany vlády, aby sa dáta vôbec mohli zverejniť. Janssen [8] totižto uvádza, že vládne dáta často nie sú v dobrej, na zverejnenie pripravenej, podobe a kvalite. Vysvetľuje, že systémy, ktoré vlády používajú neboli navrhované s myšlienkou následného sprístupňovania dát, preto ešte predtým ako môžu byť dáta zverejnené, je potreba veľa manuálnych úprav. Databázy týchto systémov obsahujú rôzne typy záznamov, obsahujú aj citlivé dáta, ktoré je treba pred zverejnením vymazať z cieľového datasetu. Pri takýchto úpravách môžu nastať nezrovnalosti v dátach, vo vzťahoch jednotlivých tabuliek databáz (ak napríklad tieto citlivé dáta predstavovali referenčné hodnoty). Takéto úkony spomaľujú zverejňovanie nových dát a proces zverejňovania si vyžaduje väčšie úsilie.

### 2.1.2 Zúčastnené strany vo VOD

Viaceré zdroje [17, 14] **TODO: sem pridať ďalšie zdroje čo definujú strany** sa zhodujú pri definovaní strán, vystupujúcich v oblasti VOD, sú to:

- **vláda**
- **sprostredkovateľ**, niekedy nazývaný aj ako technická komunita [14]
- **občan**, resp. občania

#### Vláda

Opisuje vládu ako subjekt zverejňujúci svoje datasety. Nemá to vždy jednoduché, musí dáta čistiť od citlivých informácií, resp. osobných informácií. Taktiež musí vynakladať úsilie na vytváranie datasetov. Zároveň sa však predpokladá že štát môže vďaka VOD aj získať, na základe výhod spomenutých vyššie sa dá povedať, že zodpovedná vláda môže získať na kredibilitu, môže získavať väčšiu dôveru svojich občanov a zároveň vie vďaka VOD lepšie zapájať ľudí do verejného diania. and more effectively and efficiently deliver

smarter, innovative and more personalised public services, while improving the quality of interactions between the governments and the users. [17]

### Sprostredkovateľ

*"VOD sú nepoužiteľné pokiaľ neporozumieme ich zmyslu v reálnom svete"* [8]. Dáta, ktoré vláda sprístupňuje sú často príliš "technické", napríklad v rámci VOD má Slovensko viaceré otvorené aplikačné rozhrania ( API ), ktoré sú pre bežných občanov nepoužiteľné, pretože s nimi nevie manipulovať. Sprostredkovateľom sa rozumie práve taká osoba, ktorá technicky alebo inak nepoužiteľné ( napríklad znečistené, zobrazené v nečitateľnej podobe a pod.) zverejnené vládne dáta pretvára a zobrazuje v pochopiteľnej podobe pre občanov.

Secondly, intermediaries can also be producers of data. In some cases, they are located close to data sources – such as communities – and they possess technical knowledge about dataset manipulation. Intermediaries may therefore collect data in the field – sometimes merging it with existing datasets – in order to create new open government data. [6]

Strana sprostredkovateľov so sebou taktiež nesie zodpovednosť, nakoľko občania z veľkej časti prezerajú VOD skrze sprostredkovateľov, vlastnosti ako použiteľnosť, technická podpora, jednoduchosť, používateľský zážitok a pod. dokážu priamo ovplyvniť motiváciu občanov používať VOD [14]. Ak sa stane, že sprostredkovateľský portál, ktorý zverejňuje VOD je často pomalý, nedostupný, má zlú navigáciu, príp. ďalšie problémy, záujem občanov o VOD môže často klesať [12].

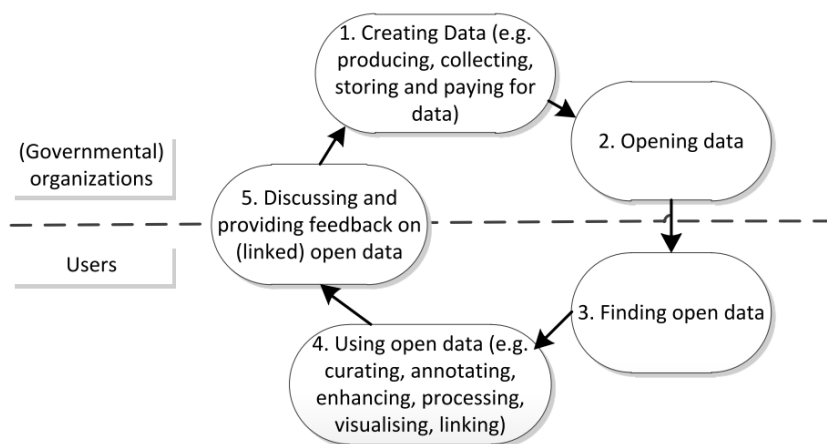
### Občan

Občan vo VOD vystupuje ako osoba získavajúca informácie od vlády, príp. od sprostredkovateľa dát. Vďaka VOD sa dokáže lepšie angažovať na dani v štáte, umožňujú mu lepšie sa rozhodovať, napríklad ak štát zverejňuje informácie o kriminalite v určitej oblasti, dokáže občan zvážiť sťahovanie sa do danej oblasti [17].

## 2.2 Merateľnosť kvality otvorených dát

Kvalita otvorených dát nám podáva doplňujúcu informáciu, vďaka ktorej dokážeme porovnávať viaceré datasety medzi sebou [18]. V časti 2.1.1 Prínos VOD sme hovorili o probléme pri zverejňovaní dát, kedy vládne systémy neboli navrhované s myšlienkou sprístupňovania ich dát [8]. V takýchto prípadoch sú potrebné manuálne úpravy dát, aby sa dáta vyčistili od údajov ktoré nesmú byť zverejnené, ako napríklad citlivé či osobné údaje.

Pri manuálnych úpravách je však v hre aj chybovosť ľudského faktora, ktorá môže viesť k zníženiu kvality a hodnovernosti finálnych zverejnených dát [17]. Pri rôznych úpravách sa môže chybovosť prejaviť inak, napríklad



Obr. 2: Akcie vykonávané jednotlivými subjektami vo VOD [20]

pri prepisovaní numerických hodnôt sa z nepozornosti môže človeku stať, že posunie desatinné miesto prepisovanej hodnoty, čo môže mať za následok vytvorenie vychýlených hodnôt. Výsledný zverejnený dataset potom môže mať preukázateľne nižšiu kvalitu ako pôvodný. Túto vlastnosť nazývame presnosť dát [18].

Mnohé práce [17, 8, 13, 18, 16, 4] sa zaoberajú aj ďalšími metódami merania kvality VOD a následného porovnávania datasetov. Tieto práce sa pozerajú na formát v akom sú dáta zverejňované, ako často sú dáta zverejňované, resp. aktualizované, ako presné sú, prip. ako dobre sa dajú strojovo spracovávať, či sú dostupné a ďalšie.

