Strategická priorita

**Otvorené údaje**

Pracovná verzia

(Verzia 0-9)

Informácia o dokumente

|  |  |
| --- | --- |
| Názov: | Strategická priorita Otvorené údaje |
| Stav: | Pracovná verzia |
| Pripravil: | Pracovná skupina K9.4 Lepšie dáta |
| Verzia: | 0.9 |
| Dátum: | 10.09.2017 |
| Pripomienkoval: |  |
| Dátum revízie: |  |

Distribučný zoznam

| Od | Dátum | Kontakt |
| --- | --- | --- |
| K9.4 |  |  |

| Pre | Akcia\* | Dátum (do) | Kontakt |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |

\* Akcia: Schváliť, Pripomienkovať, Informovať, Realizovať, iné (uveďte)

História verzií

| Verzia | Dátum verzie | Pripravil/ Zmenil | Pripomienkoval | Kľúčové zmeny |
| --- | --- | --- | --- | --- |
| 0.9 | 10.09.2017 | K9.4 |  | Prvá verzia |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Obsah

[1 Úvodný pohľad 4](#_Toc495334249)

[1.1 Cieľ dokumentu 4](#_Toc495334250)

[1.2 Definícia strategickej priority 4](#_Toc495334251)

[1.2.1 Kontext pre strategickú prioritu Manažment údajov 4](#_Toc495334252)

[1.2.2 Vysvetlenie strategickej priority 5](#_Toc495334254)

[1.2.3 Základné úlohy 6](#_Toc495334258)

[1.2.4 Základné témy 7](#_Toc495334259)

[1.3 Analýza súčasného stavu 7](#_Toc495334264)

[1.3.1 Modul otvorených dát (MOD) 7](#_Toc495334272)

[1.3.2 Interoperabilita otvorených údajov 9](#_Toc495334273)

[1.3.3 LOD Slovakia (Linked Open Data) 10](#_Toc495334275)

[2 Ciele realizácie 11](#_Toc495334278)

[2.1 Architektonické ciele 11](#_Toc495334279)

[2.1.1 Komplexný pokrok v téme otvorených údajov 11](#_Toc495334280)

[2.1.2 Zlepšiť dostupnosť údajov verejnej správy vo forme otvorených údajov 11](#_Toc495334281)

[2.1.3 Kvalita publikovaných údajov 13](#_Toc495334282)

[2.1.4 Zvýšiť dynamiku a prínos dátového hospodárstva 14](#_Toc495334360)

[2.2 Dotknuté subjekty a ich záujmy 15](#_Toc495334361)

[3 Organizácia 18](#_Toc495334362)

[3.1 Zodpovednosť 18](#_Toc495334363)

[3.2 Posúdenie kapacitnej pripravenosti 19](#_Toc495334364)

[3.2.1 Personálna kapacita 19](#_Toc495334365)

[3.2.2 Kapacita informačných systémov 19](#_Toc495334366)

[3.3 Organizačné zmeny 19](#_Toc495334367)

[4 Stratégia 21](#_Toc495334368)

[4.1 SWOT analýza 21](#_Toc495334369)

[4.2 Vízia 22](#_Toc495334370)

[4.3 Výber prístupu 23](#_Toc495334372)

[4.3.1 Otvorené údaje a prístup k dátovému hospodárstvu 24](#_Toc495334374)

[4.3.2 Ako vytvoriť funkčnú komunitu 24](#_Toc495334375)

[4.3.3 Model pre „governance“ otvorených údajov 26](#_Toc495334376)

[4.3.4 Spôsob zabezpečenia kvality otvorených údajov 27](#_Toc495334381)

[4.3.5 Spôsob prístupu (access) k otvoreným údajom 28](#_Toc495334382)

[Výber formátov pre sprístupnené otvorené údaje 28](#_Toc495334387)

[4.3.6 28](#_Toc495334388)

[4.3.7 Miera centralizácie 29](#_Toc495334389)

[4.3.8 Výber licenčných modelov 29](#_Toc495334390)

[4.3.9 Pôvod údajov 30](#_Toc495334391)

[4.3.10 Ochrana súkromia a otvorené údaje 31](#_Toc495334394)

[4.3.11 Spôsob zabezpečenia dôveryhodnosti 33](#_Toc495334396)

[4.4 Prehľad najlepších skúseností 33](#_Toc495334397)

[5 Architektúra riešenia 35](#_Toc495334398)

[5.1 Popis datasetov 35](#_Toc495334399)

[5.2 Princípy 35](#_Toc495334400)

[Pravidlá pre interoperabilitu otvorených údajov verejnej správy SR 36](#_Toc495334401)

[5.2.1 36](#_Toc495334402)

[5.3 Biznis vrstva 37](#_Toc495334404)

[5.3.1 Prehľad požiadaviek 37](#_Toc495334405)

[5.3.2 Prehľad procesov 37](#_Toc495334406)

[5.4 Aplikačná vrstva 40](#_Toc495334407)

[5.4.1 Koncept architektúry 40](#_Toc495334408)

[5.4.2 Prehľad aplikačných funkcií a komponentov 40](#_Toc495334409)

[5.4.3 Aplikačné požiadavky 41](#_Toc495334410)

[5.4.4 Aplikačné služby 41](#_Toc495334411)

[5.4.5 Aplikačné komponenty 42](#_Toc495334412)

[5.5 Technologická vrstva 44](#_Toc495334413)

[6 Realizácia 45](#_Toc495334415)

[6.1 Riziká a otvorené otázky 45](#_Toc495334416)

[6.2 Plánovanie a migrácia (v súlade s akčným plánom) 47](#_Toc495334417)

[6.2.1 Aktivity a úlohy pre centrálnu úroveň 47](#_Toc495334418)

[6.2.2 Aktivity a úlohy pre inštitúcie verejnej správy 49](#_Toc495334427)

[6.2.3 Identifikácia projektov, ktoré je vhodné realizovať 49](#_Toc495334432)

[6.2.4 Požiadavky na legislatívne zmeny 51](#_Toc495334433)

[7 Udržateľnosť a financovanie 54](#_Toc495334434)

[7.1 Zmysluplná investícia 54](#_Toc495334435)

[7.2 Dopady navrhovaných zmien 54](#_Toc495334437)

[7.3 Odporúčania z ohľadom na udržateľnosť 56](#_Toc495334438)

[8 Prílohy 57](#_Toc495334439)

[8.1 Skratky a definície 57](#_Toc495334442)

# Úvodný pohľad

Nasledujúci dokument predstavuje pracovnú verziu Strategickej priority Otvorené údaje. Ide o  výstup pracovnej skupiny K9.4 Lepšie dáta a je možné vidieť návrh riešení. K dokumentu je možné sa ľubovoľne vyjadrovať a rozumné návrhy budú zvážené a zapracované.

## Cieľ dokumentu

Cieľom tohto dokumentu je podrobnejšie rozpracovanie strategickej priority 7 Otvorené údaje: Definovanie základných cieľov v rámci strategickej priority, načrtnutie možných alternatív ich dosiahnutia ako aj výber najvhodnejšej alternatívy na realizáciu. Definované ciele musia priniesť pozitívny vplyv na zvýšenie dostupnosti údajov verejnej správy pre spracovanie tretími stranami na základe princípov definovaných v Národnej koncepcii informatizácie verejnej správy.

## Definícia strategickej priority

### Kontext pre strategickú prioritu Manažment údajov

Národná koncepcia informatizácie verejnej správy z roku 2016 (ďalej tiež „NKIVS“) ustanovuje 10 strategických priorít informatizácie verejnej správy:

1 Multikanálový prístup,

2 Interakcia s verejnou správou, životné situácie a výber služby navigáciou,

3 Integrácia a orchestrácia,

4 Rozvoj agendových informačných systémov,

5 Využívanie centrálnych spoločných blokov,

6Riadenie údajov a Big data (Manažment údajov),

**7 Otvorené údaje,**

8 Vládny cloud,

9 Komunikačná infraštruktúra,

10 Kybernetická bezpečnosť.

NKIVS ku každej strategickej priorite informatizácie verejnej správy vysvetľuje jej cieľ, prístup k riešeniu a tiež rámcový architektonický model. Tento dokument nadväzuje na NKIVS a predstavuje návrh riešenia pre Otvorené údaje vo verejnej správe a rozširuje tak kapitolu **6.2.7 Otvorené údaje.** Tento dokument ešte môže byť z pozície architektonickej kancelárie verejnej správy upravovaný a dopĺňaný aj na základe pripomienok a komunikácie s gestorom tejto strategickej priority.

Strategická priorita Otvorené údaje by sa mala prakticky realizovať ihneď, ako to bude možné (prvé projekty nového programového obdobia). Za základné dôvody je možné považovať najmä:

* Všetky modernizované informačné systémy verejnej správy by mali byť schopné automatizovane publikovať otvorené údaje (s výnimkou údajov, pri ktorých je takéto sprístupnenie vylúčené osobitným zákonom).
* Z hľadiska efektívneho spracovania otvorených údajov sú publikované údaje na nízkej úrovni interoperability, čo podstatne limituje ich akékoľvek použitie. Súčasné publikované dáta nepoužívajú jednotné referencovateľné identifikátory a ani jednotný model údajov, čím sú prakticky všetky údaje samostatné a pri ich použití je ich nutné vždy spájať, respektíve integrovať nanovo.

Táto strategická priorita rozširuje aplikáciu princípu pristupovania k dátam ako k vzácnemu zdroju, pretože zverejňovanie otvorených údajov umožňuje nájsť ďalšie využitie dát aj mimo prostredia verejnej správy. Základným typom zverejňovaných údajov sú takzvané informácie verejnej správy (Public Sector Information - PSI), ktoré verejné orgány vytvárajú, zbierajú alebo za ne platia – teda sú ich vlastníkom alebo správcom. Príkladom sú geografické dáta, štatistiky, informácie o počasí, údaje z výskumných projektov financovaných z verejných zdrojov a digitalizované knihy z knižníc. Priorita predstavuje plynulé pokračovanie aktivít smerujúcich k zverejňovaniu otvorených údajov naprieč verejnou správou v rámci aktuálneho Akčného plánu Iniciatívy pre otvorené vládnutie na roky 2017 až 2019. Navrhuje tiež nadstavbu nad nástrojmi pre zverejňovanie otvorených dát, ktoré sa vybudovali v projekte eDemokracia v rámci Operačného programu Informatizácia spoločnosti (OPIS). Väčšinu zo zaujímavých dát je však v súčasnosti:

* nemožné jednotne referencovať (identifikovať) a používať jednotný model údajov pre zabezpečenie interoperability,
* problematické licencovať pod otvorenou licenciou umožňujúcou jednoduché a bezodplatné opakované použitie aj na komerčné účely.

Dôvodom sú predovšetkým nevysporiadané autorské práva s tvorcami informačných systémov a nedostatočná transpozícia európskej direktívy na opakované využívanie informácií verejného sektora (Directive 2003/98/EC, známa ako 'PSI Directive') do samostatného komplexného zákona.

### Vysvetlenie strategickej priority

Publikovanie otvorených údajov pre opakované použitie predstavuje samostatný špecifický cieľ programového obdobia 2014 až 2020, ktorý sa zameriava hlavne na rozsah a kvalitu publikovaných dát, ako aj na intenzitu vyžívania týchto dát podnikateľmi a kreatívnymi komunitami. Primárnym zámerom je využiť ekonomický potenciál otvorených údajov verejnej správy, ktorých odhadovaná trhová hodnota v Európe je 32 miliárd EUR (čo pre Slovensko znamená približne 250 miliónov EUR).

Strategická priorita je v súlade s princípom pristupovania k dátam ako k vzácnemu zdroju, pretože ich zverejňovanie umožňuje nájsť ich optimálne využite aj mimo prostredia verejnej správy. Očakávané výsledky úspešnej realizácie strategickej priority sú podľa operačného programu nasledovné:

* Vďaka využívaniu otvorených údajov sa výrazne zvýši miera transparentnosti verejnej správy, čím vznikne pozitívny dosah na jej efektivitu;
* Na Slovensku vznikne dostatočne silné odvetvie pracujúce s otvorenými dátami, pričom vzniknú kompetencie v progresívnych oblastiach informačných technológií, ako sú dátové analýzy;
* Inovatívne použitie otvorených údajov vygeneruje pridanú hodnotu pre používateľov, ktorá sa premietne do ekonomického dopadu;
* Vďaka vzájomnej interoperabilite otvorených údajov bude možné s nimi pracovať ako s jedným celkom, to je výrazne sa zvýši ich hodnota, čo sa premietne do kvalitnejších analýz či poskytovaných elektronických služieb.

Otvorené údaje (Open Data) predstavujú údaje, ktoré sú dostupné v strojovo čitateľnom formáte a hocikto ich môže voľne využívať na komerčné aj nekomerčné účely, modifikovať, obohacovať, spájať a zdieľať maximálne s podmienkou uvedenia autora. V kontexte verejnej správy znamenajú otvorené dáta nasledovné:

* pre každú organizáciu je dostupný zverejnený zoznam jej dátových zdrojov, z ktorých možno publikovať otvorené dáta, a ich základný popis,
* údaje sú zverejnené v datasete (zostave údajov), ktorý je popísaný metadátami založených na štandardizovaných dátových prvkov uvedených v MetaIS (Centrálny model údajov verejnej správy) a zaevidovaný v centrálnom katalógu otvorených údajov data.gov.sk,
* obsah dátového zdroja (samotné údaje) sú proaktívne bezplatne sprístupnené neobmedzenému okruhu záujemcov cez referenciu na miesto sprístupnenia datasetu a schémy údajov,
* údaje sú sprístupnené v štruktúrovanom a štandardizovanom tvare umožňujúcom ďalšie strojové spracúvanie, čo sa dosiahne dodržaním stanovených technologických štandardov, ktorých základom je používanie otvorených a technologicky neutrálnych riešení,
* na údaje sa možno právne spoľahnúť, pretože majú presne uvedeného pôvodcu, informáciu o aktuálnosti, o začiatku a konci účinnosti datasetu alebo jeho údajov, najmä ak je účinnosť odlišná od platnosti,
* ďalšiemu používaniu údajov (napríklad ich spracovaniu, zverejneniu, spájaniu s inými údajmi) nie sú kladené žiadne prekážky, ak sú splnené legislatívne požiadavky pre prácu s údajmi.
* Ku každému datasetu musí byť priradená licencia,
* Používateľ údajov musí dodržiavať GDPR,
* inštitúcie verejnej správy musia pri zverejňovaní otvorených údajov dodržiavať osobitné nariadenia EU (GDPR) ako aj ostatné všeobecne záväzné právne predpisy. a získajú nástroje pre anonymizáciu datasetov.

### Základné úlohy

V rámci priority budú rozpracované opatrenia v nasledujúcich oblastiach.

Centrálna koordinácia publikovania otvorených dát:

* Jasne sa zadefinuje centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených dát v centrálnom katalógu data.gov.sk a opakovaným použitím otvorených dát.
* Na identifikáciu údajov VS sa zavedie používanie Jednotných referencovateľných identifikátorov (URI) a na ich popis sa použije Centrálny dátový model verejnej správy, ktorý bude pozostávať z množiny spoločných doménových ontológií (spravovaných centrálne ontológmi v Centrálnej kapacite) spĺňajúce aj medzinárodné štandardy otvorených prepojených dát a pravidlá sémantickej a syntaktickej interoperability v európskom priestore.
* Implementuje sa dereferenciácia pre ľubovoľnú entitu reprezentovanú URI (dátový prvok, dataset, organizácia, číselník, priestorový údaj, …), to je fyzický používateľ alebo strojový spracovateľ poznaním URI dokáže k danému objektu pristúpiť.
* Zavedú sa ďalšie automatizované metódy pre publikovanie otvorených dát vďaka implementácii konsolidovanej analytickej vrstvy a platformy integrácie údajov.
* Zavedie sa dohľad nad kvalitou a štandardizáciou metadát a popisných štruktúr.
* Do praxe sa uvedie platforma pre štandardizované vytváranie a spravovanie aplikačných rozhraní (API) pre strojový prístup k otvoreným dátam.
* Začnú sa aplikovať nové postupy pre riadenie kvality datasetov a rizík ich publikovania, predovšetkým pre prepojené údaje.
* Vytvorí sa centrálna právna kapacita pre riešenie právnych problémov so zverejňovaním a opakovaným použitím.

Zverejňovanie otvorených údajov u povinných osôb:

* Vďaka zavedeniu automatizovaných metód pre publikovanie otvorených dát budú povinné osoby značne odbremenené.
* Vyrieši sa problém efektívneho licencovania datasetov otvorených dát, najmä v oblasti informácií verejného sektora (Public Sector Information). Prideľovanie licencií bude podporovať jednoduché opakované použitie doma aj v zahraničí a v relevantných prípadoch, pri ktorých sa nejedná o PSI, bude účinne chrániť autorské práva.
* Zvýši sa kompetencia pre publikovanie prepojených údajov.
* Budú dostupné neustále sa zlepšujúce a interaktívne metodiky pre publikovanie otvorených údajov.
* Vďaka vysokej interoperabilite publikovaných dát bude možné efektívne pracovať s rôznymi datasetmi publikovanými rôznymi subjektmi ako s jedným celkom.