Viacere z nich [17, 18] používajú Tim Berners-Leeho [1] "hviezdičkový model", ktorý hodnotí kvalitu zverejnených datasetov v piatich bodoch:

1. dáta sú dostupné online a zverejnené pod otvorenou licenciou,
2. dáta sú v štruktúrovanej, strojovo spracovateľnej podobe (napr. formáte xlsx),
3. dáta sú dostupné v nekomerčných, otvorených formátoch (napr. formát csv namiesto xlsx)
4. pre lepšie porozumenie dát sa vytvárajú metadata použitím štandardov z W3C<sup>1</sup> (RDF a SPARQL) a záznamy majú svoje unikátne URI,
5. dáta sú prepojené, resp. prepožiteľné s dátami od iných poskytovateľov dát.

---

<sup>1</sup>World Wide Web Consortium, link: <https://www.w3.org>

Podľa toho koľko bodov zverejený dataset spĺňa (postupne od prvého bodu k poslednému), je ocenený prislúchajúcim počtom hviezdíčiek (ak teda dataset spĺňa prvé 4 body, ide o 4 hviezdíčkový dataset), pričom 1 hviezdíčka je najmenej a 5 najviac. Pri každom z bodov sa navyiac predpokladá splnenie všetkých predchádzajúcich bodov.

Viscusi, Spahiu a Maurino[19] na základe vyššie uvedeného hviezdíčkového modelu zhrnuli a zaradili známe formáty otvorených dát do 4 kategórií (bez hviezdíčky, 1, 2 a 3 hviezdíčky) ako je zobrazené v tabuľke č.1. Berners-Lee však ešte pridáva formát RDF a posúva ho na úroveň štyroch hviezdíčiek[1].

Skóre	N/A	*	**	***
Formát	nestiahnuteľný záznam na inter- nete, HTML	Word, PDF, PowerPoint	Excel	CSV, XML, JSON

Tabuľka 1: Kategorizácia formátov podľa Viscusi, Spahiu a Maurino[19]

V časti 3.3 Diskusia budeme porovnávať slovenské VOD aj na základe týchto kvalitatívnych vlastností:

- formát zverejnených dát, podľa hviezdíčkového modelu Berners-Leeho [1],
- časové obdobie, ako často sú dáta pridávané a existujúce datasety aktualizované.

### 2.3 Prepájanie datasetov VOD

VOD ako sme mohli vidieť v časti 2.1.1 Prínos VOD, prinášajú do spoločnosti mnohé výhody ako aj cenné informácie. Zatiaľ sme sa pozerali na VOD ako na jednotlivé datasety uverejnené na internete, z ktorých sme mohli čerpať informácie. Tim Berners-Lee [1] však hovorí o myšlienke prepájania jednotlivých datasetov VOD. Cieľom prepájania dát je agregovať, kontrolovať, integrovať, obohacovať a vytvárať dáta pre ďalšie použitie [7]. Taktiež informácie v našom datasete ktoré sa nám doteraz zdali irelevantné, môžu nabráť iný rozmer keď sa prepoja s ďalšími datasetmi [10].

V predošlej kapitole bol definovaný 5 hviezdíčkový model [1], ktorý poukazoval na kvalitu zverejnených datasetov. Štvrtá hviezdíčka (dáta vo forme RDF a využitie URI) súvisí s formátom, akým sa dáta prepájajú. Zatiaľ čo veľa datasetov je zverejňovaných vo formáte CSV, kedy ide iba o 3 hviezdíčkové dáta, formát RDF (z angl. Resource Description Framework) dokáže vďaka spôsobu akým využíva URI odkazy dáta identifikovať podľa referencií a prepájať s inými dátami na základe ďalších atribútov [14].

## 2. OTVORENÉ DÁTA

---

Piata hviezdička zasa hovorí o možnosti zverejnenia takto pripraveného vlastného datasetu na prepojenie s inými, podobne pripravenými datasetmi. Takéto vzájomné prepájanie poskytuje ďalšiu pridanú hodnotu VOD, a to znovu použitie dát. Podľa Janssena [20] datasety VOD nie sú iba na jedno použitie a tým, že sa prepájajú, je možné nachádzať závislosti medzi nimi.

## 3 Verejné otvorené dáta vo svete a na Slovensku

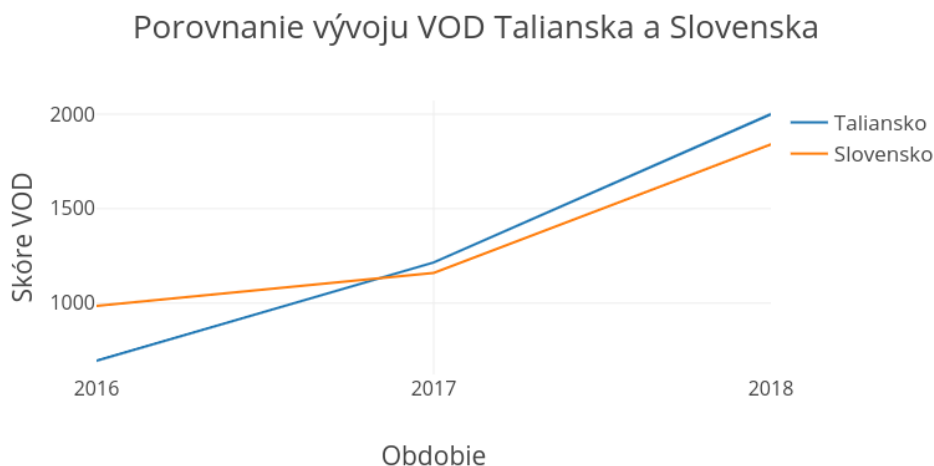
Verejné otvorené dáta na Slovensku nie sú iba veľkou neznámou. Dalgleish et al.[4] znázorňujú štatistiky v rámci Európskej únie, kde sa Slovensko umiestnilo vo viacerých z nich na popredných priečkach.

Dve hlavné metriky sú vyhodnocované vo viacerých štatistikách[4], a to:

- Vyspelosť portálu (z angl. Portal Maturity) - táto metrika je priamo závislá na verejnom portáli otvorených dát a jej hodnota rastie s kvalitou tohoto portálu.
- Pripravenosť (z angl. Open Data Readiness) - zasa poukazuje na rozsah adaptácie politiky otvorených dát a taktiež na mieru prepojenia medzi vládou a regionálnymi štatútmi ako VÚC, obecné politiky a pod.

Graf na obrázku č.4 predstavuje závislosť vyspelosti portálu ku pripravenosti. Krajiny Európskej únie sa zhlukovali do 4 kategórií od nováčikov po tzv. trendsetters, resp. udávačov tempa a trendu vo vývoji VOD. Slovensko sa v roku 2017 umiestnilo práve v kategórii trendsetters.

Európska únia vytvorila zároveň Európsky dátový portál<sup>2</sup>, ktorý okrem iných služieb, vyhodnocuje pravidelne pre členské štáty jednotlivé štatistiky VOD a na ich základe zostavuje rebríček krajín<sup>3</sup>.