Zverejňovanie otvorených údajov v samospráve:

* Aplikovanie štandardizovaných formátov otvorených dát a API rozhraní do prostredia samosprávy.
* Vytvorenie centrálneho úložiska datasetov samosprávy v prostredí DCOMu s využitím centrálnych nástrojov pre publikovanie.
* Vytvorenie dátového modelu samosprávy, štandardizujú sa samotné popisné štruktúry prioritných datasetov a metadáta pre samosprávu.
* Koordinovaná tvorba a aktualizácia odporúčaní a metodík pre fungovanie samospráv rôznych veľkostí tak, aby proces publikovania otvorených údajov prebiehal efektívne a výsledné dáta boli prínosom pre spoločnosť.
* Publikované údaje samosprávy sa budú riadiť rovnakými pravidlami interoperability ako údaje štátnej správy, to je na identifikáciu údajov sa používajú Jednotné referencovateľné identifikátory URI a údaje v najvyššej kvalite z pohľadu interoperability sú popísané Centrálnym modelom údajov verejnej správy založeným na ontológiách.

Synergia zverejňovania otvorených priestorových údajov:

* Otvorené priestorové údaje sú všetky priestorové údaje zo zdrojovej evidencie vedenej povinnou osobou pod otvorenou licenciou.
* Dokumentácia otvorených priestorových údajov metaúdajmi bude zabezpečená Registrom priestorových informácií, ktorý bude integrovaný s Centrálnym metainformačným systémom verejnej správy (MetaIS) a Portálom otvorených dát pre zabezpečenie požiadavky 1x a dosť.
* Pre zvýšenie miery sémantickej interoperability budú podporené harmonizačné aktivity prostredníctvom Centrálneho modelu údajov verejnej správy.
* Zvýši sa potenciál vyhľadateľnosti a využitia otvorených priestorových údajov ich prepojením s ďalšími údajmi, nepriestorovej povahy.

Podpora kreatívnych komunít a inovatívnych podnikov:

* Na základe hodnotenia spätnej väzby sa budú publikovať tie otvorené dáta a API, ktoré dokážu podniky a komunity využiť pre lepšie služby spoločnosti, zvyšovanie transparentnosti a pre rast digitálnej ekonomiky.

### Základné témy

Dokument prináša odpovede v nasledujúcich témach:

* Princípy dátového hospodárstva,
* Pravidlá interoperability otvorených údajov,
* Legislatívny pohľad:
* „Open data first“ a vízia zákona o údajoch.
* Pohľad architektúry:
* Čo má byť riešené na centrálnej úrovni a čo na lokálnej úrovni, aké sú kombinácie.
* Pravidlá pre oprávnené používanie otvorených údajov:
* Licenčné modely pre otvorené údaje.
* Ochrana súkromia a otvorené údaje:
* Rozsah zverejňovania údajov,
* Anonymizácia údajov.
* Procesy pre publikovanie otvorených údajov.

## Analýza súčasného stavu

Slovensko sa v roku 2012 prihlásilo k akčnému plánu pre otvorené vládnutie (OGP) a postupne prijíma akčné plány pre progres v danej oblasti. Reálne sa začalo so systematickým riešením otvorených údajov v rámci projektu e-Demokracia a otvorená vláda v rokoch 2013 až 2015, keď bol implementovaný Modul otvorených údajov (MOD).

### Modul otvorených dát (MOD)

V rokoch 2013 až 2015 prebiehal Národný projekt „Elektronické služby Úradu vlády SR – eDemokracia a otvorená vláda“, ktorým bola vytvorená centrálna infraštruktúra pre zverejňovanie otvorených údajov a v jej rámci služby pre poskytovateľov aj používateľov údajov.

Pre poskytovateľov údajov sú určené komplexné služby asistencie pri sprístupňovaní otvorených údajov najmä v nasledovných oblastiach:

* transformácie formátov a štruktúr údajov,
* ukladanie údajov, ich publikácia a riadenie súvisiacich kapacít a bezpečnosti,
* zaistenie súladu, najmä so štandardami informačných systémov verejnej správy a licencie   
  na používanie údajov,
* sprístupnenie metodík, poskytovanie poradenstva,
* podporná aplikácia – pre prácu s otvorenými údajmi na strane poskytovateľa, dostupná bezplatne.

Pre používateľov údajov sú dostupné služby podpory procesov sprístupňovania údajov najmä   
v nasledovných oblastiach:

* katalóg datasetov vrátane metadát,
* prístup k údajom prostredníctvom konzistentných API a formátov,
* nástroje pre základnú prezentáciu údajov,
* podpora komunikácie s používateľmi,
* žiadosti o sprístupnenie údajov,
* evidovanie aplikácií používajúcich otvorené údaje.

Začiatkom marca 2017 bolo v rámci portálu data.gov.sk evidovaných 1 065 datasetov z 35 organizácií.

Komponenty modulu otvorených údajov

Projekt eDemokracia nasadil nasledujúce komponenty, založené na osvedčených medzinárodných praktikách. Pri výbere použitých súčastí boli zároveň brané do úvahy technológie použité v európskom linked data softvérovom stacku LOD2 <http://lod2.eu>.

Tabuľka : Prehľad nasadených komponentov a ich funkcionalita

| **Použité komponenty** | **Skratka** | **Funkcionalita** | **Licencia** |
| --- | --- | --- | --- |
| CKAN | CKAN | Katalóg otvorených dát | Opensource AGPL 3.0 |
| Pool Party | PPT | Ontologický subsystém, údržba riadených slovníkov, extrakcia konceptov | Proprietárna |
| Virtuoso Universal DB | Virtuoso | multi-model relational database management (SQL Tables a/alebo RDF Statement Graphs) | Proprietárna |
| Unified Views | UV | Transformácie dát | Opensource GPL |
| XWiki |  | Crowdsourcing, komunikácia metodík | Opensource LGPL |
| Open data node | ODN | Podporná aplikácia pre povinné osoby | Opensource AGPL, LGPL |

Použité komponenty možno mapovať na nasledujúce komponenty MOD.

Tabuľka : Komponenty MOD

| **Použitý komponent** | **Komponent MOD** |
| --- | --- |
| CKAN | čiastočne Správa katalógov, Repozitár datasetov, čiastočne Prezentačná aplikácia |
| PPT | Ontologický editor, Správa ontológií a identifikátorov |
| Virtuoso | RDF úložisko, čiastočne “SPARQL” Query engine |
| UV | Transformačný komponent, čiastočne Správa transformácií |
| ODN | Podporná aplikácia pre povinné osoby |

Ontologický editor pre prepojené údaje

Ontologický editor umožňuje správu dátových slovníkov (ontológií) , ktoré predstavujú najvyšší stupeň formálnej reprezentácie znalostí. Ontológia poskytuje možnosť definovať triedy, a vlastnosti medzi týmito triedami, pričom prostredníctvom uplatnenia strojového odvodzovania je možné spájať , odvodzovať, respektíve overovať rôzne údaje podľa logických pravidiel.

Ontologický editor spolu s publikačným nástrojom pre XML schémy po zvážení môže vystupovať ako úložisko ontológií a ich elementov na báze linke data.

Transformačný modul

Spravuje a vykonáva transformácie z textových formátov, z databázových zdrojov do RDF, interných relačných dát. Eviduje typy a postupy transformácií. Slúži na zaradenie rôznych zdrojov údajov do katalógu RDF alebo relačných datasetov. Umožňuje zvyšovanie kvality vstupných dát, obohacovanie datasetu, rozdielové analýzy, vyťažovanie neštruktúrovaných súborov.

Podporná aplikácia

Poskytuje Povinnej osobe nasledovné funkcionality:

* funkcionalita na zber (harvestovanie) a transformáciu údajov do open data formátov.
* funkcionalita na internú evidenciu zozbieraných a transformovaných dát (open data, ale aj non-open data) - interný katalóg
* funkcionalita na publikovanie (synchronizáciu) dát z internej evidencie do MODu

Na zabezpečenie týchto funkcionalít sú použité a upravené open source balíky Unified Views, CKAN a iné.

V dobe realizácie projektu MOD sa sémantické dátové štandardy len tvorili, a preto je nutné overiť použitie MODu v súlade s aktuálnymi Sémantickými dátovými štandardmi, a to najmä s použitím pri správe Centrálneho modelu údajov verejnej správy založeného na ontológiách. Keďže jednotlivé ontológie Centrálneho modelu sú štandardizované OWL ontológie, je vysoká pravdepodobnosť že použitie MODu pre ich správu bude bez výrazných obmedzení.

### Interoperabilita otvorených údajov

Obmedzenia v interoperabilite súčasných otvorených údajov

V súčasnosti medzi najvýznamnejšie obmedzenia v kvalite publikovaných otvorených údajov patrí ich vzájomná nekompaktibilita, resp. absentujúca interoperabilita. Historický Katalóg dátových prvkov používaný na popis údajov je jednak reprezentovaný ako slabá ontológia, to je v nízko strojovo spracovateľnej podobe, a súčasne popisuje len malý zlomok údajov verejnej správy. Ďalším problémom obmedzujúci interoperabilitu otvorených údajoch je používanie rôznych identifikátoroch pre rovnaké entity, pričom mnohé údaje sú súčasne spravované rôznymi gestormi samostatne, a tie nezdieľajú rovnaký model. Pracovať s údajmi ako s jedným konzistentným celkom je nemožné a prakticky pre každé ISVS sa musia integrovať nanovo.

Sémantické dátové štandardy

Riešenie zabezpečenia požadovanej interoperability údajov verejnej správy je riešené postupne od roku 2013 prostredníctvom zavádzania tzv. Sémantických dátových štandardov pre údaje verejnej správy SR. Kľúčovými časťami uvedených štandardov je Metóda tvorby Jednotného referencovateľného identifikátora, Centrálny model údajov verejnej správy založený na ontológiách a množina registrovaných URI identifikátorov v Centrálnom metainformačnom systéme. Agenda štandardizácie spadá pod PS1. Tieto štandardy popisujú nosné prvky a metodiky potrebné na systematické použitie prepojených údajov v priestore SR. Zároveň je rozpracovaných značné množstvo ontológií popisujúcich rôzne oblasti štátnych dát (RFO, RPO, RA, ...). Tieto svojím rozsahom ako aj významom už v súčasnosti plnohodnotne nahrádzajú a rozširujú pôvodnú ideu KDP (Katalóg dátových prvkov).

Nakoľko je problematika komplexná, predpokladá sa neustále rozširovanie a spresňovanie štandardov.

### LOD Slovakia (Linked Open Data)

LOD Slovakia je množina otvorených prepojených údajov verejnej správy SR, ktorý predstavuje „cieľový“ stav kvality definujúci požadovanú interoperabilitu . Databáza LODu obsahuje rôzne údaje verejnej správy so zameraním na centrálne údaje (referenčné registre, základné číselníky, centrálny model údajov, register právnických osôb, register adries a mnohé iné), ktoré vychádzajú so schválených sémantických dátových štandardov a slúžia aj ako pomoc pri tvorbe reálnych projektov informatizácie štátnej správy. Predpokladá sa, že LOD Slovakia bude obsahovať aj otvorené datasety publikované inými producentmi údajov ako len verejný sektor, pričom LOD Slovakia bude prepojený i s globálnym projektom LOD, ktorý predstavuje implementáciu takzvaného Sémantického webu.

# Ciele realizácie

## Architektonické ciele

Táto kapitola obsahuje špecifikáciu architektonických cieľov , ktoré by mali byť dosiahnuté v oblasti manažmentu údajov verejnej správy do konca roku 2020. Ciele určujú konkrétnu pozitívnu zmenu, ktorá sa vo verejnej správe očakáva, ktorá bude podporovaná. Pre každý cieľ sú jasne stanovené merateľné ukazovatele, ktoré budú pravidelne vyhodnocované v rámci systematického monitoringu e-Governmentu.

Ciele sú záväzné, stav ich plnenia bude centrálne monitorovaný, posudzované projekty budú hodnotené aj z pohľadu príspevku k plneniu týchto cieľov.

### Komplexný pokrok v téme otvorených údajov

Tabuľka : Postavenie krajiny v téme otvorených údajov

|  |  |  |
| --- | --- | --- |
| Ciel | Ukazovateľ cieľa | Výsledok |
| Zlepšiť postavenie SR v medzinárodnom porovnávaní a stať sa lídrom v téme otvorených údajov v EU | Open Data Maturity index | Umiestnenie SR v prvej päťke |

**Open Data Maturity index** meria pripravenosť na realizáciu otvorených údajov, dopad využívania otvorených údajov a vyspelosť národného portálu pre otvorené údaje. SR dosahuje v súčasnosti v rebríčku Open data postavenie na 9. mieste a patrí do skupiny „fast-trackers“.

### Zlepšiť dostupnosť údajov verejnej správy vo forme otvorených údajov

Tabuľka 4: Dostupnosť údajov v štátnej správe

| Ciel | Ukazovateľ cieľa | Výsledok |
| --- | --- | --- |
| Zvýšiť počet inštitúcií verejnej správy, ktoré publikujú otvorené údaje | Podiel inštitúcií štátnej správy, ktoré publikujú otvorené údaje | 100 % |
| Podiel inštitúcií štátnej správe, ktoré publikujú aktuálne otvorené údaje | 75 % |
| Zvýšiť rozsah publikovaných údajov štátnej správy | Podiel **povinne zverejňovaných údajov** štátnej správy, dostupných vo forme otvorených údajov | 100 % |
| Podiel verejne prístupných údajov publikovaných ako otvorené údaje | 98 % z vhodných |
| Umožniť používanie otvorených údajov na právne účely | Podiel **prioritných datasetov**, ktoré sú použiteľné na právne účely | 100 % z vhodných |
| Podiel údajov publikovaných ako otvorené údaje, ktoré sú použiteľné na právne účely | 90 % z vhodných |

Tabuľka : Dostupnosť údajov v samospráve

|  |  |  |
| --- | --- | --- |
| Ciel | Ukazovateľ cieľa | Výsledok |
| Zvýšiť počet inštitúcií samosprávy, ktoré publikujú otvorené údaje | Podiel inštitúcií samosprávy, ktoré publikujú otvorené údaje | VÚC 100 %  Mestá 100 %  Obce 70 % |
| Podiel inštitúcií samosprávy, ktorých dáta sú publikované ako **publikačné minimum pre samosprávu** | VÚC 100 %  Mestá 100 %  Obce 90 % |

Prioritné datasety

Prioritné datasety sú najžiadanejšie datasety od verejnosti, ktoré sú určené na základe pravidelného prieskumu medzi verejnosťou a zainteresovanými subjektami. Prioritné datasety budú schválené v procese manažmentu údajov. Zoznam je dynamicky aktualizovaný každoročne, pričom cieľom je jeho redukcia. Zoznam pripravuje Úrad splnomocnenca pre rozvoj občianskej spoločnosti a je schvaľovaný Dátovou kanceláriou verejnej správy.

Publikačné minimum pre samosprávu

Publikačné minimum pre samosprávu predstavuje zoznam datasetov podľa jednotlivých kategórií miest a obcí, ktoré by mala prioritne publikovať inštitúcia samosprávy. Tento zoznam datasetov definuje aj požiadavky na kvalitu, aktuálnosť a formát údajov (pre jednotlivé typy datasetov). Ako také bude toto publikačné minimum definované Úradom podpredsedu vládu SR pre informatizáciu a investície v spolupráci so splnomocnencom vlády SR pre rozvoj občianskej spoločnosti a predsedom Združenia miest a obcí Slovenska, pričom publikačné minimum bude aktualizované.

Podiel inštitúcií verejnej správy, ktoré publikujú otvorené údaje

Ukazovateľ vyjadruje podiel inštitúcii verejnej správy na celkovom počte inštitúcií, ktoré publikujú otvorené dáta, ktoré sú zaevidované v centrálnom katalógu otvorených údajov. Dôležité je, aby inštitúcie dodržiavali procesy publikovania otvorených údajov.

Podiel inštitúcií verejnej správy, ktoré publikujú aktuálne otvorené údaje:

Aktuálne údaje sú údaje, ak respektíve ak sa údaje zhodujú s platnými údajmi v informačnom systéme verejnej správy a informácia o tejto zhode je dostupná respektíve ak požadovaná frekvencia zverejňovania datasetu je dodržiavaná.

Podiel údajov publikovaných ako otvorené údaje, ktoré sú použiteľné na právne účely

Dataset je použiteľný na právne účely v prípade, ak má zmysel na základe neho tvoriť rozhodnutia, ktoré ovplyvňujú vzťahy subjektov v občianskom alebo obchodnom práve, alebo sú informácie z datasetu relevantné v konaniach orgánov verejnej moci, respektíve ide o referenčné údaje.

Počet datasetov publikovaných vo formáte s vysokým potenciálom na znovupoužitie

Vysokým formátom na znovupoužitie sú myslené prepojené údaje: na identifikáciu entít sa používajú Jednotné referencovateľné identifikátory v podobe URI, jednotlivé údaje sú popísané prostredníctvom odporučených ontológií EÚ na popis údajov respektíve prostredníctvom národných ontológií, ak také nie sú k dispozícií. Údaje tak vedia byť strojovo spracované s ohľadom na význam.

### Kvalita publikovaných údajov

| Ciel | Ukazovateľ cieľa | Výsledok |
| --- | --- | --- |
| Zvýšiť kvalitu publikovaných údajov štátnej správy | Podiel datasetov publikovaných minimálne v úrovni kvality 3★ (<http://5stardata.info/> ) | 100 % |
| Podiel datasetov s vysokým potenciálom na znovupoužitie publikovaných minimálne v úrovni kvality 4★ ([http://5stardata.info/](http://4stardata.info/)) | 90 % |
| Podiel datasetov s vysokým potenciálom na znovupoužitie publikovaných minimálne v úrovni kvality 5★ (<http://5stardata.info/> ) pre novovznikajúce a inovované systémy | 100 % |
| Zvýšiť kvalitu publikovaných údajov samosprávy | Podiel datasetov publikovaných v úrovni kvality 3★ (<http://5stardata.info/> ) | 100 % |
| Podiel datasetov s vysokým potenciálom na znovupoužitie publikovaných vo formáte 4★ (<http://5stardata.info/>) | 80 % |
| Podiel datasetov s vysokým potenciálom na znovupoužitie publikovaných vo formáte 5★ (<http://5stardata.info/>) pre datasety pre novovznikajúce a inovované systémy | 100 % |

Tabuľka 6: Definovanie úrovní kvality otvorených verejných údajov SR z pohľadu ich interoperability[[1]](#footnote-2)

| Úroveň | Popis |
| --- | --- |
| 1★ | Sprístupnenie údajov (v akomkoľvek formáte) pod otvorenou licenciou |
| 2★ | Údaje sú sprístupnené v štruktúrovanej forme (napríklad Excel namiesto obrázka s tabuľkou) |
| 3★ | Sprístupnenie údajov v otvorenom formáte (napríklad CSV alebo ODS namiesto Excelu) |
| 4★ | Použitie URI na identifikáciu ľubovoľných entít (dátové prvky, údaje) tak, aby ich bolo možné referencovať. Údaje sú popísané ontológiami, avšak tie nepatria do Centrálneho modelu údajov verejnej správy. |
| 5★ | Zladenie/prepojenie údajov s ostatnými, aby bol jasný ich význam. V prípade otvorených verejných údajov SR ide o ich popísanie Centrálnym modelom údajov verejnej správy jednotne mapovaným na medzinárodné ontologické štandardy. |

Rozhodnutie o tom, v akej kvalite sú, alebo majú byť vypublikované otvorené údaje je v kompetencii Centrálnej dátovej kancelárie (pridelenie úrovne kvality v stupňoch interoperability 1★-5★). Pre existujúce otvorené údaje, ktoré sú reprezentované vo forme s nízkym stupňom interoperability (max 3★), avšak majú vysoký potenciál na znovapoužitie, Centrálna dátová kancelária odporučí ich postupné skvalitnenie na dosiahnutie požadovaného stupňa interoperability. V prípade nových, alebo inovovaných ISVS publikujúcich otvorené údaje Centrálna dátová kancelária určí, či sú dáta povahy z vysokým stupňom znovapoužitia a je nutné zabezpečiť ich súlad z Centrálnym modelom údajov založených na ontológiách (5★), alebo údaje tento charakter nemajú (4★).

### Prístupnosť publikovaných údajov

| Ciel | Ukazovateľ cieľa | Výsledok |
| --- | --- | --- |
| Zlepšiť prístupnosť publikovaných údajov verejnej správy | Podiel datasetov publikovaných prostredníctvom aplikačného rozhrania (API) | 70 % |
| Podiel datasetov v súborovom formáte | 90 % |

### Zvýšiť dynamiku a prínos dátového hospodárstva

Tabuľka 6: Dátové hospodárstvo

|  |  |  |
| --- | --- | --- |
| Ciel | Ukazovateľ cieľa | Výsledok |
| Zvýšiť dynamiku a prínosnosť dátového hospodárstva | Počet stiahnutí otvorených údajov ročne | 950 000 |
| Počet aplikácií, ktoré kreatívne využívajú otvorené údaje | 200 |
| Zlepšiť zapojenie Slovenska do jednotného digitálneho trhu (DSM) | Podiel datasetov dostupných v anglickom jazyku | 30 % |
| Katalóg datasetov predložený v anglickom jazyku | 100 % |

Počet stiahnutí otvorených údajov

Ide o stiahnutie datasetov cez katalóg otvorených údajov aplikáciami alebo používateľmi, evidované na platforme pre otvorené údaje.

Počet aplikácií, ktoré kreatívne využívajú otvorené údaje

Aplikácia počítaná v prípade, ak je registrovaná na platforme data.gov.sk a pre svoju funkčnosť využíva dostupné údaje verejnej správy cez centrálny katalóg.

## Dotknuté subjekty a ich záujmy

V tejto časti sú identifikované dotknuté subjekty, ktorých sa daná strategická priorita (alebo jej časť) týka a ich relevantné záujmy.

Štúdia otvorených dát (The Open Data Study (2010)) zistila, že na úspešnú implementáciu programu otvorených dát verejnej správy (OGD) sú nevyhnutné tri kľúčové skupiny stakeholderov, alebo inak povedané tri vrstvy spoločnosti:

* Vplyvná a aktívna občianska spoločnosť, ktorá vytvára tlak zdola na zmenu tradičných postupov informovania verejnosti a vytvára inovatívne služby, produkty a webové stránky, ktoré ukazujú spôsob využitia OGD.
* Zamestnanci verejnej správy, ktorí vnímajú OGD ako spôsob zvýšenia efektivity a transparentnosti, predstavujú strednú vrstvu.
* Vysoko postavení politickí lídri ako ministri tvoria tretiu vrstvu a tvoria víziu otvoreného vládnutia.

Ako v rámci spomenutej štúdie odpozoroval Sir Tim Berners-Lee: *„Úspešná realizácia OGD musí začať zhora, v strede aj zospodu!”* Tento prístup je všeobecne označovaný za trojstupňový prístup implementácie OGD a zahŕňa angažovanie sa všetkých vrstiev dotknutých subjektov.

Úrad podpredsedu vlády SR pre informatizáciu a investície

* Mať politický mandát, právnu a odbornú kompetenciu (v oblasti licencií, legislatívy, technológií a dátových analýz) riešiť problémy a spory pri publikovaní otvorených dát, napríklad v oblasti licencií alebo platnosti dát.
* Zabezpečiť vysoký počet užitočných datasetov otvorených dát vo vysokej kvalite, ktorú budú intenzívne využívané inovatívnymi podnikmi a komunitami (rast dátového hospodárstva).
* Efektívne vykonávať dohľad nad zverejňovaním otvorených dát a dodržiavaním štandardov a najlepších praktík s podporou Centrálneho modelu údajov VS.
* Zodpovedať za plnenie stanovených cieľov v tomto dokumente

Úrad splnomocnenca vlády pre rozvoj otvorenej spoločnosti

* Stimulovať rozvoj inovatívnej, aktívnej a participatívnej občianskej spoločnosti, ktorá používa otvorené údaje.
* Aby dochádzalo k plneniu úloh plánu iniciatívy pre otvorené vládnutie (OGP) a ich koordinácii s aktivitami e-Govermentu.

Správca informačného programu verejnej správy

* Mať možnosti pre plnenie povinností súvisiacich s rolou správcu informačného systému verejnej správy a poskytovateľa údajov (vrátane jasného návodu pre postup, ako zverejňovať otvorené údaje),
* Mať k dispozícii dostatok prostriedkov a kapacít na tvorbu a publikovanie kvalitných údajov.

Ministerstvá, samospráva a ďalšie štátne organizácie

* Získavať konštruktívnu spätnú väzbu od verejnosti a neziskového sektora.
* Legitimizovať pred verejnosťou svoje rozhodnutia a schopnosť zavádzať pozitívne opatrenia do praxe.
* Zaviesť princípy štátu, fungujúceho na základe využívania dát v praxi.
* Vykonávať dohľad nad efektívnosťou opatrení.
* Mať kontrolu nad efektívnosťou vynakladaných zdrojov a investícií.
* Efektívne plniť povinnosti vyplývajúce z všeobecne záväzných právnych predpisov (napríklad zákon o slobodnom prístupe k informáciám či príslušných európskych nariadení).
* Mať k dispozícii dostatok prostriedkov a kapacít na tvorbu a publikovanie kvalitných údajov.

Úrad vlády SR a NASES

* Viesť centrálny katalóg otvorených dát prostredníctvom agentúry NASES a prevádzkovať centrálne komponenty platformy otvorených dát.
* Vytvoriť inovatívnu komunitu pracujúcu s otvorenými údajmi.

Verejnosť

* Mať zabezpečenú ochranu súkromia,
* Mať jednoduchý, zrozumiteľný a centrálny prístup ku všetkým otvoreným dátam verejnej správy a samosprávy.
* Efektívne aplikovať práva vyplývajúce zo Zákona o slobodnom prístupe k informáciám.
* Zvyšovať transparentnosť vo verejnej správe a v procesoch verejného obstarávania.
* Mať prehľad o predpokladaných dopadoch politík a regulácií.
* Mať prehľad o zdôvodnení investícii vo verejnom záujme.
* Mať prehľad o výkone verejnej správy v jednotlivých sektoroch (až na úroveň jednotlivých inštitúcií) – vrátane ratingov.
* Možnosť participatívnej účasti.
* Mať prístup k navrhovaným štandardom a definíciám schválených štandardov prostredníctvom zverejňovania v systéme MetaIS.

Občianske komunity a inovatívne komunity

* Okrem vyššie uvedeného, chcú mať k dispozícií kvalitné údaje v strojov čitateľnom formáte a nástroje (platformy a aplikačné rozhrania), aby mohli vytvárať aplikácie klasifikované do troch základných kategórií:
* Transparentnosť a zodpovednosť, ktoré nútia verejnú správu deklarovať svoju zodpovednosť tým, že informácie a procesy budú transparentné. Projekty v tejto oblasti často smerujú k poradnej činnosti komunít a k ovplyvňovaniu politík a rozhodovania.
* Interakcia občanov a verejnej správy, ktorá uľahčuje komunikáciu a dáva jej nový rozmer, tieto projekty majú za cieľ nahradiť tradičné kanály a spôsoby komunikácie elektronickými verejnými službami, ktoré sú orientované na používateľa a poskytujú vyšší používateľský komfort.
* Digitálne nástroje pre každodenný život občanov, ktoré ich život uľahčujú, tieto projekty sa často orientujú na malé, ale dôležité aspekty nášho každodenného života, s cieľom poskytnúť riešenia jednoduché na použitie.

Táto skupina subjektov sa považuje vždy za pionierov v oblasti rozvoja programov OGD, hlavne vďaka implementácií pilotných projektov preukazujúcich úžitok zo zverejňovania otvorených dát. Odporúča sa tiež orientovať na vytváranie takzvaných verejno-komunitných partnerstiev, ktoré budú experimentovať s novými typmi spolupráce medzi týmito veľmi rozdielne fungujúcimi subjektami[[2]](#footnote-3).

Podnikateľské prostredie

* Mať dostupné nástroje, API a otvorené dáta a pre poskytovanie inovatívnych komerčných služieb,
* Regulácie, ktoré podporujú dátové hospodárstvo a využitie údajov (vrátane jasného modelu licencií),
* Mať prístup ku kvalitným analýzam a prognózam o vývoji ekonomiky a o plánovaných opatreniach a obstarávaniach verejnej správy,
* Zvyšovať transparentnosť verejnej správy pre zlepšovanie podnikateľských príležitostí.

# Organizácia

Návrh modelu zodpovedností a potrebné organizačné zabezpečenie vrátane organizačných zmien.

## Zodpovednosť

Zodpovedný gestor

Gestorom strategickej priority Otvorené údaje je **Úrad podpredsedu vlády SR pre informatizáciu a investície** (ÚPPVIaI).

ÚPPVIaI je ústredným orgánom štátnej správy zodpovedným za centrálne riadenie informatizácie spoločnosti a tvorbu politiky jednotného digitálneho trhu, rozhodovanie o využívaní finančných zdrojov vo verejnej správe pre informačné technológie, centrálnu architektúru integrovaného informačného systému verejnej správy a koordináciu plnenia úloh v oblasti informatizácie spoločnosti. Úrad tiež sleduje stav a hodnotenia rozvoja informatizácie spoločnosti a usmerňuje tvorbu koncepcií rozvoja informačných systémov verejnej správy. Riadi implementáciu prioritnej osi 7 Operačného programu Integrovaná infraštruktúra ako sprostredkovateľský orgán pod riadiacim orgánom. Z pohľadu organizačného zabezpečenie je nevyhnutné vytvoriť Centrálnu kapacitu pre koordináciu publikovania otvorených dát vo verejnej správe a v samospráve a dopracovanie Metainformačného systému do podoby, ktorá zabezpečí systematický, udržateľný a efektívny rozvoj otvorených prepojených údajov.

ÚPPVIaI tiež zodpovedá za sémantickú interoperabilitu údajovej základne verejnej správy:

* Tvorba a štandardizácia Centrálneho modelu údajov verejnej správy založeného na ontológiách v Centrálnom metainformačnom systéme.
* Schvaľovanie a vydávanie URI identifikátorov pre referencovateľnosť údajov verejnej správy v Centrálnom metainformačnom systéme.
* Prideľovanie existujúceho a požadovaného stupňa interoperability pre otvorené údaje.

Zapojené organizácie

Zoznam organizácií, ktoré budú dotknuté realizáciou priority:

**Úrad vlády SR:**

* je správcom centrálneho katalógu otvorených údajov.

**NASES**:

* prevádzkuje Centrálny katalóg otvorených údajov, ďalej „centrálny katalóg“, na Ústrednom portáli verejnej správy. Centrálny katalóg je informačný systém verejnej správy, ktorý obsahuje údaje o informáciách sprístupňovaných na opakované použitie a otvorených údajoch, najmä ich obsah, meta-údaje, podmienky použitia a obmedzenia použitia.
* Prevádzkuje Centrálne služby pre otvorené údaje.
* zodpovedá za dereferenciáciu URI.

**Úrad splnomocnenca** **Úrad vlády SR**

* koordinuje plnenie Iniciatívy pre otvorené vládnutie (OGP). Zámerom tejto iniciatívy je zlepšovanie vládnutia a spravovania vecí verejných cez zvyšovanie transparentnosti, efektivity a zodpovednosti, ako aj vytváraním inštitúcií, ktoré občanom umožnia participovať na vládnutí.
* Zabezpečuje spoluprácu s inými štátnymi inštitúciami v téme otvorených údajov.

**Všetky organizácie verejnej správy a samosprávy**, pretože na základe pripravovaného právneho stavu majú byť sprístupnené povinnými osobami všetky údaje, okrem tých, pre ktoré to špecificky určitý zákon vylučuje. Inštitúcie zverejňujú informácie v zmysle zákona č. 211/2000 Z. z. o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov v znení neskorších predpisov. Sprístupňujú verejnosti údaje z informačných systémov verejnej správy, ak osobitný predpis neustanovuje inak v zmysle zákona o informačných systémoch verejnej správy.

## Posúdenie kapacitnej pripravenosti

### Personálna kapacita

Jednotlivé zložky štátnej správy ako aj samosprávy nie sú z hľadiska personálnej kapacity na zmeny pripravené. Je potrebné zabezpečiť dodatočné zdroje na realizáciu úloh, ktoré vyplývajú z tohto dokumentu. Príslušné zdroje je však možné získať (ušetriť) aplikáciou racionalizačných a optimalizačných opatrení.

### Kapacita informačných systémov

Informačné systémy verejnej správy sú vo všeobecnosti schopné publikovania otvorených údajov. Je však potrebná ich modifikácia, najmä v súvislosti:

* s implementáciou princípov pre prepojené údaje (Linked Data),
* efektívnym zabezpečením využívania mechanizmov správy a zdieľania údajov, najmä riadenie kvality údajov a využitie platformy integrácie údajov,
* zabezpečenie automatizovaného generovania datasetov otvorených údajov,
* využitím nástrojov a služieb pre publikovanie otvorených údajov.