Obr. 3: Graf vývoja vyspelosti VOD Slovenska a Talianska

V nasledujúcich dvoch podkapitolách sa budeme venovať aktuálnemu stavu VOD na Slovensku a v Taliansku, v časti 3.3 Diskusia potom zhrnieme

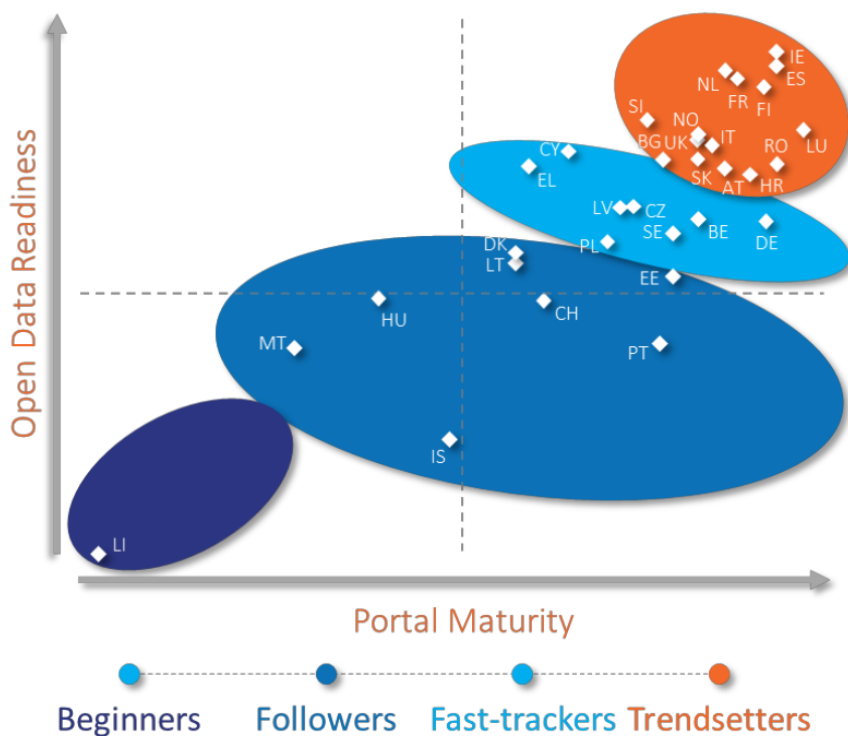
---

<sup>2</sup>EU Data Portal: <https://www.europeandataportal.eu/>

<sup>3</sup>Porovnanie VOD: <https://www.europeandataportal.eu/en/dashboard#2018>

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

a porovnáme vývoj a situáciu oboch krajín. Vybrali sme si Taliansko ako krajinu na porovnanie, pretože ako môžeme vidieť na obrázku č.3 a obrázku č.4, Slovenské aj Talianské VOD zapadá medzi trendsetters a na grafe vývoja vyspelosti môžeme vidieť, že Taliansko dokonca jemne predbehlo Slovensko, preto analyzujeme registre aj Slovenska, aj Talianska.



Obr. 4: Graf závislosti vyspelosti portálu a pripravenosti[4]

#### 3.1 Verejné otvorené dáta na Slovensku

Na Slovensku je v rámci VOD vytvorených niekoľko registrov, informačných systémov, ktoré zverejňujú v rôznych podobách svoje dáta pre verejnosť. Medzi známe slovenské VOD registre patrí: Centrálny register zmlúv, Portál otvorených dát, Register a identifikátor právnických osôb, podnikateľov a orgánov verejnej moci, Register partnerov verejného sektora, Otvorené dáta zverejňované na stránkach Ministerstva vnútra a ďalšie.. Tieto budú ďalej rozpisované v nasledujúcich podkapitolách.

#### 3.1.1 Portál otvorených dát

Podobne ako ostatné krajiny sveta adaptujúce si myšlienku verejných otvorených dát, aj Slovensko disponuje Portálom otvorených dát,<sup>4</sup> ktorý predstavuje centrálné miesto pre zverejňovanie nových datasetov verejnými inštitúciami.

Prevádzku portálu zabezpečuje Národná agentúra pre sieťové a elektronické služby (NASES) radiaca sa ako príspevková organizácia Úradu vlády Slovenskej republiky. NASES spustila prevádzku portálu 30. novembra 2015.

K 3. decembru 2018 zverejnilo 72 verejných organizácií 1572 datasetov, z čoho najväčší podiel má Štatistický úrad SR, ktorý vyprodukoval 648 z týchto datasetov, nasleduje ho Úrad pre verejné obstarávanie (73 datasetov) a Ministerstvo pôdohospodárstva a rozvoja vidieka SR (70 datasetov). Ďalšie dve priečky si delia Ministerstvo vnútra SR a mesto Prešov s rovnakým počtom (58) zverejnených datasetov. Iné organizácie už majú počet datasetov menší ako 50. Z pohľadu ministerstiev má najmenej datasetov zverejnených Ministerstvo práce, sociálnych vecí a rodiny SR (5 datasetov) a Ministerstvo obrany SR (6 datasetov). Iné ministerstvá disponujú minimálne 12 datasetmi.

V kapitole 2.2 Merateľnosť kvality otvorených dát sme zoraďovali formáty zverejňovaných datasetov podľa ich kvality tak, ako ich zaraďovali viacerí autori [1, 19]. Na obrázku č.5 môžeme vidieť rozdelenie jednotlivých zverejnených formátov. Jeden dataset môže byť na portáli otvorených dát zverejnený vo viacerých formátoch, preto súčet všetkých záznamov presahuje počet zverejnených datasetov. Najmohutnejšie skupiny datasetov tvoria datasety zverejňované vo formátoch csv a xml (1607 datasetov), ktoré sme definovali ako 3-hviezdičkové formáty. Nemalá skupina datasetov (449) je však zverejňovaných aj vo formáte HTML a PDF, ktoré majú veľmi nízku strojovú spracovateľnosť. Ďalšiu väčšiu skupinu datasetov (309) tvoria formáty z programu Microsoft Excel (xls, xlsx), definované ako 2-hviezdičkový formát. Celý portál obsahuje iba 6 datasetov vo formáte RDF, ktorý predstavuje 4-hviezdičkové dáta.