## Organizačné zmeny

Aby bolo možné realizovať strategickú prioritu Open Data, sú navrhované nasledujúce organizačné zmeny:

* Doplnenie Dátovej kancelárie verejnej správy o nové kompetencie, ktorá bude mať na starosti 1) prideľovanie Jednotných referencovateľných identifikátorov (URI) pre údaje verejnej správy, 2) správu Centrálneho modelu údajov verejnej správy založeného na ontológiách, 3) Centrálnu koordináciu zverejňovania otvorených údajov v správnom formáte**,** 4) prideľovanie a definovanie požadovaného stupňa interoperability pre všetky zverejňované datasety. Vhodné by bolo tiež disponovať právnymi odborníkmi v oblasti dát a licencovania.
* Zriadenie Centrálnej kapacity pre koordináciu publikovania otvorených údajov v štátnej správe.
* Zriadenie Centrálnej kapacity pre koordináciu publikovania otvorených údajov v samospráve v spolupráci so Zdužením miest a obcí Slovenska.

Nové role:

Na centrálnej úrovni a lokálnej úrovni budú menované nasledujúce role.

Tabuľka 7: Prehľad organizačného zabezpečenia

| Úroveň | Zodpovednosť | Rola | Počet |
| --- | --- | --- | --- |
| Centrálna  (ÚPPVII) | * Celková zodpovednosť za otvorených údajov vo verejnej správe | Vedúci Dátovej kancelárie | 1 |
| * Celková zodpovednosť za kvalitu otvorených údajov | Hlavný dátový kurátor | 1 |
| * Celková zodpovednosť za sémantickú interoperabilitu | Správca URI identifikátorov  Správca Centrálneho modelu | 1 |
| Centrálna  (NASES) | * Centrálne služby pre otvorené údaje | Architekt otvorených údajov | 1 |
| * Kvalita údajov | Dátový kurátor pre data.gov.sk | 1 |
| Centrálna  (DEUS) | * Kvalita údajov | Hlavný dátový kurátor pre samosprávu | 1 |
| Väčšia inštitúcia[[3]](#footnote-4) | * Životný cyklus údajov * Kvalita údajov | Vlastník údajov | >1 (pre každý segment údajov) |
| Dátový kurátor | >1 (pre každý segment údajov) |
| * Dôveryhodnosť údajov | Manažér dôveryhodnosti otvorených údajov | 1 |
| Menšia inštitúcia | * Životný cyklus údajov * Kvalita údajov | Vlastník údajov | 1 |
| Dátový kurátor | 1 |

# Stratégia

Účelom tejto kapitoly je výber strategického prístupu, ktorý nás privedie k dosiahnutiu cieľov a želaného zlepšenia. Jednotlivé témy, ako referenčné údaje, kvalita údajov sú podrobnejšie vysvetlené v kontexte slovenského prostredia. Identifikované sú základné varianty, ako k problematike pristúpiť, vrátane výberu vhodného riešenia.

## SWOT analýza

Tabuľka 8: SWOT analýza pre manažment údajov

|  |  |
| --- | --- |
| Silné stránky | Slabé stránky |
| * Spustenie portálu data.gov.sk * Implementácia projektu Elektronické služby Úradu vlády SR - eDemokracia a otvorená vláda * Rastúci počet datasetov verejnej správy * Zabezpečené financovanie ďalších aktivít v oblasti otvorených údajov v rámci OP II a OP EVS * Podpora vlády SR a vytvorenie samostatného ÚOŠS, ktorý komplexne riadi informatizáciu * Zapojenie odbornej verejnosti a mimovládnych organizácií do riešenia problematiky otvorených údajov * Postavenie SR v rámci krajín EÚ | * Chýbajúca komplexná legislatíva a metodická podpora pre oblasť otvorených údajov * Chýbajúce jednotné pravidlá a princípy pre poskytovanie otvorených údajov v rámci verejnej správy * Nízky počet zverejnených datasetov v pomere k informáciám, ktorými verejná správa disponuje * Chýbajúci komplexný prehľad o existujúcich údajoch vo verejnej správe * Chýbajúca kontrola poskytovania otvorených údajov a nedostatočná vymáhateľnosť * Nedostatočné kapacity (finančné a personálne) na strane verejnej správy * Nedostatočná komunikácia medzi verejnou správou a socioekonomickými partnermi * Nekvalitné alebo chýbajúce metadáta * Nejednotné používanie slovníkov * Nízka kvalita zverejnených datasetov * Chýbajúca osveta voči občanom a podnikateľom * Nedostatky týkajúce sa používateľskej prívetivosti a odozvy |
| Príležitosti | Hrozby |
| * Prínos otvorených údajov do rozvoja ekonomiky štátu * Zvýšenie konkurencieschopnosti SR * Zvýšenie transparentnosti a dôveryhodnosti verejnej správy * Zvýšenie spokojnosti občanov s verejnou správou * Väčšie zapojenie verejnosti do správy veci verejných * Nové služby pre občanov zo strany verejnej správy * Nové služby a produkty zo strany súkromného sektoru * Podpora inovatívnych služieb * Podpora začínajúcich podnikateľov * Zlepšenie procesov vo verejnej správe na základe nových údajov * Zlepšenie politík a legislatívy na základe nových údajov * Možnosť znovupoužitia údajov, resp. využitia existujúcich údajov bez potreby vytvárať duplicitné údaje * Nové údaje a podnikateľské možnosti získané kombináciou a spojením verejných a súkromných údajov * Vytvorenie centrálneho portálu, kde by mohli podnikatelia zverejňovať aplikácie založené na otvorených údajoch | * Odmietavý postoj úradníkov verejnej správy * Zmeny politiky v oblasti otvorených údajov v rámci politického cyklu * Nejednotná politika v oblasti otvorených údajov * Nejednotná licenčná politika * Chýbajúce štandardy * Nedostatok finančných zdrojov na realizáciu zverejňovania otvorených údajov obzvlášť u menších úradov * Zverejňovanie len otvorených údajov, ktoré nemajú potenciál na znovupoužitie * Ekonomický model založený na spoplatnení otvorených údajov (priamy príjem z otvorených údajov vs. nepriamy príjem v podobe vyšších daňových príjmov) * Nedostupnosť otvorených údajov, resp. ich obmedzená prístupnosť * Nepripravenosť podnikateľov na využívanie otvorených údajov * Slabá odozva verejnej správy na podnety verejnosti * Nedostatočná ochrana osobných a citlivých údajov * Nedostatočná aktualizácia datasetov * Neschopnosť verejnej správy poskytovať nadstavbové služby * Potencionálny výpadok príjmov štátneho rozpočtu |

## Vízia

Slovensko bude transformované na „Data-driven state“ – štáte, fungujúcom na základe využívania dát a presnom riadení celého životného cyklu dát Štát bude prijímať rozhodnutia na základe najlepších znalostí, ktoré sú k dispozícii. Aby bolo možné dosiahnuť takýto stav:

* Údaje budú spravované ako vzácny zdroj.
* Vytvoria sa podmienky pre maximálne využitie potenciálu, ktorý je možné vyťažiť z údajov.
* Verejná správa bude schopná komplexne pokryť životný cyklus údajov.

Znamená to tiež, že pre jednotlivé oblasti rozhodovania budú k dispozícií modely, dáta a nástroje, ktoré umožnia vytvoriť analýzy, na základe ktorých sa podporí rozhodovanie. Znamená to tiež transformáciu fungovania organizácií a procesov vo verejnej správe tak, aby boli tieto analýzy efektívne používané a zároveň zverejňované vo vhodnej vizuálnej podobne pre verejnosť.

Treba zaviesť jednotné riešenie, ktoré však nebude narúšať autonómiu jednotlivých inštitúcii verejnej správy v oblasti tvorby údajov a riadenia prístup k nim.

V prípade otvorených údajov táto vízia znamená, že:

* Vznikne silná dátová ekonomika, ktorá bude otvorené údaje verejnej správy využívať na poskytovanie služieb s pridanou hodnotou,
* Každá inštitúcia verejnej správy bude schopná a ochotná publikovať kvalitné údaje v otvorenom formáte,
* Otvorenosť zmení kultúru vo verejnej správe smerom k participatívnosti a spolupráci s angažovanou komunitou.

Téma otvorených údajov je široká a zahŕňa viacero iniciatív. Pri obmedzenom čase a zdrojoch je potrebné prioritizovať a chápať relatívnu dôležitosť jednotlivých tém. Nasledujúca tabuľka udáva poradie týchto priorít.

Tabuľka : Prioritizácia tém

| Priorita | Téma |
| --- | --- |
| 1 | Dostupnosť a rozsah údajov |
| 2 | Podpora komunity pracujúcich s otvorenými údajmi |
| 3 | Kvalita údajov |
| 4 | Nové centrálne služby a platforma |
| 5 | Metodiky a štandardy |

## Výber prístupu

Prístup k systematickému riešeniu Open Data je možné zhrnúť ako:

* Implementovať „lessons learned“ z projektu eDemokracia:
* najťažšie je sprostredkovanie komunikácie medzi poskytovateľmi údajov a používateľmi údajov – musí vzniknúť funkčná platforma pre spoluprácu,
* minimalizovať formálne náležitosti,
* zvládnuť sofistikovanosť implementácie (testovaním prototypov, zapojením čo najširšej používateľskej základne, otvoreným zverejňovaním informácií o stave projektu a všetkých výstupov projektu).
* Ako súčasť reformy verejnej správy, vytvorí sa centrálna kapacita schopná centralizovane a systematicky riadiť a koordinovať zverejňovanie a používanie otvorených dát vo verejnej správe a v samospráve:
* Manažment dátového modelu verejnej správy a samosprávy(vrátane ontológie a definície prepojení),
* Spolupráca pri návrhu analytických produktov, definícia požiadaviek na dátové zdroje,
* Riadenie zberu dát z informačného prostredia verejnej správy a samosprávy,
* Licencovanie dátových zdrojov pre otvorené dáta.
* Na jednotlivých ministerstvách budú zriadené analytické jednotky zodpovedné za konkrétne oblasti verejnej správy (politiky, regulácie, výkonnosť, investície) a so schopnosťami zverejňovať otvorené dáta (obdobné kapacity budú vybudované aj pre samosprávu).
* Podporované budú segmentové iniciatívy s cieľom komplexného zabezpečenia publikovania v segmente a v samospráve:
* Identifikácia datasetov, ktoré sa budú zverejňovať v rámci segmentu,
* Nastavenie procesov zverejňovania dát,
* Zabezpečenie, že datasety sú pravidelne publikovanie cez portál „data.gov.sk“.
* Jednotlivé informačné systému verejnej správy budú pripájané k Centrálnemu dátovému skladu cez dátovú integráciu a ETL procesy.

### Otvorené údaje a prístup k dátovému hospodárstvu

Ak chceme, aby naše dátové hospodárstvo vytváralo rast a pracovné miesta, musia sa využívať dáta. Ale na to, aby sa využívali, je tiež potrebné, aby boli dostupné a aby sa analyzovali.

Dátové hospodárstvo predstavuje súbor ekonomických aktivít, ktoré súvisia so zberom, spracovaním, analyzovaním údajov a tvorbou vyplývajúcej pridanej hodnoty. Hlavným motorom dátovej ekonomiky v súčasnosti je spracovanie veľkých údajov a aplikácia umelej inteligencie, ako i fenomén digitálnych platforiem, ktoré vytvárajú nové a disruptujú tradičné biznis modely.

Tabuľka

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Údaje sa zverejnenia ako otvorené v maximálne možnom rozsahu - štát "nezaujíma" použitie údajov, to je úlohou dátového hospodárstva. Autorské práva sú riešené otvorenou licenciou. | X | * Schválené už v Stratégii a akčnom pláne sprístupnenia a používania otvorených údajov verejnej správy Akčnom pláne OGP a NKIVS a v smernici PSI |
| Údaje by sa nemali používať na komerčné účely. |  | * Variant ide proti princípom otvorených údajov (o použití by sa nemalo rozhodovať) |
| Dátové hospodárstvo nie je podporované otvorenými údajmi. |  | * Variant ide proti princípom Stratégie 2020, NKIVS a politickým prioritám ÚPPVIaI |

### Ako vytvoriť funkčnú komunitu

Funkčná komunita, ktorá pracuje s otvorenými údajmi je nevyhnutnou podmienkou na získanie prínosov z dát. Komunita (resp. komunity) by mala vznikať samoorganizačne. S otvorenými údajmi by mal pracovať neziskový sektor, kreatívny priemysel, inovatívne MSP, či akademický sektor. Kľúčovými faktormi pri podpore komunity sú:

* aby zahŕňala dostatočný počet (kritický) členov,
* aby vytvárala inovatívne aplikácie pracujúce s otvorenými údajmi,
* aby bola v čo najvyššej miere prepojená so zahraničnými a globálnymi komunitami.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Dopytové výzvy na tému otvorené údaje, kde žiadateľom respektíve partnerom budú neziskové organizácie a MSP - stimulácia vytvárania aplikácií verejnými zdrojmi | X | * Je potrebné správne nastaviť kritéria výberu * Zamerať sa aj na hodnotné aplikácie pre spoločnosť (nie len komerčné) |
| Urýchlené zverejnenie zaujímavých a kvalitných otvorených údajov (prioritných datasetov) | **X** | * Prioritné datasety pre zverejnenie sú pravidelne skúmané ÚSpROS na základe reálneho dopytu odbornej verejnosti a komunity používateľov * Dá sa predpokladať, že sprístupnenie práve týchto datasetov výrazne zvýši hodnotu zverejnených otvorených údajov a motiváciu a priestor tvoriť aplikácie |
| Medzinárodná spolupráca a integrácia s nadnárodnými bázami údajov | X | * Pozitívny príklad sú aktivity v oblasti priestorových údajov * Zámerom je, aby slovenské otvorené údaje boli interoperabilné a aby aplikácie vyvíjané v jednotnom digitálnom trhu fungovali so Slovenskými údajmi |
| Organizácia hackathonov a komunitných podujatí | X | * Ideálny formát pre podporu „spolupráce“ používateľov údajov a poskytovateľov údajov * Kreatívna atmosféra môže zmeniť **kultúru verejnej správy**, podľa princípov tohto dokumentu |
| Fórum, komunikácia a právna istota (aj zber spätnej väzby) | X | * Efektívny spôsob, ako zabezpečiť prístup a používanie otvorených údajov |
| Podpora spracovania otvorených údajov univerzitami. Výskum, s ktorými výskumníci pracujú by mali byť otvorené |  | * Je nevyhnutné zapojiť univerzity do témy otvorených údajov **a začať pravidelnú komunikácií** |
| Podpora komunity zo strany štátu nie je potrebná |  | * Prílišná administratívna zaťaž vo vzťahu k efektivite takejto pomoci |
| Grantová agentúra a pôžičkový systém |  | * Prílišná administratívna zaťaž vo vzťahu k efektivite takejto pomoci * Potrebné finančné nástroje nie sú pripravené |

### Model pre „governance“ otvorených údajov

Riadenie publikovania otvorených údajov má viacero úrovní, pričom každá je dôležitá a musí byť systematický riešená. Je potrebné určiť:

* Spôsob tvorby politiky v oblasti údajov a údajov verejného sektora (legislatíva, pravidlá a štandardy, koncepcie),
* Zodpovednosť za publikovanie údajov,
* Zodpovednosť za riešenia a ekosystém,
* Model kontroly a sledovania dosahovania výsledkov.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| „laissez-faire":   * bez spoločnej dátovej politiky, publikovanie na základe dobrovoľnosti a spolupráce |  | * Tento model fungoval do roku 2013 zo zlými výsledkami * Súčasná kultúra verejnej správy na Slovensku nie je nastavená pre dobrovoľné zdieľanie údajov a spoluprácu v dostatočnej miere |
| Hierarchický model („command and control“):   * centrálne určovanie politík a lokálne publikovanie údajov na základe nastavených cieľov |  | * Tento model teoreticky dokáže dosiahnuť stanovené KPI, zároveň však potláča kreativitu a iniciatívu zdola (zo strany inštitúcie) |
| Zmiešaný model:   * centrálne určovanie dátovej politiky a koordináciu * **iniciatíva a zodpovednosť** na strane inštitúcií * možnosť využívať centrálne nástroje pre identifikáciu a transformáciu údajov * možnosť využívať centrálne nástroje pre publikovanie údajov * možnosť využívať centrálne dátové služby | X | * pre celú verejnú správu * kombinácia dobrých vlastností jednotlivých modelov: centrálna zodpovednosť za výsledok a politiky a zároveň podpora iniciatívy jednotlivých inštitúcií * nadväzuje na súčasné riešenia (MOD) * možnosť zvoliť pre každý prípad vhodné riešenie na základe hodnoty za peniaze |
| Centralistický model:   * centrálne určovanie politík * a **centrálna autorita zodpovedná** za publikovanie údajov |  | * V rozpore s princípom, že za publikovanie údajov zodpovedá ich správca * Takéto riešenie **bude zvolené ako núdzové** v prípade, že sa nepodarí implementovať Akčný plán OGP 2017-2019 a na konci roka 2019 bude jasné, že nesplníme ciele tohto dokumentu |
| Model „outsourcing“:   * publikovanie údajov môže byť "outsourcované"  súkromným "data providerom" | X | * v prípade verejnej správy * v rámci otvorenosti bude možné využívať služby data providerov – ktorých vznik podporí dátovú ekonomiku |

### Spôsob zabezpečenia kvality otvorených údajov

Kvalita údajov má priamy dopad na možnosti využitia údajov v aplikáciách. Na základe definovaných kritérií v dokumente SP Manažment údajov je nastavený systém pre kontinuálne zlepšovanie kvality údajov v informačných systémoch verejnej správy. Zásadná otázka je, akým spôsobom sa bude sledovať, kontrolovať a zlepšovať kvalita publikovaných otvorených údajov.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Kvalita otvorených údajov bude riešená rovnakými nástrojmi ako Manažment údajov (ÚPPVIaI resp. NASES) | X | * Využitie nastavovaného mechanizmu pre Manažment údajov |
| Kvalita bude posudzovaná nezávislou komisiou (aj automatizovane) – kompetencia by bola odovzdaná komunite |  | * Takéto opatrenie bude zvolené, ako náhradné, ak sa nepodarí dosiahnuť stanovené ciele kvality údajov v roku 2020 |
| Kvalita bude posudzovaná nezávislou komisiou (aj automatizovane) – nezávislý Úrad |  | * Administratívne náročné opatrenie |
| Kvalita bude zvyšovaná službami pre spätnú väzbu a jej zakomponovanie do procesov Manažmentu údajov | X | * Spätná väzba komunity a používateľov údajov dokáže rýchlo a účinne zlepšiť kvalitu údajov zdola |

### Spôsob prístupu (access) k otvoreným údajom

Akým spôsobom sa používatelia dostanú k otvoreným údajom? Zámerom je čo najširšie možnosti pre rôzne typy používateľov.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| prostredníctvom štandardizovaného aplikačného rozhrania (API) – v kombinácií s dávkovým súborom | X | * Preferované |
| použitím news-feeedu (RSS/Atom) | X |  |
| cez SPARQL uzol | X |  |
| dávkové súbory | X |  |
| CD |  |  |

### Výber formátov pre sprístupnené otvorené údaje

| Variant | Úroveň interoperability | Príklad | Výber | Zdôvodnenie |
| --- | --- | --- | --- | --- |
| Proprietárne formáty | 1★, 2★ | PDF, XLS |  |  |
| Otvorené formáty | 3★ | CSV, XML, ODS, JSON | X | Pre existujúce otvorené údaje by bolo potrebné alokovať dodatočné finančné zdroje. |
| 4★, 5★ | RDF, OWL, TriX, JSON | X | Pre nové alebo inovované ISVS publikujúce otvorené údaje sa použijú na to alokované zdroje. |

### Miera centralizácie

Ktoré komponenty majú byť riešené na centrálnej úrovni a ktoré na lokálnej úrovni, aké sú prípustné kombinácie?

| Variant | Riešenie | Zdôvodnenie |
| --- | --- | --- |
| Model údajov | * Jeden centrálny: evidovaný v MetaIS | * Podľa SP Manažment údajov * Zabezpečenie interoperability v rámci SR |
| Manažment identifikácie | * Jeden centrálny: evidovaný v MetaIS | * Podľa SP Manažment údajov |
| Katalóg pre otvorené údaje | * Jeden centrálny „data.gov.sk“ | * Spoľahlivý jeden priestor pre komunitu používateľov, * Umožní rýchle hľadanie |
| Dátové úložisko | * Kombinácia: centrálne úložisko môže existovať, ale inštitúcie majú kompetenciu rozhodnúť o použití nástrojov, na základe: kritérií efektivity a kritérií kvality | * Možnosť voľby podporí efektivitu a kvalitu |
| Nástroje pre integráciu, transformáciu a publikovanie údajov | * Kombinácia: centrálny nástroj môže existovať, ale inštitúcie majú kompetenciu rozhodnúť o použití nástrojov, na základe: kritérií efektivity a kritérií kvality – pričom je možné využívať aj riešenia hybridného cloudu | * Možnosť voľby podporí efektivitu a kvalitu |

### Výber licenčných modelov

Licenčný model pre zverejňovanie údajov v otvorenej forme určuje, aké sú reštrikcie na použitie údajov (prepoužitie, kombinácia údajov, predaj údajov a podobne). V prípade dodržania princípov otvorených údajov sú povolené dva typy reštrikcií:

* attribution (by): používateľ údajov musí označiť pôvod údajov (tvorcu) a prisúdiť mu kredit,
* share-alike (sa): odvodený obsah z údajov musí byť zverejnený pod rovnakou licenciu.

**Licencia otvorených údajov musí byť vybraná povinne pre každý dataset.**

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| **Creative Commons CCZero (CC0)** | **X** | * **Bude zvolené ako „Default“** * Odporúčanie EK[[4]](#footnote-5) * Výborné pre potreby zabezpečená princípov otvorených údajov a podporu digitálnej ekonomiky |
| Open Data Commons Public Domain Dedication and Licence (PDDL) |  | * Vhodná pre databázy, bez výrazných výhod oproti CC0 |
| Creative Commons Attribution 4.0 (CC-BY-4.0) | X | * Iba na základe vážneho zdôvodnenia |
| Open Data Commons Attribution License (ODC-BY) |  |  |
| Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0) |  | * Nedostatočná podpora dátovej ekonomiky |
| Špeciálna licencia (na mieru) |  | * Neodporúčané riešenie vďaka lokalizmu, čo znamená nedostatočná podpora dátovej ekonomiky |

### Pôvod údajov

Údaje sú spravované v zdrojovom systéme a pre jednoduché použitie v rámci procesov verejnej správy a referencovanie údajov vznikli 4 referenčné registre:

* Register právnických osôb,
* Register fyzických osôb,
* Register adries,
* Register priestorových informácií.

Referenčné údaje (nie len z referenčných registrov) sú zároveň dostupné cez platformu integrácie údajov. Je potrebné navrhnúť mechanizmus, kto by mal prioritne zverejňovať otvorené údaje v prípade previazaných zdrojov údajov.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Zdrojový register | X | * Priamy dopad na kvalitu údajov má správca zdrojového registra |
| Referenčný register |  |  |
| Platforma integrácie údajov | X | * Možnosť jednoduchého a automatizovaného generovania datasetov z platformy (otázky integrácie budú vyriešené) |

### Ochrana súkromia a otvorené údaje

Zverejňovanie údajov vo formáte otvorených údajov naráža na ochranu súkromia osôb, ktorých sa týkajú tieto údaje. Veľa otázok vyvoláva najmä nové nariadenie GDPR. Je preto potrebné navrhnúť jasný mechanizmus, ako systematicky k tejto téme pristúpiť.

Rozsah zverejňovania údajov

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| **Zverejňujú sa anonymizované údaje** | X | * Údaje musia byť ochránené natoľko, aby ak nie je žiadúce nebolo možné identifikovať jednotlivca a jeho citlivé údaje |
| Osobné údaje a citlivé údaje sú vypustené z datasetu |  | * Otvorené údaje by stratili veľkú časť svojej hodnoty, čo povedie k menším možnostiam ich využitia |
| Osobné údaje a citlivé údaje sú vypustené z datasetu, okrem výnimiek:   * dostupné vo verejných listinách, * údaje štátnych zamestnancoch z dôvodu transparentnosti a podobne. |  | * Otvorené údaje by stratili veľkú časť svojej hodnoty, čo povedie k menším možnostiam ich využitia |

Anonymizácia údajov

Teoreticky existuje viacero spôsobov anonymizácie údajov a každý z nich má dopad na ďalšie využitie údajov, ako i mieru ochrany osobných údajov.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Agregácia údajov |  | * Účinná metóda, avšak veľmi znižujúca informačnú hodnotu údajov |
| „Top and bottom coding“ |  | * Úplne nepostačujúca metóda, kedy sa ochránia iba extrémne prípady a ostatné zostanú neanonymizované |
| Odstránenie osobných údajov z datasetov |  | * Otvorené údaje by stratili veľkú časť svojej hodnoty, čo povedie k menším možnostiam ich využitia |
| **Použitie pseudo-identifikátorov (crypographic hash)** | **X** | * Výrazne zvýšenie hodnoty otvorených údajov |
| Zahmlenie údajov (odchýlky, aby bolo možné štatistické spracovanie) |  | * Je potrebná komplexná metodika, nakoľko nedostatočné zahmlenie nezabezpečí anonymizáciu údajov |
| Vytvorenie metodiky pre minimálne štandardy ochrany a jej použitie | X | * Vzhľadom k veľkému počtu publikovaných údajov v čase, nie je možné kontrolovať ich prepojiteľnosť a tým možné odhalenie citlivých údajov prostredníctvom viacerých zdrojov. Je nevyhnutné metodicky nastaviť minimálne štandardy ochrany (napr. počet jednotlivcov v rôznych kombináciách, najvyššia prípustná úroveň detailu, ...) pri ktorých nenastane únik citlivých údajov aj pri použití a prepojení viacerých tabuliek. |

Prepájanie údajov, vo vzťahu k osobným údajom

Často sa stáva, že je možné osobné údaje získať vďaka prepájaniu údajov z rôznych dátových zdrojov. Je potrebné zvoliť prístup, ako riešiť tento problém a eliminovať tak zverejňovanie analýz s osobnými údajmi.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Prepájanie osobných údajov je limitované |  | * Takéto riešenie by brzdilo inovácie a rozvoj dátovej ekonomiky |
| Prepájanie osobných údajov je v zodpovednosti osoby, ktorá prepája údaje | X | * V súlade s platnou legislatívou na ochranu osobných údajov |
| Systematický dohľad nad prepájaním údajov (s využitím spätnej väzby) | X | * Výrazné zvýšenie dôveryhodnosti riešenia a samotnej ochrany založené na ex-post analýze zverejňovaných údajov * Zmiernenie obáv inštitúcií pre publikovanie otvorených údajov |

### Spôsob zabezpečenia dôveryhodnosti

Aby bolo možné zabezpečiť použiteľnosť otvorených údajov na právne účely, je potrebné zabezpečiť ich nepopierateľnosť. V praxi to znamená, že je potrebné vedieť, aké údaje boli publikované v danom časovom okamihu (bude tak možné odlíšiť v prípade sporu, či sa prípadnej chyby dopustila aplikácia interpretujúca údaje alebo inštitúcia, ktorá údaje interpretuje). Existuje viacero technologických a organizačných prístupov a voľba riešenia ovplyvňuje následnú budúcu architektúru.

| Variant | Výber | Zdôvodnenie |
| --- | --- | --- |
| Podpisovanie KEP | X | * Riešenie sa javí ako príliš ťažkopádne * Problematická dlhodobá udržateľnosť (nutnosť repodpisovania) * Riešenie bude zvolené v prípade, ak sa zvolená alternatíva ukáže ako nefunkčná |
| „blockchain“ | X | * Podpora inovácií a digitálnej ekonomiky * Bude si vyžadovať overenie vhodnosti - „Proof of Concept“ |
| Bez špecifických technických opatrení (Organizačné opatrenia – deklarácia, legislatíva, ...) |  | * Kultúralny problém s akceptáciou riešenia * V prípade sporu nebude jasný pravdivý stav |

## Prehľad najlepších skúseností

Nasledujúca kapitola uvádza prehlaď najlepších skúseností (best-practice) zo zahraničia, ktoré sú relevantné pre strategickú prioritu, respektíve sa z nich vychádza.

Portály pre otvorené údaje

* Open data portály v EU, UK a USA,
* Medzinárodný projekt COMSODE,
* European Public Sector Information (PSI) Platform.

Prepojené údaje (Linked dáta) a ontológie

* Ontológie - SEMIC odporučené ontológie pre popis Fyzickej osoby, Organizácie, Lokácie a podobne.

Dátová kvalita

* Metodika Five Star Open Data.
* Vychádza sa z odporúčaní the **Open Data Foundation, the DDI Alliance**, the International Household Survey Network, ktoré sa zameriavajú na kvalitu dát aj z pohľadu sekundárneho využitia.

Metadáta

* Kľúčové je propagovať dôležitosť kvalitnej dokumentácie dát pomocou štandardov na metadáta na základe najlepších praktík. Metadáta sa musia štandardizovať naprieč celou verejnou správou jednotným prístupom podporujúcim rôzne štandardizované metamodely (ISA2 odporučené ontológie, INSPIRE), pričom kľúčovým prvkom v ich správe je Centrálny metainformačný systém. Zakomponovanie popisu metadát do každej etapy životného cyklu – je potrebné dokumentovať v okamihu udalosti Metadáta možno ďalej využiť na automatické generovanie štatistických skriptov alebo vytváranie dokumentácie, čo znižuje celkové náklade a zvyšuje kvalitu.

# Architektúra riešenia

V kapitole sú navrhnuté princípy, ktoré budú vyžadované od riešení vo verejnej správe z pohľadu otvorených údajov.

Dôležitým predpokladom k dosiahnutiu systematického rozvoja otvorených dát, je dopracovanie nosných prvkov metainformačného systému. Medzi kľúčové považujeme doplnenie funkcionalít v oblasti štandardizácie dátových prvkov vo forme URI a ontológií. Táto zmena zaručí, že zverejnené datasety, budú môcť vychádzať zo štandardizovaných dátových prvkov, čo v konečnom dôsledku výrazným spôsobom zjednoduší integráciu a tým aj efektivitu použitia otvorených dát.

## Popis datasetov

Na popísanie datasetu publikovaného v minimálnej kvalite 4★ bude použitý medzinárodný štandard DCAT (Data Catalog Vocabulary) , ktorý poskytuje prostriedky pre definíciu katalógov, datasetov a ich distribúcii v čase. S ohľadom na interoperabilitu a odporúčania Európskej komisie budú požadované vlastnosti publikovaného datasetu formálne vyjadrené prostredníctvom tzv. Aplikačného profilu pre publikáciu otvorených dát verejnej správy Slovenskej republiky DCAT-AP-SK. Štandardizácia uvedených požadovaných vlastností datasetu obsahujúceho dáta s vysokou interoperabilitou je v súčasnosti predmetom pracovnej skupiny PS1.

Detailná špecifikácia datasetov a prislúchajúcich metaúdajov bude štandardizovaná v rámci pracovnej skupiny PS1 a zverejnená v podrobnej metodike.

## Princípy

Na základe spresnení Open Data Working Group, údaje publikované ako otvorené údaje musia spĺňať nasledovné princípy a podmienky, ktoré majú byť transponované aj do záväzných predpisov:

* Sú verejné.
* Sú prístupné.
* Sú úplné.
* Sú bez technologických obmedzení.
* Sú nediskriminačné.
* Sú opakovane použiteľné.
* Sú popísané.
* Sú včasné.
* Majú správcu.

Riešenie otvorených údajov bude tiež postavené na princípoch z Charty otvorených údajov (<http://opendatacharter.net/>), ktoré rozširujú tieto princípy o:

* Otvorené údaje prednostne:
* Je zabezpečená interoperabilita (prepojenie údajov):
* Podpora správy vecí verejných a participácie občanov:
* Podpora inkluzívneho rozvoja a inovácií:

Znamená to, že každý projekt, ktorý rieši otvorené údaje je navrhnutý tak, aby zabezpečil tieto princípy a každá inštitúcia, ktorá spravuje informačné systémy verejnej správy bude v súlade s týmito princípmi do konca roka 2020.

### Pravidlá pre interoperabilitu otvorených údajov verejnej správy SR

Aby bolo možné naplniť celkové stanové ciele pre kvalitu otvorených údajov definovanú v kapitole 2 ciele, je nutné uplatniť nasledovné pravidlá v súvislosti s novými datasetmi nasledovne:

Tabuľka : Pravidlá pre úrovne interoperability otvorených údajov verejnej správy

|  |  |
| --- | --- |
| Požadovaná úroveň | Dotknuté otvorené údaje |
| 3★ | Všetky existujúce otvorené údaje, ktoré nie sú súčasťou nových alebo inovovaných ISVS publikujúcich otvorené údaj |
| 4★ | Všetky nové a inovované ISVS publikujúce otvorené údaje  Maximálna úroveň pre identifikátory fyzických osôb |
| 5★ | Pre všetky nové a inovované ISVS publikujúce otvorené údaje s centrálnou povahou (referenčné údaje, centrálne registre, údaje v rozsahu MetaIS, resp. prioritné datasety, publikačné minimum samosprávy, publikované údaje a poskytované elektronické služby miest a obcí - DCOM) |

**Definícia:** „Všetkými novými a inovovanými ISVS publikujúcimi otvorené údaje“, sú pre účely definície dostupnosti údajov v štátnej správe označené také programové vybavenia, ktoré zabezpečujú tvorbu otvorených údajov a sú spolufinancované prostredníctvom verejných zdrojov, akými sú rôzne operačné programy (OPII, OPEVS a podobne), respektíve štátny rozpočet.