#### 3.1.2 Centrálny register zmlúv

Centrálny register zmlúv<sup>5</sup> (CRZ) je online portál, na ktorom sa uverejňujú zmluvy uzatvorené vládnymi organizáciami. CRZ eviduje zmluvy od 1. januára 2011 a k 1. decembru 2018 už obsahuje vyše 1 761 000 zmlúv. Zverejňované sú nielen platné zmluvy, no taktiež dodatky k týmto zmluvám, zrušené zmluvy a doplnené zmluvy. Prevádzkovateľom portálu je Úrad vlády Slovenskej republiky, ktorý zabezpečuje prevádzku, kontrolu, bezpečnosť, vývoj a technickú podporu, no za obsah stránok jednotlivých zmlúv, tak ako aj za

---

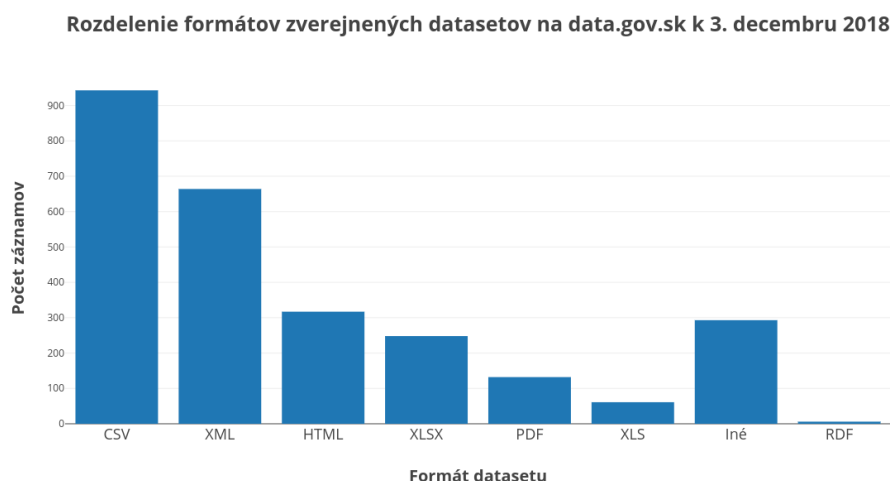
<sup>4</sup>Portál otvorených dát, odkaz: <https://data.gov.sk/>

<sup>5</sup>CRZ, Centrálny register zmlúv: <http://www.crz.gov.sk/>



### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---



Obr. 5: Rozloženie formátov datasetov na data.gov.sk

údaje v zmluvách, zodpovedajú jednotlivé vládne organizácie, ktoré zmluvu uzatvárajú.

Pre každú zmluvu je na CRZ vytvorená unikátna stránka a priradené unikátne ID číslo. Vďaka ID číslu dokážeme danú zmluvu jednoduchšie vyhľadávať, odkazovať sa a tiež pristupovať na jej stránku, nakoľko URI stránky konkrétnej zmluvy vyzerá nasledovne:

*<http://www.crz.gov.sk/index.php?ID=3784379>*

kde ID identifikátor predstavuje práve číslo zmluvy, v našom prípade má zmluva ID číslo 3784379. Zmluvy majú na portáli zverejnené tieto údaje: meno objednávateľa, dodávateľa, ich adresy, identifikačné IČO čísla, ďalej dátumy (dátum zverejnenia, uzavretia, účinnosti, platnosti a poslednej zmeny), názov, typ a stav zmluvy, ID číslo priradené z CRZ, dohodnutá čiastka a celková čiastka, vystavovateľa zmluvy, počet zobrazení zmluvy na CRZ za posledných 30 dní a ako stiahnuteľnú prílohu samotnú zmluvu (formát pdf buď naskenovanej zmluvy alebo zmluvy v elektronickej podobe).

CRZ tiež obsahuje podporu pre automatické sťahovanie a strojové spracovanie, každý deň ráno (približne o 2:00) zverejní stiahnuteľný zip súbor obsahujúci informácie o zverejnených zmluvách za predošlý deň. Zverejnený zip súbor obsahuje jeden xml súbor zahŕňajúci nové zmluvy. Tento súbor sa dá stiahnuť na nasledujúcej adrese:

*<http://www.crz.gov.sk/export/rrrr-mm-dd.zip>*

kde sa rok-mesiac-deň nahradia želaným dátumom. Prvý stiahnuteľný záznam je z dátumu 2011-01-01, a teda od začiatku funkčnosti portálu.