## Biznis vrstva

### Prehľad požiadaviek

Všeobecné požiadavky

* Publikovanie dát bude efektívne, jednoduché a transparentné, a čo v najvyššej možnej miere automatizované.
* Údaje budú identifikované pomocou URI, pričom údaje centrálnej povahy budú popísané prostredníctvom Centrálneho modelu údajov.
* Otvorené údaje musia spĺňať pravidlá pre úrovne interoperability otvorených údajov verejnej správy SR.
* Podnikatelia, verejnosť a kreatívne komunity sa budú môcť jednoducho zapojiť do procesu otvoreného vládnutia.

Požiadavky na centrálne riešenie

* Riešenie obsahuje centrálne automatické transformácie dát
* Riešenie umožňuje jednoducho prideliť licenciu dátovému zdroju.
* Existuje centrálny dátový model verejnej správy vrátane ontológií pre otvorené dáta.
* Bude efektívne monitorovaný dopad opatrení otvorených dát a v prípade nedosahovania cieľov budú dostupné nápravné opatrenia.
* Datasety je možné referencovať, zdieľať, vyhľadávať, komentovať a sú dostupné cez jednotný prístup.
* Budú dostupné neustále sa zlepšujúce a interaktívne metodiky pre publikovanie otvorených dát.
* Pomoc s interpretáciou pre identifikáciu vhodných datasetov.
* **Služby pre zverejňovanie otvorených údajov budú dostupné pre všetky subjekty, nie len pre orgány verejnej moci (**súkromné subjekty, neziskové organizácie a podobne**).**

Požiadavky na inštitúcie verejnej správy

* Implementácia procesov pre publikovanie otvorených údajov (v rámci čoho je kľúčový zber požiadaviek a ich vyhodnocovanie)

### Prehľad procesov

V nasledujúcej kapitole sú identifikované základné procesy pre správu otvorených údajov, ktoré je potrebné nastaviť: definícia, ako sa budú procesy vykonávať a kto bude za nich zodpovedný.

Obrázok : Biznis procesy pre otvorené údaje



Tabuľka 12: Procesy pre životný cyklus údajov

| Proces | Centrálne | Lokálne |
| --- | --- | --- |
| * Riešenie vecí verejných |  |  |
| * Informovanie verejnosti prostredníctvom otvorených údajov | * Realizácia | * Používanie |
| * Pripomienkovanie a hodnotenie datasetov | * Realizácia | * Používanie |
| * Poskytovanie otvorených údajov |  |  |
| * Príprava plánu pre publikovanie údajov | * Prieskum v rámci OGP * Je nevyhnutné vykonávať prieskumy v rámci kreatívnych komunít a inovatívnych podnikov, o ktoré dáta by mali najväčší záujem, a podľa toho prioritizovať v pláne publikovanie. Odporúča sa tiež sledovať Globálny index otvorených dát <http://index.okfn.org/place/>, ktorý formou crowd-sourcingu vyhodnocuje rozsah, kvalitu a užitočnosť datasetov zverejňovaných naprieč verejnými správami. | * Reakcia na priority prieskumu OGP * Návrh vlastných priorít |
| * Spracovanie údajov pre publikovanie | * Metodika | Realizácia:   * Príprava prístupu k dátovým zdrojom, extrakcia dát, * Transformácia dát na otvorené datasety, prípadne linked dáta, * Popis datasetov a dátových zdrojov metadátami, * Pridelenie licencie. |
| * Publikovanie údajov na portáli. | * Metodika * Správa katalógu | * Zverejnenie datasetov a ich sprístupnenie cez API a centrálny katalóg |
| * Aktualizácia publikovaných údajov. | * Metodika | * Aktualizácia datasetov podľa navrhnutých pravidiel |
| * Archivácia publikovaných údajov | * Realizácia |  |
| * Koordinácia spracovania otvorených údajov |  |  |
| * Správa centrálnych katalógov | * Otvorených údajov; * API; * Aplikácií nad otvorenými údajmi |  |
| * Návrh a spravovanie aplikačných rozhraní | * Metodika * Správa | * Používanie |
| * Dohľad nad kvalitou a dodržiavaním štandardov | Dohľad nad touto aktivitou bude mať Dátová kancelárie verejnej správy a pracovná skupina PS1 (štandardy).  Možno vydávať certifikáty kvality <https://certificates.theodi.org> |  |
| * Riadenie rizík | Je nevyhnutné nastaviť kompetencie a zodpovednosti za riadenie rizík a realizáciu plánovaných aktivít pre mitigovanie rizík. Pre každé riziko vedené v registri rizík musí byť definované, či je možné sa riziku vyhnúť, mitigovať ho alebo preniesť zodpovednosť na tretiu stranu napríklad prostredníctvom poistenia. Riadenie rizika však nesmie vyžadovať viac zdrojov ako očakávaná strata, spôsobená s istou pravdepodobnosťou a dopadom rizika. Preto pri výbere vhodných stratégií je nevyhnutné zobrať do úvahy aj očakávaný dopad a pravdepodobnosť rizika |  |
| * Správa URI identifikátorov | * Realizácia | * Používanie |
| * Správa centrálneho modelu údajov | * Realizácia | * Používanie |

## Aplikačná vrstva

V kapitole je uvedený prehľad aplikačných komponentov, ich funkcií a požiadaviek na nich.

### Koncept architektúry

Architektúra pre riešenie otvorených údajov je integrálnou súčasťou celkovej architektúry e-Govermentu a implementuje takzvaný hybridný model: centrálne riešenie katalógu otvorených údajov je postavené na báze open-source riešenia CKAN a kombinované s možnosťou využiť vlastné úložisko (na strane ISVS) alebo využiť ako službu možnosti centrálneho úložiska. Nástroje pre vyhľadávanie v datasetoch a archivácia datasetov bude riešená centrálne. K dispozícii bude tiež súbor nástrojov (pre transformácie, prepojené údaje a podobne) dostupné ako služby na centrálnej platforme, alebo ako framework pre inštaláciu do informačného prostredia inštitúcie. Znamená to, že každá inštitúcia získa k dispozícii súbor možností, s ktorých si môže vybrať podľa svojich potrieb z ohľadom na hodnotu za peniaze. Služby pre prácu s otvorenými údajmi (transformácie, uloženie a publikovanie otvorených údajov) je možné realizovať aj spôsobom SaaS. Poskytovateľ datasetu má možnosť vyhodnotiť ktorá možnosť je v jeho prípade najefektívnejšia a použiť ju.

### Prehľad aplikačných funkcií a komponentov

Obrázok 2: Prehľad aplikačných funkcií a komponentov



### Aplikačné požiadavky

**Zoznam požiadaviek:**

* Prideľovanie URI pre akékoľvek entity ISVS je riešené centrálne prostredníctvom Centrálneho metainformačného systému
* Riešenie centralizuje katalógy pre otvorené údaje, aplikácie a API a štandardizuje popisné metadáta,
* Riešenie podporuje ontológie ako aj centrálny a ad hoc dátový model,
* Nové dátové prvky je možné navrhovať dynamicky,
* Riešenie poskytuje nástroje na čistenie dát, obohacovanie dát, prelinkovanie dát a dátové transformácie,
* Riešenie obsahuje základné štatistické a analytické nástroje (aj nad RDF úložiskom),
* Riešenie podporuje distribuovanú prácu s dátami.

**Prepojené údaje:**

* Riešenie poskytuje nástroje interaktívnych vizualizácií pre linked dáta,
* Riešenie využíva URIs ako mená vecí,
* Riešenie využíva HTTP URIs, aby si používatelia mohli dané mená vyhľadať,
* Ak si niekto vyhľadá URI, riešenie poskytne užitočné informácie s využitím štandardov (RDF, SPARQL),
* Riešenie umožní pridať linky na ďalšie URIs, aby používatelia mohli objaviť ďalšie veci a súvislosti,
* Riešenie zabezpečí jednoznačné označenie objektov a zdrojov,
* Riešenie využije štruktúru webu,
* Riešenie uľahčí získavanie informácií o menovanom objekte alebo zdroji, pričom sa tieto informácie dajú ľahko parsovať,
* Riešenie zabezpečí dereferenciáciu URI prostredníctvom integrácie na Metainformačný systém,
* Riešenie uľahčí prelinkovanie akéhokoľvek súvisiaceho objektu alebo zdroja.

### Aplikačné služby

Služby pre verejnosť (pre prístup k datasetom)

Vyhľadávanie:

* Sémantické vyhľadávanie nad datasetmi, ktorý zvýši nájditeľnosť dát využitím sémantických vzťahov v dátach,
* Vyhľadávanie v dokumentoch.

Prístup k údajom (rozvoj existujúcich služieb):

* Poskytovanie raw datasetov,
* Poskytovanie obohatených datasetov,
* Poskytovanie prepojených údajov,
* Poskytovanie datasetov a služieb prostredníctvom open API,
* Poskytovanie dokumentov,

Použitie údajov:

* Overenie dôveryhodnosti datasetov (pre daný čas).
* Dereferenciácia URI prostredníctvom integrácie s Metainformačným systémom.

Tvorba aplikácií nad otvorenými údajmi (rozvoj existujúcich služieb):

* Správa aplikácií (v kontajneri pre aplikácie).

Služby pre prácu s údajmi (pre inštitúcie verejnej správy a ich informačné systémy)

Rozvoj a zlepšenie existujúcich služieb:

* Extrahovanie dát z interných a externých systémov,
* Spracovanie dát:
* konverzie a transformácie,
* čistenie,
* anonymizácia,
* obohacovanie,
* prelinkovanie.
* Uloženie publikovaných datasetov a metadát.

Nové služby:

* Archivácia datasetov,
* Zabezpečenie dôveryhodnosti dát,
* Validácia otvorených údajov (automatizovaná)
* Klasifikácia otvorených údajov na základe ich kvality (automatizovaná)
* Poskytovanie transformácií pre špecifické účely z 4-5 hviezdičiek aj do tabuľkových procesorov typu MS Excel, ktoré sú v štátnej správe veľmi rozšírené.

Nadstavbové služby

* Vizualizovanie dát (štandardná podľa CKAN) a interpretácia dát (existuje modul COGNOS, spokojnosť, neexistuje proces pre prenosu informácií z katalógu do BI nástroja)
* Náhľady na otvorené údaje, ktoré súčasne obsahujú vrstvu nezobrazovaných metadát. Tieto metadáta slúžia na sémantickú indexáciu obsahu, aké používa Google. Alebo Bing. Funguje to na základe poskytnutie schema.org schémy priamo na zobrazovanej stránke. Vyhľadávače tak majú dodatočné štruktúrované údaje čo spôsobí, že údaj bude nájditeľný aj cez všeobecné vyhľadávače.

### Aplikačné komponenty

Zoznam komponentov je nasledovný.

Centrálny katalóg (data.gov.sk)

* Aktualizácia katalógu CKAN.
* Katalóg obsahujúci štandardizované metadáta o položkách na [data.gov.sk](file:///C:/Users/Juraj/Downloads/data.gov.sk):
* **Otvorené údaje (datasety);**
* *API* ( bude nový komponent, ako inšpirácia slúži <https://www.data.gov/developers/apis>);
* **Aplikácie nad otvorenými dátami,**
* Vyhľadávanie datasetov a v datasetoch (aj **SPARQL-Query**) – ale žiadne dáta nie je možné nájsť
* **Funkcionalita pre hodnotenie datasetov, diskusie** *a prieskumy* (tvorba komunity) – čiastočne -*jednoduché prihlasovanie,*
* **Modul udalostí: upozorňovanie na nové datasety, nové verzie datasetov a podobne**,
* Požiadavky na publikovanie datasetov – je implementované, ale veľmi zložité na použitie (žiadosť podpisovaná ZEPom)

Úložisko otvorených údajov (sklad otvorených údajov)

Bezpečné, stabilné a dostupné úložisko pre otvorené údaje, ktoré bude podporovať:

* Katalóg metadát dátových zdrojov,
* Repozitár datasetov,
* RDF úložisko (distribuované),
* Archív otvorených údajov pre uchovávanie a interpretáciu historických údajov.

Úložisko pre otvorený zdrojový kód (štátny github).

Úložisko pre otvorené dokumenty (neštruktúrované dokumenty).

Centrálne služby pre prácu s údajmi

* „Blockchain“ mechanizmus: pôjde o „Proof of concept“ zavedenia technológie blockchain do prostredia slovenského e-Governmentu s dôrazom na trvácnosť, integritu a dôveryhodnosť poskytovania otvorených údajov. Využije sa tak distributed ledger technology (DLT) - “technológiu distribuovanej účtovnej knihy”, do ktorej budú odkladané jednotlivé datasety otvorených údajov. Základnou požiadavkou na funkčnosť je overenie platnosti údajov pre daný časový okamžik.
* Transformačný nástroj pre prepojené údaje: prevod údajov zo zdrojovej podoby (SQL databázy, CSV, XLS a podobne) do podoby prepojených údajov (a naopak). Tento problém nie je triviálny a riešenie musí ponúknuť vizualizačný nástroj, ako i prvky strojového učenia pre automatizáciu úlohy na základe údajov a ich schémy (aby dokázali vyberať prvky z centrálneho modelu údajov respektíve určovať pravdepodobnosť možných vhodných datasetov).
* Anonymizačné nástroje: prevod „citlivých údajov“ do anonymizovanej podoby na základe legislatívneho nastavenia (podporuje použitie pseudo-identifikátorov (crypographic hash) a zahmlenie údajov).
* Validačné služby
* Vizualizačné služby: tabuľky, grafy, mapovanie údajov.
* Správa aplikačných rozhraní pre tvorbu, publikovanie a spravovanie open API: vrátane nástroje pre tvorbu API pre import údajov do data.gov.sk.
* **Nástroje priestorové údaje:** možnosť vizualizácie priestorových údajov (aj časový vývoj a interpretácie priestorových údajov), tvorba priestorových vrstiev a objektov a automatizované exportovanie priestorových údajov do formátu pre Openstreetmaps (táto opensource platofma by potom mohla byť používaná ako mapové podklady pre webové sídla).
* Správa ontológií.
* Nástroje umelej inteligencie a dátová analýza: pokročilé nástroje pre interpretáciu údajov, hľadanie vzorov, vrátane anonymizácie a konverzie z pohľadu osobných údajov.
* Online chat, kde je možné sa dopytovať na pomoc pri tvorbe otvorených údajov. Samotný chat môže byť tvorený ako Question Answering systém, kde sa prostredníctvom metód strojového učenia stroj naučí odpovedať na otázky potenciálnych prispievateľov (vládna Siri pre otvorené údaje).

Meta-informačný systém verejnej správy

* Centrálny model údajov VS.
* Register referencovateľných identifikátorov.
* Register API pre služby.
* Dereferenciácia URI.

Kontajner aplikácii obsahujúci aplikácie vytvorené kreatívnymi komunitami a podnikateľmi

* Aplikácie môžu byť umiestnené priamo v dedikovanom kontajnery. Takéto aplikácie môžu pre svoj beh využívať:
* výpočtové prostriedky centrálnej platformy,
* framework zo základnými funkcionalitami.

Podporná aplikácia pre inštitúcie verejnej správy

**Podporná aplikácia pre povinné osoby**, ktorá obsahujúce:

* Nástroje na dopytovanie a transformáciu dát (ETL nástroje),
* Publikovanie datasetu do centrálneho katalógu,
* Nástroje na vizualizáciu otvorených dát,
* Framework (vo formáte otvoreného kódu, napríklad v Java) pre publikovanie otvorených údajov.

Služby dátovej integrácie

* Služby dátovej integrácie umožnia inštitúcií automaticky získať údaje zo svojho ISVS a nahrať ich priamo do centrálneho úložiska.

Komponenty na strane inštitúcie verejnej správy

* Lokálne úložisko datasetov.

## Technologická vrstva

Nové a aktualizované informačné systémy budú prevádzkované vo vládnom cloude. Služby integrácie údajov budú nasedené vo forme iPaaS, pričom k dispozícii budú centrálne poskytované služby dostupné cez definované aplikačné rozhrania a podporné služby realizované formou SaaS.

# Realizácia

Kapitola sa venuje návrhu praktických krokov, ktoré zabezpečia realizáciu navrhnutej koncepcie pre Otvorené údaje v praxi. Plánovaniu zavedenie správneho publikovania otvorených údajov sa venujú dva vládou schválené dokumenty:

* Akčný plán Iniciatívy pre otvorené vládnutie v Slovenskej republike na roky 2017 – 2019 (ďalej aj Akčný plán OGP).
* Stratégia a akčný plán sprístupnenia a používania otvorených údajov verejnej správy (ďalej aj Stratégia otvorených údajov).

Zámerom tohto dokumentu nie je nahradiť schválené stratégie, ale doplniť ich o detaily z pohľadu realizácie cieľov Národnej koncepcie informatizácie verejnej správy (NKIVS). Kapitola začína pomenovaním hlavných rizík a otvorených otázok, s ktorými bude nutné sa vysporiadať. Následne sú pomenované hlavné míľniky, ktoré je potrebné dosiahnuť. Nasleduje identifikácia kľúčových aktivít (z Akčného plánu OGP a Stratégie otvorených údajov), ktoré je vhodné vykonať na centrálnej a lokálnej úrovni. Inštitúcie verejnej správy tak získavajú jednoduchý prehľad čo môžu v nasledujúcom období očakávať a čo by mali vykonať. Na záver sú zhrnuté potrebné legislatívne požiadavky a zmeny, ktoré bude nutné presadiť.

## Riziká a otvorené otázky

Tabuľka : Prehľad rizík a spôsobov ich zmiernenia

| Riziko | Spôsob zmiernenia |
| --- | --- |
| Zadefinovanie datasetov nebude správne | * Zabezpečiť školenia v oblasti dátového modelovania pre zamestnancov verejnej správy (program školení, harmonogram, cieľová skupina) * Aktualizácia príručky pre publikovanie otvorených údajov spôsobom, aby bola zrozumiteľná a interaktívna * Dátový audit, dopyt |
| Návrh dátového modelu bude príliš komplexný a nepoužiteľný v praxi | * Zabezpečiť školenia v oblasti dátového modelovania pre zamestnancov verejnej správy |
| Nepodarí sa získať otvorenú licenciu pre kľúčové datasety zaujímavé pre inovatívne podniky ako meteorologické a geografické informácie | * Komplexné riešenie otázky * Investovať do zabezpečenia licencií a dát * Vysporiadanie licenčných otázok ako podmienka realizácie investícii z OPII |
| O datasety a open API nebude záujem (verejnosti, občianskych komunít a podnikateľov) | * Prioritne sa budú zverejňovať dôležité a žiadané údaje * Správna komunikácia a PR * Organizácia hackathonov * Súťaže a mikrogranty * Podpora vzniku inovatívnych aplikácií cez dopytovú výzvu (mikro projekty, rádovo 10 000 EUR), nízka bariéra, orientácia na HUBy * Prepojené údaje pre medzinárodnú interoperabilitu datasetov |
| Kvalita údajov bude nízka a údaje nebudú aktuálne | * Prepojené údaje „by default“ v nových projektoch * Čistenie údajov ako podmienka realizácie modernizačného IT projektu * Podpora tvorby interných kapacít pre Manažment údajov * Systematický benchmarking kvality údajov |
| Organizačné zabezpečenie sa nepodarí (na dátových kurátorov nebudú vyčlenené finančné prostriedky) | * Pre malé inštitúcie možnosť využitia zdieľaného dátového kurátora * Ekonomický prínos otvorených údajov bude vypočítaný na základe kvalitnej analýzy, ukáže sa hodnota za peniaze a začne sa rokovať s MF SR |
| Dátová kancelária nebude dostatočne pružná | * Dôraz na **synergiu** súvisiacich aktivít * Modelovať Dátovú kanceláriu na otvorených princípoch, flexibilite a agilite |

Tabuľka : Prehľad otvorených otázok a spôsobov ich riešenia

| Otvorené otázky | Spôsob riešenia |
| --- | --- |
| Ako vytvoriť efektívnu platformu s uspokojivou SLA pre zapojenie kreatívnych komunít a inovatívnych podnikateľov? | * Zabezpečiť otvorenú a transparentnú prevádzku a správu platformy otvorených údajov * Zabezpečiť použitie opensource nástrojov a vytvoriť komunity vývojárov * Podpora SaaS služieb pre publikovanie otvorených údajov |
| Ako zabezpečiť ochranu osobných údajov a údajov týkajúcich sa národnej bezpečnosti? | * Anonymizácia údajov pred publikovaním v otvorenom formáte * Jasné kompetencie dátových kurátorov v oblasti * Nastaviť proces pre riešenie problémov s osobnými údajmi (možnosť jednoducho poukázať na problém cez formulár) |
| Ako získať plnú podporu zo strany autorov informačných systémov pre opakované používanie PSI pre komerčné aj nekomerčné účely? | * Transparentný spôsob financovania publikovania otvorených údajov (cez dopytové výzvy) * Jasne deklarované architektonické požiadavky * Kvalitné metodické materiály * Redefinícia legislatívy ohľadom údajov |
| Ako získať dostatočné zdroje a kapacity, schopné zverejňovať a spravovať otvorené dáta a riadiť inovačné snahy? | * Potreba diskusie a rokovania s MF SR o pravidelnom financovaní otvorených údajov z rozpočtu |

## Plánovanie a migrácia (v súlade s akčným plánom)

Vo všetkých relevantných informačných systémoch verejnej správy bude realizované sprístupnenie údajov v otvorenej forme vhodnej na znovu-použitie v maximálnej možnej miere. Referenčné údaje budú automaticky prístupné ako otvorené dáta (tieto dáta tak bude možné použiť na právne účely). S cieľom maximalizovať úžitkovú hodnotu už vykonaných investícii do ISVS a s cieľom plniť akčný plán OGP, budeme dôsledne vyžadovať poskytovanie dát ako otvorené dáta od všetkých inštitúcii cez existujúce platformy, pričom bude prebiehať vylepšenie centralizovaného prístupu k týmto údajom.

Kľúčové míľniky

**Všetky inštitúcie verejnej správy sprístupnia otvorené údaje** (do konca roka 2020):

* Metodické materiály budú dostupné a schválené (do apríla 2018).
* Referenčné údaje budú prístupné ako otvorené údaje (do júla 2018) – znamená to, že z referenčných údajov bude automatizovane generovaný dataset, ktorý bude obsahovať údaje, ktoré je možné zverejniť.
* Prioritné datasety budú zverejnené (do januára 2018 ).
* Nové centrálne služby pre otvorené dáta budú prístupné (**do júla 2019**).
* Prepojenie inštitúcii verejnej správy s katalógom otvorených dát (do konca roka 2019).
* **Automatizované** zverejňovanie všetkých možných datasetov (do konca roka 2020).

### Aktivity a úlohy pre centrálnu úroveň

Tabuľka 15: Aktivity a úlohy pre centrálnu úroveň ÚPPVII

| Aktivita | Východisko | Dátum |
| --- | --- | --- |
| Vypracovať analýzu trhovej hodnoty a ekonomického potenciálu otvorených dát | AP OGP | 31.12.2017 |
| Zriadenie medzirezortnej pracovnej skupiny | Stratégia | 31.07.2017 |
| Vyhlásiť dopytovú výzvu pre publikovanie otvorených údajov | SP | **31.10.2017** |
| Vypracovať a predložiť návrh zákona o údajoch na rokovanie vlády SR | AP OGP | 31.12.2018 |
| Vypracovať štandardy pre tvorbu datasetov (novelizácia výnosu o štandardoch) | Stratégia | 31.12.2017 |
| Aktualizovať metodiky pre publikovanie otvorených údajov (vypracovať metodický pokyn k výnosu o štandardoch, týkajúci sa tvorby a zverejňovania otvorených údajov, obsahujúci vzorové príklady, ktorý dostatočne vysvetlí zásady publikovania datasetov) | Stratégia | 31.12.2017 |
| Navrhnúť licenčný rámec pre otvorené údaje | Stratégia | 31.12.2017 |
| Stanoviť rozsah údajov pre samosprávu | AP OGP | 30.6.2018 |
| Zabezpečiť, aby všetky referenčné údaje boli prístupné ako otvorené údaje | SP | 30.06.2018 |
| Vypracovať analýzu stavu v oblasti dátových kurátorov v rámci inštitúcií verejnej správy SR | Stratégia | 30.11.2017 |
| Menovať hlavného dátového kurátora (ako súčasť dátovej kancelárie verejnej správy) | Stratégia | 31.03.2018 |
| Zaistiť aplikáciu ontológií | Stratégia | 31.12.2017 |

Tabuľka 16: Aktivity a úlohy pre NASES (Úrad vlády SR)

| Aktivita | Východisko | Dátum |
| --- | --- | --- |
| Zaistiť odborné zaškolenie | Stratégia  SP | 31.12.2017 a každoročne |
| Realizovať dátový audit a publikácia výsledkov dátového auditu (ako otvorených dát) | Stratégia | 31.3.2018 |
| Navrhnúť nové centrálne služby (podľa aplikačnej architektúry) a model prevádzky a rozvoja centrálnych služieb pre publikáciu otvorených údajov | SP | 31.12.2017 |
| Implementovať zabezpečenie dôveryhodnosti datasetov | SP | 30.6.2019 |
| Implementovať zabezpečenie archivácie datasetov | SP | 30.6.2019 |

Tabuľka 17: Aktivity a úlohy pre ÚSpOS

| Aktivita | Východisko | Dátum |
| --- | --- | --- |
| Uskutočniť verejnú kampaň na propagáciu využitia datasetov zverejnených na portáli otvorených údajov a na podporu vývoja inovácií. | AP OGP | Priebežne  k 31.12. |
| Vykonať prieskum dopytu verejnosti po najžiadanejších datasetoch otvorených údajov | AP OGP | každoročne do 31. marca |

Tabuľka 18: Aktivity a úlohy pre DEUS

| Aktivita | Východisko | Dátum |
| --- | --- | --- |
| Navrhnúť riešenie pre zverejnenie otvorených údajov samospráv | SP | 01.03.2018 |
| Pripraviť riešenie pre zverejnenie otvorených údajov samospráv | SP | 31.12.2019 |

### Aktivity a úlohy pre inštitúcie verejnej správy

Nasledujúce kroky

Tabuľka 19: Aktivity a úlohy pre inštitúcie verejnej správy

| Aktivita | Východisko | Dátum |
| --- | --- | --- |
| Vytvoriť rolu “dátového kurátora” | Stratégia | 31.03.2018 |
| Zabezpečiť organizačnú jednotku pre riešenie otázok manažmentu dát a publikovania otvorených údajov | SP | 30.06.2018 |
| Realizovať dátový audit | Stratégia | 31.03.2018 |
| Vypracovať časový harmonogram vytvárania a sprístupňovania datasetov | Stratégia  AP OGP | 31.12.2017 a každoročne |
| Zverejniť prioritné datasety | Stratégia  AP OGP | 30.06.2018 |
| Vytvoriť dátový inventár a vytvoriť katalóg otvorených údajov | Stratégia | 30.06.2018 |
| Zaviesť procesy manažmentu údajov a publikovania otvorených údajov (realizovať procesné zmeny) | SP | 31.12.2018 |
| Vyriešiť licenčné otázky ohľadom údajov, ktoré spravuje subjekt | SP | 31.12.2018 |
| Zvýšiť kvalitu údajov spôsobom, aby ich bolo možné použiť na právne účely | SP | 31.12.2019 |
| Zabezpečiť automatizované prepojenie s katalógom otvorených údajov | SP | 31.12.2019 |
| Zverejniť otvorené údaje ako prepojené údaje (Sprístupňovať datasety s otvorenými údajmi v minimálnej kvalite úrovne 4 v zmysle štandardov ISVS) | Stratégia | 31.03.2018 a priebežne |

### Identifikácia projektov, ktoré je vhodné realizovať

Nasledujúci zoznam predstavuje projekty, ktoré sa javia ako vhodné, aby bolo možné dosiahnuť ciele definované v tomto dokumente. Pre identifikované projekty sa očakáva, že gestor posúdi ich vhodnosť, vypracuje reformný zámer (ak je potrebné) a štúdiu uskutočniteľnosti. Rozhodnutie o realizácii projektu sa vykoná práve na základe štúdie uskutočniteľnosti.

Programový zámer: Otvorené údaje (súčasť Programu Manažment údajov)

Tabuľka : Programový zámer Otvorené údaje

| Projekt | Možný zdroj financovania | Aktivity | Garant |
| --- | --- | --- | --- |
| *Každý modernizačný IT projekt* | OPII | * Kategorizácia údajov a prepojenie údajov * Zabezpečenie kvality údajov * Generovanie prioritných datasetov * Integrácia údajov * Zabezpečenie automatizovaného zverejňovania otvorených údajov | Inštitúcie verejnej správy |
| Projekty: zabezpečenia publikovania otvorených údajov | OPII (Dopytová výzva) | * Kategorizácia údajov a prepojenie údajov * Zabezpečenie kvality údajov * Generovanie prioritných datasetov * Integrácia údajov * Zabezpečenie automatizovaného zverejňovania otvorených údajov (v prioritnom rozsahu) | ÚPPVIaI |
| Inovatívne využitie otvorených údajov | OPII (Dopytová výzva) | * Použitie otvorených údajov v zmysle cieľov: * pre aplikácie s pridanou hodnotou, * pre zvýšenie transparentnosti verejnej správy, * pre obohatenie údajov. | ÚPPVIaI |
| Centrálne služby pre otvorené údaje | OPII | Podpora rozvoja služieb:   * Aktualizácia a modernizácia centrálneho katalógu * Centrálne úložisko pre datasety * Služby a nástroje pre prepojené údaje * Služby pre anonymizáciu * Služby pre spracovanie priestorových údajov * Služby archivácie a zabezpečenia dôveryhodnosti a právoplatnosti datasetov | NASES |
| Dátová integrácia | OPII | * Zabezpečenie generovania otvorených údajov z referenčných údajov * Realizácia integrácie údajov | ÚPPVIaI |
| Otvorené údaje v zdravotníctve (ako súčasť konsolidačného projektu údajovej základne) | OPII | * Kľúčové datasety zo sektora zdravotníctva budú dostupné ako otvorené údaje * Zabezpečenie automatizovaného zverejňovania otvorených údajov | MZ SR |
| Otvorené údaje v školstve (ako súčasť konsolidačného projektu údajovej základne) | OPII | * Kľúčové datasety zo sektora školstva budú dostupné ako otvorené údaje * Zabezpečenie automatizovaného zverejňovania otvorených údajov | MŠŠVaV SR |
| Manažment údajov Sociálnej poisťovne | OPII | * Kľúčové datasety zo Sociálnej poisťovne budú dostupné ako otvorené údaje * Zabezpečenie automatizovaného zverejňovania otvorených údajov | Sociálna poisťovňa |
| Otvorené údaje v samospráve | OPII | * Publikačné minimum pre samosprávu bude prístupné ako otvorené údaje (pre subjekty zapojené do DCOM) * Zabezpečenie automatizovaného zverejňovania otvorených údajov | DEUS |
| Otvorené údaje v sektore životného prostredia | OPII | * Kľúčové datasety zo sektora životného prostredia budú dostupné ako otvorené údaje * Zabezpečenie automatizovaného zverejňovania otvorených údajov | MŽP SR |

### Požiadavky na legislatívne zmeny

Legislatíva

Zoznam požiadaviek na legislatívne zmeny, ktoré sú potrebné pre realizáciu strategickej priority (vrátane podzákonných predpisov):

| Legislatívny predpis | Legislatívna požiadavka | Garant | Termín ukončenia |
| --- | --- | --- | --- |
| Metodika „Open Data Ready“ | Ide o súbor kritérií podstatných pre možnosť publikácie otvorených dát automaticky informačnými systémami povinnej osoby. Metodika obsahuje podrobný popis týchto kritérií, postupy ich dosahovania a postup vyhodnotenia pre plnenie týchto kritérií. Kľúčovým kritériom metodiky je vysporiadanie si autorských práv údajov uložených v databázach informačných systémov s ich dodávateľom ako aj získanie súhlasu dodávateľa na zverejňovanie údajov uložených v databázach ako otvorených údajov pod licenciou CC0. Dodržiavanie metodiky Open Data Ready bude podmienkou pre čerpanie financií zo štrukturálnych fondov EÚ. | ÚPPVII | December 2018 |
| Nový zákon o údajoch vo verejnej správe | Nový komplexný zákon, ktorý prehľadne transponuje európsku direktívu (Directive 2003/98/EC, známa ako 'PSI Directive') a **navrhne povinnosť pre inštitúcie verejnej správy zverejňovať všetky údaje, ktorými verejná správa disponuje (okrem výnimiek) vo formáte prepojených otvorených údajov.** | ÚPPVII | Február 2018  Účinnosť zmien je žiadúca od Januára 2019 |
| Aktualizácia Výnosu 55/2014 o štandardoch ISVS | Upravuje výnos na základe trendov rozvoja otvorených dát, predovšetkým v oblasti linked dáta a ontológií. | ÚPPVIaI | 31.12.2017. |
| Licenčný model | Bude obsahovať presné vymedzenie používania licencií pre PSI, pričom primárne sa odporúča používať CC0 a následne CC-BY.  Pripraví sa aktualizovaný licenčný model, ktorý bude odrážať nastavené princípy a vyrieši otázky verejných dát, ktoré v súčasnosti nie sú pod otvorenou licenciou. | ÚPPVII | 31.12.2018 |

Požiadavky, ktoré by mali byť zabezpečené v rámci Operačného programu Efektívna verejná správa

* Návrh a nastavenie nových procesov pre manažment údajov vo verejnej správe a samospráve pre potreby otvorených údajov.
* Podpora zriadenia a vybudovanie kapacít Dátovej kancelárie verejnej správy pre problematiku otvorených údajov.
* Zefektívnenie komunikácie s občianskymi kreatívnymi komunitami a inovatívnymi podnikateľmi.
* Návrh opatrení pre podporu občianskych komunít a inovatívnych podnikateľov.