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---

CRZ taktiež na technickej stránke<sup>6</sup> definuje šablónu zverejňovaného xml, no tá sa žiaľ nezhoduje s aktuálne zverejňovanými xml súbormi, aktuálny formát zverejňovaného xml dokumentu je uvedený v ukážke kódu 1.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <zmluvy>
3    <zmluva>
4      <nazov>3002146</nazov>
5      <ID>1478404</ID>
6      <zs1>Pranie a zehlenie, s. r. o. </zs1>
7      <zs2>Bratislavská teplarenská, a.s.</zs2>
8      <predmet>Zmluva o dodávke a odbere tepla </predmet>
9      <datum_ucinnost>2014-07-10</datum_ucinnost>
10     <datum_platnost_do>0000-00-00</datum_platnost_do>
11     <suma_zmluva>0</suma_zmluva>
12     <suma_spolu>0</suma_spolu>
13     <id>0</id>
14     <poznámka/>
15     <rezort>114499</rezort>
16     <datum_zverejnene>2014-08-12 11:38:31</datum_zverejnene>
17     <ico>35823542</ico>
18     <stav>2</stav>
19     ...
20   </zmluva>
21   ...
22 </zmluvy>
```

Listing 1: Ukážka xml súboru zverejneného v CRZ

#### 3.1.3 Register a identifikátor právnických osôb, podnikateľov a orgánov verejnej moci

Register a identifikátor právnických osôb, podnikateľov a orgánov verejnej moci<sup>7</sup> (RPO) predstavuje agregatívny portál, ktorý združuje informácie o všetkých podnikateľoch, právnických osobách a orgánoch verejnej moci z viacerých iných menších registrov (napr. Obchodný register SR, Živnostenský register SR, Register politických strán a hnutí a ďalších vládnych registrov).

Správcom tohoto portálu je Štatistický úrad Slovenskej republiky, ktorý na použitie sprístupnených údajov uplatňuje medzinárodnú verejnú licenciu Creative Commons Attribution License (cc-by) 4.0.

---

<sup>6</sup>CRZ, štruktúra xml: [http://www.crz.gov.sk/data/files/42\\_crz.xsd](http://www.crz.gov.sk/data/files/42_crz.xsd)

<sup>7</sup>RPO, odkaz: <https://rpo.statistics.sk>

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---

RPO ponúka tieto údaje o právnych subjektoch: identifikačné IČO číslo, obchodné meno, adresu, právnu formu, dátum vzniku, predmet činnosti, štatutárny orgán, zainteresovanú osobu, oprávnenie konať, základné imanie, vklady, taktiež poskytuje informáciu o zdrojovom registri a dátum aktualizácie údajov.

RPO vznikol 1. novembra 2015, obsahuje viac ako 1,4 milióna záznamov o právnických osobách, no momentálne neposkytuje žiadne nástroje pre jednoduché strojové spracovanie, avšak v časti 3.1.5 Projekty vzniknuté na základe VOD na Slovensku budeme hovoriť o projekte Slovensko.digital, ktoré vytvorilo verejné API na poskytovanie služieb RPO v strojovo spracovateľnej podobe.

#### 3.1.4 Register partnerov verejného sektora

Register partnerov verejného sektora<sup>8</sup> (RPVS) je registrom, ktorý združuje informácie o partneroch verejného sektora. RPVS definuje partnera verejného sektora takto:

*"Partnerom verejného sektora je fyzická osoba a právnická osoba, ktorá prijíma od štátu, samosprávy a iných subjektov verejného sektora peňažné plnenie alebo majetok nad zákonom určený limit".*

RPVS je spravovaný Ministerstvom spravodlivosti Slovenskej republiky<sup>9</sup> (MSSR). Od 1. februára 2017 agreguje taktiež dáta z predošlého registra konečných užívateľov výhod, ktorých MSSR definuje podľa §6 zákona č. 297/2008 Z. z. ako

*"fyzické osoby, ktoré skutočne ovládajú alebo kontrolujú právnickú osobu, fyzickú osobu – podnikateľa alebo združenie majetku a každú fyzickú osobu, v prospech ktorej tieto subjekty vykonávajú svoju činnosť alebo obchod".*

Okrem konečných užívateľov výhod (KUV) a partnerov verejného sektora eviduje RPVS ešte aj verejných funkcionárov, verifikačné dokumenty (o zápise konečného užívateľa výhod do RPVS) a oprávnené osoby. Celkovo portál obsahuje vyše 25 200 záznamov.

V rámci VOD ponúka RPVS strojovo spracovateľné otvorené REST API rozhranie.<sup>10</sup> Dopyty sú špecifikované pre každú cieľovú skupinu osôb, ktorú chceme z registra získať (konečný užívateľia výhod, partneri verejného sektora, ...). Pre každú skupinu máme na výber 2 typy dopytov:

1. dopyt, ktorý vracia zoznam výsledkov. Pri tomto dopyte si môžeme

---

<sup>8</sup>RPVS, Register partnerov verejného sektora: <https://rpvs.gov.sk>

<sup>9</sup>Ministerstvo spravodlivosti SR, odkaz: <https://www.justice.gov.sk>

<sup>10</sup>RPVS REST API: <https://rpvs.gov.sk/OpenData/swagger/>

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---

zadávať tri kritéria: (1) skip - počet záznamov koľko chceme preskočiť od prvého výsledku, (2) filter - filtrovanie záznamov podľa atribútov, (3) expand - atribúty pre vnorenie sa hlbšie vo výsledkoch. Kritéria pridávame za seba postupne s oddeľovačom "\$"

GET: .../KonecniUzivatelVyhod?\$filter=CisloVlozky eq 117

2. dopyt, ktorý vracia konkrétny výsledok. Pri dopytovaní sa na konkrétny záznam môžeme zadávať dve kritériá: (1) id - ID číslo záznamu (dokážeme zistiť z volania predošlého dopytu), (2) expand - atribúty pre vnorenie sa hlbšie vo výsledkoch. Kritéria pridávame za seba postupne s oddeľovačom "\$"

GET: .../KonecniUzivatelVyhod(283)?\$expand=Partner

Názov "*KonecniUzivatelVyhod*" sa mení podľa želanej skupiny a tri bodky sa v dopytoch nahradia za <https://rpvs.gov.sk/OpenData>.

V oboch prípadoch dopytov dostaneme ako odpoveď json. Ukážka kódu č.2 zobrazuje odpoveď volania druhého ukázkového dopytu.

```
1  {
2    "@odata.context": "..#KonecniUzivatelVyhod",
3    "value": [
4      {
5        "Id": 283,
6        "Meno": "Adriana",
7        "Priezvisko": "Brunnerová Schweinzer",
8        "DatumNarodenia": null,
9        "TitulPred": null,
10       "TitulZa": null,
11       "JeVerejnyCinitel": false,
12       "ObchodneMeno": null,
13       "Ico": null,
14       "PlatnostOd": "2017-02-01T00:00:00+01:00",
15       "PlatnostDo": "2017-09-25T00:00:00+02:00",
16       "Partner": {
17         "Id": 117,
18         "CisloVlozky": 117
19       }
20     ]
21   }
22 }
```

Listing 2: Ukážka JSON výsledku z RPVS API

#### 3.1.5 Projekty vzniknuté na základe VOD na Slovensku

Na Slovensku vďaka zverejneným VOD vzniklo niekoľko projektov s rôznymi zámermi, pracujúcich s rôznymi zverejnenými datasetmi a slúžiacich rôznym skupinám používateľov. Niektoré slovenské VOD projekty:

- *Slovensko.digital*.<sup>11</sup> Občianske združenie so snahou zvýšiť kvalitu VOD na Slovensku.
- *Otvorenezmluvy.sk*. Školský tímový projekt prepájajúci viaceré verejné registre.
- *Finstat.sk*.<sup>12</sup> Komerčné využitie VOD poskytujúce informácie o právnych subjektoch.
- *OpenSpending mesto Prešov*.<sup>13</sup> OpenSpending platforma aplikovaná na VOD z mesta Prešov.
- *Verejne.digital*.<sup>14</sup> Trénovanie umelej inteligencie na slovenských VOD.
- *Uvostat.sk*.<sup>15</sup> Verejné obstarávania v prehľadnej forme.
- *Corobiapolitici.sk*.<sup>16</sup> Štatistika práce politikov a politických strán.
- ďalšie projekty je možné nájsť na stránkach *utopia.sk*<sup>17</sup>

V nasledujúcej časti bližšie popisujeme projekty Slovensko.digital, Finstat.sk a OpenSpending mesta Prešov.

#### Slovensko.digital

Slovensko.digital ako občianske združenie vytvorilo už niekoľko projektov v oblasti VOD a eGovernmentu a aktívne sa zapája do digitalizácie na Slovensku. Svojimi projektami v oblasti VOD vytvorili tzv. ekosystém projektov, skrze ktoré vystupujú ako sprostredkovatelia dát. Medzi ich hlavné projekty v otvorených dátach patria:

- *Register právnických osôb*<sup>18</sup> predstavuje otvorené API zo spracovaného RPO opísaného v kapitole 3.1.3 Register a identifikátor právnických osôb, podnikateľov a orgánov verejnej moci. Tento projekt zjednodušuje strojovú spracovateľnosť údajov z RPO jednoduchými volaniami API dopytov v tvare:

---

<sup>11</sup>Slovensko.digital, odkaz: <https://slovensko.digital>

<sup>12</sup>Finstat, odkaz: <https://finstat.sk>

<sup>13</sup>OpenSpending Prešov, odkaz: <https://openspending.org/s/?q=presov>

<sup>14</sup>Verejne.digital, odkaz: <https://verejne.digital>

<sup>15</sup>Uvostat, odkaz: <https://uvostat.sk>

<sup>16</sup>Corobiapolitici, odkaz: <http://corobiapolitici.sk/>

<sup>17</sup>Ďalšie VOD projekty: <https://utopia.sk/wiki/display/opendata/SK+a+CZ+opendata+aplikacie+a+projekty>

<sup>18</sup>RPO Slovensko.digital: <https://ekosystem.slovensko.digital/otvorene-api#rpo>

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---