# Udržateľnosť a financovanie

## Zmysluplná investícia

Prínos realizovanej investície

Navrhnuté zmeny budú realizované ako program, financovaný vo veľkej miere z Európskych štrukturálnych fondov. Vďaka tejto investícii dokážeme:

* Vybudovať potrebné kapacity (ľudské zdroje) a priniesť „know-how“ v oblasti Otvorených údajov (spoločne s Manažmentom údajov),
* Realizovať zmeny v procesoch (v každej inštitúcií verejnej správy) a prijať potrebné legislatívne opatrenia (nový zákon o údajoch),
* Navrhnúť štandardy a metodiky,
* Implementovať potrebné nové komponenty a upraviť existujúce informačné systémy v zmysle navrhnutých princípov a požiadaviek.

Požiadavky na udržateľný prístup k projektom

Pre zmysluplné nastavenie projektov v rámci **programu Otvorené údaje** je potrebné (na základe týchto kritérií sa budú posudzovať projekty, ktoré riešia Otvorené údaje):

* aby bola dosiahnutá maximálna **hodnota za peniaze**,
* aby riešenie podporilo **inovačný potenciál** dátovej ekonomiky,
* aby projekt maximalizoval **príspevok k navrhnutým cieľom a merateľným ukazovateľom** SP Manažment údajov a SP Otvorené údaje,
* aby **prevádzkové náklady** riešenia boli minimálne (zdrojový kód výsledkov projektu pod otvorenou licenciou, využitie open source komponentov, flexibilne nastavené SLA, úpravy v IS, ktoré nevyvolajú ďalšie náklady),
* aby vďaka výsledkom projektu boli datasety generované automaticky aj po skončení projektu,
* **agilné riadenie implementácie:** aby riešenie bolo vyvíjané na základe agilnej metodológie a vývoj stál na testovaní prototypov.

Pre podporu udržateľného prístupu k projektom z programu Otvorené údaje sa zavedie **transparentný benchmarking** projektových a prevádzkových nákladov (medzi inštitúciami verejnej správy).

## Dopady navrhovaných zmien

Organizačné dopady Otvorené údaje

Na inštitúciu verejnej správy:

* Pozícia dátového kurátora respektíve útvar pre publikovanie otvorených údajov:
* 1 dátový kurátor pre každú inštitúciu, možnosť zdieľania kurátorov,
* veľkosť útvaru od 0 do 6 podľa veľkosti inštitúcie
* Zavedenie nových procesov pre publikovanie (je potrebné, aby sa procesy vykonávali pravidelne a neustále aj po skončení údajov).

Na garanta politiky Otvorených údajov (ÚPPVIaI):

* Pozícia hlavného dátového kurátora,
* Pozícia zodpovedná sa **politiky otvorených údajov,**
* Pozícia zodpovedná za realizáciu akčného plánu OGP,
* Pozícia zodpovedná za Programový manažment Otvorených údajov
* Rozšírenie Dátovej kancelárie verejnej správy o tím zodpovedný za otvorené údaje (plus 3).

Na prevádzkovateľa centrálnych komponentov:

* Pozícia architekta centrálnych komponentov pre otvorené údaje,
* Tím technickej podpory centrálneho riešenie.

Finančné dopady programu Otvorené údaje

Na inštitúciu verejnej správy:

* Náklady spojené s prevádzkou IT riešenia (lokálne komponenty),
* Náklady na mzdy zamestnancov (dátových kurátorov resp. jednotky pre publikovanie otvorených údajov),
* Náklady na externú asistenciu,
* Výpadok príjmov za predaj údajov,
* Náklady na podporu komunity.

Na garanta politiky Otvorených údajov:

* Náklady na mzdy zamestnancov,
* Náklady na externú asistenciu,
* Náklady na podporu komunity,

Na prevádzkovateľa centrálnych komponentov:

* Náklady na prevádzku centrálneho riešenia,
* Náklady na poskytovanie centrálnych dátových služieb inštitúciám verejnej správy (outsourcing pozícií, SaaS),
* Náklady na podporu komunity.

Situácia po skončení programu Otvorené údaje

Prínosy:

* Zverejňovanie otvorených údajov vytvára tlak transparentnejší a tým pádom **efektívnejší výkon verejnej moci** (dá sa odhadnúť úspora pre jednotlivé segmenty),
* Vyššia dôvera vo verejnú moc (kvalita podnikateľského prostredia a občianskeho práva),
* Rozvoj dátovej ekonomiky a úspešné firmy: Dátová ekonomika generuje daňové príjmy (za podmienky vzniky významných inovácií a funkčného ekosystému, ktorý spracováva otvorené údaje),
* Nad otvorenými údajmi vzniknú služby s pridanou hodnotou: pre občanov a pre biznis, ktorý sám môže vďaka otvoreným údajom fungovať efektívnejšie respektíve ponúkať kvalitnejšie služby,
* Rozvoj ľudského kapitálu a rast know-how vo vzdelanostnej ekonomike.

Prevádzka:

* Prevádzkové zmluvy systémov otvorených údajov musia brať do úvahy snahu **minimalizovať prevádzkové náklady,**
* Pravidelná organizácia hackathonov a súťaží (aj po skončení projektu).
* Tok datasetov nebude prerušený a centrálny model údajov bude neustále aktualizovaný.
* Zabezpečí sa obnova riešenia, jeho aktualizácia a potrebný nárast úložiskovej kapacity.

## Odporúčania z ohľadom na udržateľnosť

Princípy

Aby riešenie dlhodobo fungovalo, odporúčame:

* Zabezpečiť dostatočné zdroje potrebné na publikovanie otvorených údajov (minimálne 1 milión EUR ročne).
* Systematicky podporovať personálne obsadenie a kvalitu ľudského kapitálu (odborné vzdelávanie a kurzy, osobné ohodnotenie - podrobnejšie).
* Systematicky budovať „know-how“ inštitúcie (vrátane medzinárodnej spolupráce na úrovni EÚ).
* Neustále aplikovať najnovšie metódy a technológie (princípy kontinuálneho zlepšovania).
* **Zabezpečiť otvorenú a transparentnú prevádzku a správu platformy otvorených údajov (**princípy**).**
* Súťaž medzi inštitúciami v komunikácií s používateľmi údajov.

**Nasledujúce kroky pre zabezpečenie udržateľnosti**

* Zabezpečenie tabuľkových miezd dátových kurátorov sa dá **dosiahnuť rokovaním** s MF SR na základe Analýzy trhovej hodnoty otvorených údajov.
* Príprava licenčného modelu musí zohľadniť tiež ekonomický aspekt (nové finančné toky).

# Prílohy

Ako prílohy sú v dokumente zahrnuté:

* Skratky a definície.

## Skratky a definície

Definície

| Odborný termín | Definícia |
| --- | --- |
| Ontológia | Ontológia umožňuje špecifikovať koncepty a vzťahy medzi nimi, ktoré spoločne charakterizujú istú doménu záujmu, a to otvoreným a zrozumiteľným jazykom.. Napríklad v oblasti biológie môže zadefinovať tri triedy ako Zvieratá, Ryby a Cicavce, pričom Zvieratá sú nadradenou triedou. Každá ryba je tým pádom aj zvieraťom. Všeobecnou vlastnosťou triedy Zvieratá môžu byť napríklad Zmyslové orgány. Oko bude potom podriadenou vlastnosťou Zmyslových orgánov. |
| Otvorené údaje (Open Data) | Otvorené údaje (OpenData) je názov pre paradigmu, v ktorej je možné s údajmi, ktoré sú vytvárané a spracovávané v dátových zdrojoch verejnej správy, voľne pracovať. To znamená, že údaje musia byť dostupné, zrozumiteľné a práca s nimi musí byť možná bez obmedzení.  V kontexte verejnej správy to znamená nasledovné:   * Otvorené údaje môžu byť zverejňované v rôznych úrovniach kvality podľa takzvaného Tim Berners-Lee modelu označovaného ako five star model * pre každú organizáciu je dostupný zverejnený zoznam jej dátových zdrojov a ich základný popis, * obsah dátového zdroja (samotné údaje) sú proaktívne bezplatne sprístupnené neobmedzenému okruhu záujemcov, * údaje sú sprístupnené v tvare umožňujúcom ďalšie automatizované spracúvanie, čo sa dosiahne dodržaním stanovených technologických štandardov, ktorých základom je používanie otvorených a technologicky neutrálnych riešení, * ďalšiemu používaniu údajov (napr. ich spracovaniu, zverejneniu, spájaniu s inými údajmi) nie sú kladené žiadne legálne prekážky. |
| Public Sector Information (Informácie verejného sektora) | PSI sú definované v Directive 2013/37/EU, ktorá určuje jednotný trh pre informácie verejnej správy.  Informácie verejného sektora zahŕňajú akýkoľvek obsah bez ohľadu na médium (v papierovej podobe alebo elektronickej forme alebo založený na zvukových, vizuálnych alebo audiovizuálnych záznamoch) vytvorení verejným sektorom v rámci jeho agendy a kompetencií s výnimkou:   * dokumentov verejnoprávnych médií, * dokumentov vzdelávacích výskumných inštitúcií (školy, univerzity, knižnice a podobne), * dokumentov kultúrnych inštitúcií (múzea, knižnice, archívy, orchestre, opery, balety, divadlá). |
| Transformácia | Predstavuje sled transformačných predpisov reprezentujúcich celý proces spracovania otvorených dát. Predpisy umožňujú rozdelenie transformácie na spracovacie jednotky (data processing unit DPU) a ich reťazenie. Výsledkom sú spracované alebo obohatené otvorené dáta s príslušnými metadátami a prípadný záznam v katalógu. |
| Big Data | Pojem Big Data predstavuje obrovské množstvo dát rôzneho typu získavaných s vysokou rýchlosťou z vysokého počtu rôznych typov zdrojov. Big Data sa vyznačujú 4 charakteristikami: Objem, Rôznorodosť (napríklad videá, tweety, dáta zo senzorov), Rýchlosť (spracovávanie tokov dát v reálnom čase) a Hodnovernosť (s množstvom zbieraných dát sa spája neistota o ich presnosti). Spracovávanie takého množstva rôznorodých dát zbieraných v reálnom čase si vyžaduje nové nástroje a metódy ako výkonné procesory, nové softvérové riešenia a algoritmy. |
| Dataset | Ucelená a samostatne použiteľná skupina súvisiacich údajov vytvorených a udržiavaných na určitý účel a uložených spoločne podľa rovnakej schémy. |
| Dátový zdroj | Pôvodné miesto evidencie datasetu. 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. Dátovým zdrojom môžu byť aj zariadenia produkujúce dáta v internete vecí, senzory alebo aplikačné rozhrania sociálnych sietí. |
| Metadáta | Štruktúrované údaje obsahujúce informácie o primárnych údajoch – dátovom zdroji, pričom primárne údaje spravidla reprezentujú určitý hmotný objekt alebo nehmotný objekt. sú určené najmä na vyhľadávanie, katalogizáciu a využívanie primárnych údajov. |
| Mikrodáta | V kontexte štatistiky predstavujú mikrodáta informáciu na úrovni individuálnych respondentov, ktorí sa zúčastňujú prieskumov alebo sčítaní ľudu. |
| Paradáta | Paradáta sú dáta dokumentujúce proces zberu údajov. V prípade vykonania prieskumu môže ísť o informácie zahŕňajúce časy, kedy sa vykonávali pohovory, dĺžku trvania pohovorov, spôsob komunikácie, ochotu respondenta a mnohé ďalšie. |
| Prístup cez aplikačné rozhranie - API | Štandardizovaný spôsob vzájomnej strojovej komunikácie, charakteristický určením koncového bodu poskytovania služby, množinou funkcií v tomto bode dostupných, ich parametrami a formátmi vymieňaných správ. |
| Jednotný referencovateľný identifikátor | Jednotný referencovateľný identifikátor (URI) je množina znakov tvoriacich jednotným spôsobom identifikátor ľubovoľných logických a fyzických objektov. |
| Surové dáta | Surové dáta, tiež nazývané primárne dáta, sú dáta získané priamo z dátového zdroja, ktoré nie sú spracované na použitie pre určitý konkrétny účel (ešte neprešli spracovaním ani žiadnou inou manipuláciou). |

Skratky

| Skratka | Popis |
| --- | --- |
| BIA | Hodnotenie vplyvov regulácie na podnikateľské prostredie (Business impact assessment) |
| CDS | Centrálny dátový sklad verejnej správy |
| CKS | Centrálny konsolidačný systém |
| CSRÚ | Centrálna správa referenčných údajov |
| DCOM | Dátové centrum obcí a miest, teda nadrezortný informačný systém verejnej správy, ktorý poskytuje obciam technické a programové prostriedky na výkon verejnej moci elektronicky, na prevádzkovanie informačných systémov verejnej správy v ich správe a na zabezpečenie základných činností v oblasti elektronického výkonu vnútorných agend a prevádzku ostatných informačných systémov, ktoré obec používa. |
| DEUS | Datacentrum elektronizácie územnej samosprávy Slovenska, ktorého jedinými členmi sú ministerstvo financií a Združenie miest a obcí Slovenska. DEUS je zároveň správcom DCOM. |
| EIA | Hodnotenie vplyvov na životné prostredie (Environmental impact assessment) |
| ETL | (Extract, Transform, Load) |
| FIA | Hodnotenie vplyvov na rozpočet (Financial impact assessment) |
| IČ DPH | Identifikačné číslo pre daň z pridanej hodnoty |
| IAM | Systém na správu identít a oprávnení (Identity and Access Management) |
| IFP | Inštitút finančnej politiky |
| IKT | Informačno-komunikačné technológie |
| ISVS | Informačný systém verejnej správy |
| IVP | Inštitút vzdelávacej politiky |
| KDP | Katalóg dátových prvkov |
| KPI | Kľúčový výkonnostný ukazovateľ (Key performance indicator) |
| MDM | Správa kmeňových údajov (Master data management) |
| MetaIS | Metainformačný systém |
| MSP | Malí a strední podnikatelia |
| MUK | Modul úradnej komunikácie |
| OP EVS | Operačný program Efektívna verejná správa |
| OP II | Operačný program Integrovaná infraštruktúra |
| OVM | Orgán verejnej moci |
| PgK | Programová kancelária |
| RA | Register adries |
| RIA | Hodnotenie vplyvov regulácie (Regulatory impact assessment) |
| RFO | Register fyzických osôb |
| RPI | Register priestorových informácií |
| RPO | Register právnických osôb |
| SIA | Hodnotenie sociálnych vplyvov (Social impact assessment) |
| SSO | Jednotné prihlásenie (Single Sign-On) |
| URI | Jednotný referencovateľný identifikátor (Uniform resource identifier) |
| ÚPPVIaI | Úrad podpredsedu vlády SR pre informatizáciu a investície |

1. Doplňujúca informácia: Uvedená definícia stupňov interoperability (kvalita otvorených údajov z pohľadu ich strojového spracovania) spresňuje a nahradzuje pôvodnú definíciu úrovní kvality v dokumente Strategická priorita – Manažment údajov. Pravidlá priamo napomáhajú plniť stanové ciele definované v SP manažment údajov, kde je definované, že 70% publikovaných datasetov musí byť v súlade s Centrálnym modelom údajov verejnej správy. Pôvodne bola úroveň, kde bolo potrebné byť v súlade s Centrálnym modelom označená ako úroveň 4★, v súčasnosti je to správne označené ako úroven 5★. Ako je už uvedené v predchádzajúcej tabuľke, úroveň je o používaní URI pre popis a identifikáciu údajov, ktoré sú vedlajšej povahy a nepatria do centrálnych údajov s vysokým stupňom na znovapoužitie. [↑](#footnote-ref-2)
2. Aknowledgement to the European Public Sector Information (PSI) Platform [↑](#footnote-ref-3)
3. Skutočne potrebné kapacity sa môžu líšiť na základe rozsahu agendy jednotlivých inštitúcií verejnej správy (vo vzťahu k spracovaniu údajov) [↑](#footnote-ref-4)
4. <http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?action=display&doc_id=6421> [↑](#footnote-ref-5)