```
GET .../api/data/rpo/organizations/:id
```

kde id predstavuje IČO identifikátor právnickej osoby. Ako odpoveď na dopyt dostaneme json súbor s údajmi, ktoré ponúka RPO o danej právnickej osobe. Slovensko.digital ponúka taktiež na stiahnutie celú databázu záznamov z RPO, takže netreba používať API, stačí sa dopytovať do vlastnej databázy.

- *Centrálny register zmlúv*<sup>19</sup> podobne ako predošlý projekt predstavuje spracovaný verejný register, v tomto prípade Centrálny register zmlúv opísaný v kapitole 3.1.2 Centrálny register zmlúv. Slovensko.digital v tomto prípade vytvorilo API, v ktorom sa vieme dopytovať na konkrétnu zmluvu podľa jej ID čísla, ktoré získalo v CRZ. API dopyt sa vytvára v tvare:

```
GET .../api/data/crz/contracts/:id
```

Odpoveďou takéhoto dopytu je json súbor obsahujúci informácie o danej zmluve tak, ako sú uvedené v registri CRZ. Aj v tomto prípade ponúka Slovensko.digital možnosť stiahnuť si celú databázu záznamov.

- *Synchronizačné API*<sup>20</sup> vytvorilo Slovensko.digital pre všetky nimi spracované registre pre aktualizáciu údajov v stiahnutých databázach. Synchronizačné API je vhodné využiť ak máme stiahnutú databázu záznamov, napr. CRZ a chceme aktualizovať naše záznamy. Dopyt si môžeme vyskladať podľa našich preferencií troma premennými, a to: since - čas poslednej zmeny, last\_id - identifikátor záznamu, only\_ids - obmedzí a zrýchli odpoveď, v ktorej dostaneme iba základné údaje o zmene. API teda vyzerá nasledovne:

```
GET .../api/data/<zdroj>/sync{?since,last_id,only_ids}
```

namiesto zdroja sa v dopyte vyberie želaný zdroj dát (rpo, crz,...).

V každom z dopytov sa namiesto troch bodiek dopĺňa adresa <https://datahub.ekosystem.slovensko.digital>. Okrem týchto projektov má Slovensko.digital ešte projekty ako Vestník verejného obstarávania, dopyty na základe SQL, Obchodný vestník a ďalšie<sup>21</sup>, no tieto sú zatiaľ iba v alfa fáze.

#### Finstat.sk

Finstat.sk predstavuje sprostredkovateľa VOD, ktorý agreguje viaceré verejné registre dát a poskytuje množstvo prehľadných štatistík o právnických osobách svojim zákazníkom. Svoje služby poskytuje okrem platenej aj

---

<sup>19</sup>CRZ Slovensko.digital: <https://ekosystem.slovensko.digital/otvorene-api#crz>

<sup>20</sup>Sync API <https://ekosystem.slovensko.digital/otvorene-api#sync-api>

<sup>21</sup>Ďalšie projekty Slovensko.digital: <https://ekosystem.slovensko.digital/>

### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

---

v bezplatnej verzii. Na svojej stránke uvádza, že agreguje dáta z Ministerstva financií SR, Ministerstva spravodlivosti SR (RPVS), Štatistického úradu (RPO), zdravotných poisťovní, Sociálnej poisťovne, Slovenskej komory exekútorov, taktiež spracúva dáta z Českého štatistického úradu.

Okrem zobrazovania štatistík ako napríklad najziskovejšie firmy, počet nových firiem, firmy s najväčšími tržbami, najväčším počtom zamestnancov a rôznymi ďalšími informáciami<sup>22</sup>, poskytuje finstat aj možnosti jednoduchšieho strojového spracovania verejných registrov, a to:

- API<sup>23</sup> - finstat poskytuje veľké množstvo API, ktoré radí do troch kategórií: základné(napr. dopyt na číslo vložky právnickej osoby), rozšírené(napr. dopyt na finančné ukazovatele) a ultimate(napr. zoznam webových adries danej osoby). Väčšina dopytov sa tvorí podľa IČO identifikátora právnickej osoby.
- databázy na stiahnutie<sup>24</sup> - finstat taktiež podporuje sťahovanie celých datasetov obsahujúcich informácie o právnických osobách, finančných ukazovateľoch, konkurzov a reštrukturalizácií a mnohých ďalších.

Obe tieto služby sú spoplatnené a podľa zložitosti služby rozdelené do troch cenových skupín. Viac informácií ohľadom cenníka ako aj o ďalších produktoch a službách finstatu je možné nájsť na ich stránkach uvedených vyššie.

#### OpenSpending mesto Prešov

OpenSpending<sup>25</sup> predstavuje otvorenú platformu, vytvorenú ako jeden z projektov organizácie Open Knowledge International<sup>26</sup> venujúcej sa VOD. OpenSpending poskytuje možnosť zverejniť dataset VOD, na ktorom následne platforma dokáže vytvárať vizualizáciu dát. Takúto vizualizáciu si vie používateľ nastaviť úplne podľa seba.

OpenSpending mesto Prešov<sup>27</sup> teda predstavuje projekt, kedy sa zverejnený dataset z Portálu otvorených dát (data.gov.sk) spracoval a uverejnil vo vyčistenej podobe na platforme OpenSpending. Použitý dataset bol zverejnený mestom Prešov hneď vo viacerých formátoch (xml, json, xls a pdf). Ide o dataset popisujúci programový položkový rozpočet mesta a jeho organizácií, okrem funkcionálnych atribútov (napr. ID, timestamp,..) obsahuje tieto údaje: názov položky rozpočtu, rok, druh rozpočtu, program a jeho číslo v rámci rozpočtu, údaje o klasifikácii položky (napr. všeobecný materiál, bežné výdavky, tovary,..), názov akcie a schválenú sumu spoločne so skutočnou čerpanou sumou. Detailnejší popis atribútov je možné nájsť na

---

<sup>22</sup>Finstat.sk štatistiky: <https://finstat.sk/vsetky-sekcie#nase-analyzy>

<sup>23</sup>Finstat.sk API: <https://finstat.sk/api>

<sup>24</sup>Finstat.sk databázy: <https://finstat.sk/vsetky-sekcie#nase-databazy>

<sup>25</sup>OpenSpending, odkaz: <https://openspending.org>

<sup>26</sup>Open Knowledge International: <https://okfn.org/>

<sup>27</sup>OpenSpending mesto Prešov: <https://openspending.org/s/?q=presov>

stránkach Portálu otvorených dát<sup>28</sup>. Štatistiky o rozpočte mesta Prešov sú tak objaviteľné medzi ostatnými mestami zverejnenými na OpenSpending platforme. Pôvodný aj spracovaný dataset je možné stiahnuť vo formátoch csv a a json priamo zo stránky projektu.

Platforma OpenSpending je trochu ťažšia pre zvyknutie si na prácu s ňou, avšak ponúka veľkú pridanú hodnotu, nakoľko združuje informácie z viac ako 3150 datasetov z 80 krajín po celom svete.

## 3.2 Verejné otvorené dáta v Taliansku

V kapitole 3 Verejné otvorené dáta vo svete a na Slovensku sme videli, že talianske VOD v roku 2017 predbehli vývoj tých slovenských, v tejto kapitole sa pozrieme na 2 veľké talianske registre VOD, aké datasety ponúkajú a v akých formátoch.

### 3.2.1 Taliansky portál otvorených dát

Talianska verzia Portálu otvorených dát nesie názov data.gov.it<sup>29</sup>, ktorý podobne predstavuje katalóg zverejnených datasetov z verejných inštitúcií. Prevádzku portálu zabezpečuje Talianska digitálna agentúra (Digital Agency for Italy). Portál vznikol v roku 2011, odvtedy prešiel niekoľkými zmenami až do dnešnej podoby.

Portál obsahuje k 3. decembru 2018 vyše 23200 datasetov, ktoré portál pekne klasifikuje do rôznych kategórií. Kategórie s najviac zverejnenými datasetmi sú "obyvateľstvo a spoločnosť" (Popolazione e società, 2136 datasetov), "životné prostredie" (Ambiente, 1970 datasetov) a tretie patrí kategórii "vláda a verejný sektor" (Governo e settore pubblico, 1354 datasetov). Rozdelenie formátov zverejnených datasetov môžeme vidieť na obrázku č.6, formát csv (3 hviezdičky podľa Berners-Leeho schémy[1]) je najzastúpenejšou skupinou datasetov s počtom viac ako 18490 datasetov, nasleduje formát json (taktiež s tromi hviezdičkami) s počtom 8300 datasetov, xml (3 hviezdičky) s počtom 6100 datasetov a postupne ďalšie. Zaujímavosťou je, že až vyše 3890 datasetov je zverejnených vo formáte RDF, ktorý sme definovali ako 4-hviezdičkový formát.

### 3.2.2 OpenCoesione register projektov

Reggi a Dawes[12] uvádzajú, že Taliansko je s hodnotou 42,7 miliardy € pre roky 2014-2020 druhý najväčší príjemca dotácií z fondov Európskej únie, presnejšie z fondu ESIFs, ktorý spolufinancuje rozvoj infraštruktúry, obnoviteľných energií, vodných zásob a tiež rozvoj nových podnikov. Taliansko

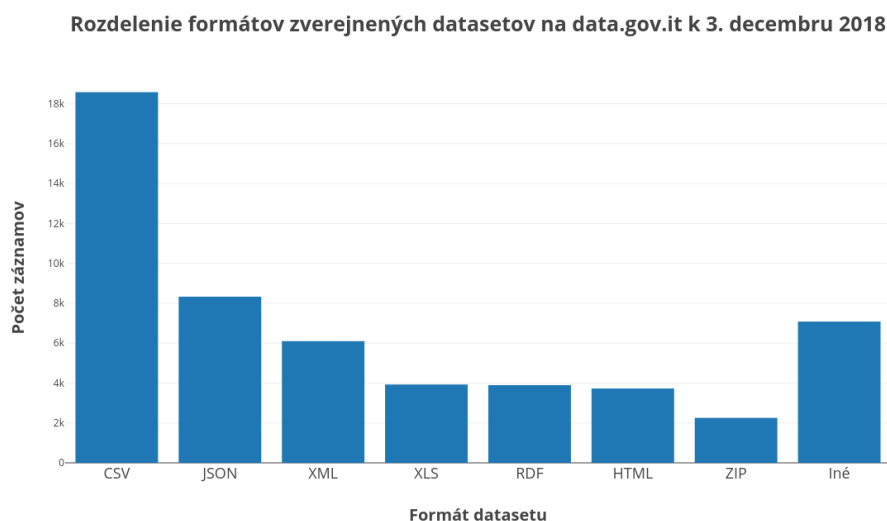
---

<sup>28</sup>Dataset prešov: <https://data.gov.sk/dataset/programovy-polozkovity-rozpocet>

<sup>29</sup>Data.gov.it: <https://data.gov.it>



### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU



Obr. 6: Rozloženie formátov datasetov na data.gov.it

samotné ešte prispieva k tejto čiastke a navyšuje ju na 123 miliárd € zameranú na rozvoj.

Verejný otvorený register OpenCoesione<sup>30</sup> sleduje využívanie týchto dotácií a podrobne dozerá na každý z projektov. Pre každý projekt ponúka detailné informácie o cieľových sektoroch, ktoré dotácie prijali, čo sa s dotáciami pre daný projekt stalo, ktoré právnické osoby získali najviac z takýchto dotácií a ktoré najviac pracovali na projekte a mnohé ďalšie štatistiky dostupné pri otvorení konkrétneho projektu na stránkach registra.

Register vznikol v roku 2012 ako projekt Ministerstva rozvoja ekonomiky (Ministero dello sviluppo economico) a k 4. decembru 2018 monitoruje viac ako 1 124 000 projektov a viac ako 122 miliárd € v dotáciách.

V rámci strojového spracovania poskytuje OpenCoesione otvorené API<sup>31</sup> a možnosť stiahnutia celej databázy. Otvorené API je implementované na vysokej úrovni, používateľ si môže zvoliť formát výsledku, v závislosti od voľby dostane odpoveď buď prehľadnú html stránku, alebo json hromadnú odpoveď. Základný API dopyt, ktorý sa používa pre získanie informácií o všetkých možných API dostupných na portály vyzerá nasledovne:

GET <https://opencoesione.gov.it/it/api/?format=api>

premennú formát na konci môžeme zmeniť na json, vtedy dostaneme ako odpoveď json so všetkými dostupnými API (výsledok možno vidieť na ukážke kódu č.3).

<sup>30</sup>OpenCoesione: <https://opencoesione.gov.it>

<sup>31</sup>OpenCoesione API: <https://opencoesione.gov.it/it/api/>



### 3. VEREJNÉ OTVORENÉ DÁTA VO SVETE A NA SLOVENSKU

Register	CRZ	RPO	RPVS	OC
Zverejňované dáta	Zmluvy	Právnické osoby	Partneri VS & KUV	Projekty
Strojové spracovanie	XML	Žiadne*	API	API + CSV
Počet záznamov	1761000	1400000	25200	1124000
Dátum vzniku	1.1.2011	1.11.2015	1.2.2017	2012

Tabuľka 3: Porovnanie verejných registrov

[illegible][illegible]

Sprostredkovateľ	Slovensko.digital	Finstat	OS Prešov
Používaný dataset	RPO, CRZ,..	RPO, RPVS,..	položkový rozpočet
Aké datasety boli prepájané	Otvorené dáta	Poistovne,..	Žiadne
Strojové spracovanie	API, CSV, SQL	API, CSV	CSV, JSON
poznámka	Viacere projekty, záleží na projekte	Poskytuje aj cenné informácie	Dá sa stiahnuť dataset

Tabuľka 4: Porovnanie VOD projektov

Dobree to uz je skoro cele hotove :))Dobree to uz je skoro cele hotove

[illegible]

## 4 Dataset Ministerstva vnútra SR

#AllForJan Hackathon priniesol, okrem iných datasetov, aj dataset objednávok a faktúr Ministerstva vnútra Slovenskej Republiky (MV SR), s ktorým sa zaoberá naša práca. Tento dataset získal Ing. Ivan Srba, PhD. postupným sťahovaním záznamov zo stránok MV SR. Kolega Martin Tonhauzer rozpracováva v rovnomenne pomenovej práci dataset faktúr, v našej práci je viac rozpracovaný dataset objednávok.

**TODO: obrázok ako vyzerá spojenie medzi objednávkami, faktúrami a ich položkami.**

Je zverejňovaný na stránkach MV SR, nie je k nemu však poskytovaná žiadna možnosť automatického sťahovania ( tak ako pri Centrálnom registri zmlúv 3.1.2 ) ani žiadne otvorené API ( tak ako pri Registri partnerov verejného sektora 3.1.4 ).

## 5 Návrh spracovania získaných VOD

Problém s chýbajúcimi objednávkami pre faktúry, problém s vychýlenými hodnotami, nie správne vypočítanými + ako by sa to dalo riešiť?

## 6 Spracovanie a zobrazenie VOD

Parsovali sme, crawlovali sme, pushujeme do DB na serveri, kde beží meta-base. Hlavne, čistili sme!

## 7 Zhodnotenie a záver

Ako sa podarilo dosiahnuť čo bolo naplánované v zimnom semestri.. Dobré to je :)

## 8 Zdroje

### Literatúra

- [1] Linked data. <https://www.w3.org/DesignIssues/LinkedData.html>. Accessed: 2018-11-30.
- [2] The open definition. <http://opendefinition.org>. Accessed: 2018-11-22.
- [3] What is open data? <http://opendatahandbook.org/guide/en/what-is-open-data>. Accessed: 2018-11-22.
- [4] T. Dalgleish, J. M. G. Williams, A.-M. J. Golden, N. Perkins, L. F. Barrett, P. J. Barnard, C. Au Yeung, V. Murphy, R. Elward, K. Tchanturia, and E. Watkins. Open data maturity in Europe. *Journal of Experimental Psychology: General*, 136(1):23–42, 2007.
- [5] T. Davies and Z. Bawa. The promises and perils of open government data (ogd). *The Journal of Community Informatics*, 8(2), Apr. 2012.
- [6] F. González-Zapata and R. Heeks. Understanding Multiple Roles of Intermediaries in Open Government Data. *13th International Conference on Social Implications of Computers in Developing Countries*, (May 2015):1–10, 2015.
- [7] E. Hyv. The Semantic Web: ESWC 2012 Satellite Events. 7540:226–230, 2015.
- [8] M. Janssen. Transparency - by - Design as a Foundation for Open Government. 2017.

- [9] M. Janssen and A. Zuiderwijk. Janssen Zuiderwijk (2014) - Infomediary Business Models for Connecting Open Data Providers and Users.pdf.
- [10] Y. C. &. A. Z. Marijn Janssen. Benefits, Adoption Barriers and Myths of Open Data and Open Government. 2012.
- [11] N. Rajshree and B. Srivastava. Open Government Data for Tackling Corruption – A Perspective Case Study : NREGA in India. *2012 AAAI Workshop - Semantic Cities*, pages 21–24, 2012.
- [12] L. Reggi and S. Dawes. Open Government Data Ecosystems : Linking Transparency for Innovation with Transparency for Participation and Accountability. 1:74–86.
- [13] I. Safarov, A. Meijer, and S. Grimmelikhuijsen. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, 22(1):1–24, 2017.
- [14] N. Shadbolt, K. O. Hara, T. Berners-lee, N. Gibbins, and H. Glaser. Government Data : Lessons from. pages 2–10, 2007.
- [15] R. E. Sieber and P. A. Johnson. Civic open data at a crossroads : Dominant models and current challenges. *Government Information Quarterly*, 2015.
- [16] Sunlight Foundation. Guidelines for Open Data Policies. (March):1–13, 2014.
- [17] B. Ubaldi. Open Government Data Towards Empirical Analysis of Open Government Data Initiatives. (22), 2013.
- [18] A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando. Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2):325–337, 2016.
- [19] G. Viscusi, B. Spahiu, A. Maurino, and C. Batini. Compliance with open government data policies: An empirical assessment of Italian local public administrations. *Information Polity*, 19(3-4):263–275, 2014.
- [20] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. Alibanks. Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2):156–172, 2012.

## 9 Prílohy

...

### 9.1 Časový harmonogram na prácu v druhom semestri

## 10 Hot ToDos

čo treba spraviť určite do BP1

- Ujednotiť názvy pre všeobecný názov ministerstiev, orgánov a úradov, niečo ako Ústredný orgán štátnej správy len jednoduchšie aby tomu čitateľ pochopil
- Pridať do 2.1 že VOD sa bežne ukladajú na jednotný portál
- Tabuľka o výhodách OGD + možno aj challenges, to by sa hodilo asi aj spraviť ako agregát do diskusie
- Celkovú sémantiku zmeniť, opraviť tak, aby pekne na seba nadväzovali kapitoly
- ak budem porovnávať ešte na základe nejakých ďalších kvalitatívnych metrík, treba to pripísať do 2.2 Merateľnosť kvality otvorených dát
- ešte možný článok ako príklad prepájania DBpedia  
<https://link.springer.com/content/pdf/10.1007%2F978-3-540-76298-0.pdf> page 754 Alebo možno skôr nejaký obrázok nakresliť kde by sa dal zastihnúť potenciál prepojenia
- používať všade "verejné organizácieä v úvode definovať čo pod tým myslím
- CHECK - done. Definovať ktoré formáty spadajú do strojovo spracovateľných, aby to potom v porovnaní bolo jasné čo je dobré a čo nie
- formát anotácie má byť iný
- napísať že zdrojom dát v trojke sú samotné portály a vlastné zistenie
- CHECK - netreba. príklady dopytov bude asi lepšie zmeniť na obrázky, aby to bolo aj krajšie
- CHECK - netreba. prefarbiť ukážky kódu na jednotnú farbu
- otázne je, či by sa nemal pridať ešte popis aj ministerstva vnútra v rámci slovenského VOD. Asi áno, no nie do detailov, iba tak, že čo všetko ponúkajú na stránke
- dať rovnakú štruktúru poznámkam pod čiarou