

Strategická priorita *Open Data*

Pracovná verzia
(Verzia 0.2)

Informácia o dokumente

Názov:	Strategická priorita Open Data
Stav:	Prvá verzia na pripomienkovanie
Pripravil:	Dagmar Bošanská, Juraj Bárdy,
Verzia:	0.2
Dátum:	30.9.2015
Pripomienkoval:	AKVS
Dátum revízie:	29.1.2016

Distribučný zoznam

Od	Dátum	Kontakt

Pre	Akcia*	Dátum (do)	Kontakt

* Akcia: *Schváliť, Pripomienkovať, Informovať, Realizovať, iné (uved'te)*

História verzii

Verzia	Dátum verzie	Pripravil/ Zmenil	Pripomienkoval	Kľúčové zmeny

Obsah

1	Úvodný pohľad	3
1.1	Skratky a definície	3
1.2	Definícia strategickej priority	7
2	Správa otvorených údajov	13
2.1	Ciele realizácie	15
2.1.1	Architektonické ciele	15
2.1.2	Stakeholderi a ich záujmy	16
2.2	Organizácia	18
2.2.1	Zodpovednosť	18
2.2.2	Organizačné zmeny	18
2.3	Stratégia a riešenie	19
2.3.1	Popis súčasného stavu	19
2.3.2	Najlepšie skúsenosti	21
2.3.3	Strategický prístup k riešeniu	22
2.3.4	Biznis vrstva	23
2.3.5	Aplikačná vrstva	25
2.3.6	Technologická vrstva	27
2.4	Problémy a riziká	27
2.5	Legislatívne požiadavky	27
2.6	Plánovanie a migrácia	28
2.6.1	Centrálny dátový sklad	29
2.6.2	Otvorené údaje pre samosprávu	29
2.6.3	Open API	31
2.6.4	Dopytový projekt pre inovatívne komunity a podnikateľov	32
2.6.5	Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov	33

1 Úvodný pohľad

1.1 Skratky a definície

Definície

Odborný termín	Definícia
Aplikačné rozhrania (API)	API popisuje SW komponent prostredníctvom jeho operácií (metód), vstupov, výstupov a dátových typov – ide o súbor služieb. API je v princípe nezávislé na konkrétnej implementácii a umožňuje tak integráciu stavebných blokov a uľahčuje vývoj aplikácií.
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 takeého množstva rôznorodých dát zbieraných v reálnom čase si vyžaduje nové nástroje a metódy ako výkonné procesory, nové softvérové riešenia a algoritmy. Big Data sú ďalej riešené v samostatnej strategickej prioritě „Riadenie údajov a Big Data“
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.
Linked dáta	Linked dáta predstavujú najvyššiu úroveň kvality Otvorených údajov (podľa Tim Berners-Lee modelu označovaného tiež ako five star model). Predstavuje spôsob publikácie štruktúrovaných údajov, ktoré sú sémanticky prepojené a môžu byť jednoducho znovu-používané a prehľadávané prostredníctvom queries. Linked dáta sú postavené na štandardných webových technológiách ako HTTP, RDF a URI, ktoré sa využívajú na automatizovanú komunikáciu informačných systémov. Základnými princípmi formátu Linked data sú: <ul style="list-style-type: none"> – Použitie URI pre pomenovanie (identifikáciu) objektov, – Použitie HTTP URIs pre prehľadávanie, – Poskytovanie informácií o tom, čo sa skrýva pod pomenovaním cez nástroje ako RDF a SPARQL, – Referencia cez HTTP URI k iným objektom.

Odborný termín	Definícia
Ontológia	<p>Ontológia umožňuje špecifikovať koncepty a vzťahy medzi nimi, ktoré spoločne charakterizujú istú doménu záujmu, a to otvoreným a zrozumiteľným jazykom.. Najjednoduchším príkladom ontológie je jazyk RDFS, ktorý umožňuje ontológovi vytvoriť jednoduchú hierarchiu konceptov a hierarchiu vlastností. Napríklad v oblasti biológie môže zadať tri triedy ako Zvieratá, Ryby a Cicavce, pričom Zvieratá sú nadradenou triedou. Každá ryba je tým pádom aj zvieratom. Všeobecnou vlastnosťou triedy Zvieratá môžu byť napríklad Zmyslové orgány. Oko bude potom podriadenou vlastnosťou Zmyslových orgánov.</p>
Otvorené údaje (Open Data)	<p>Otvorené údaje (Open Data) 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.</p> <p>V kontexte verejnej správy to znamená nasledovné:</p> <ul style="list-style-type: none"> ■ 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.
Otvorený formát	<p>Otvorený formát je publikovaná špecifikácia na uchovávanie digitálnych údajov, obvykle udržiavaná neproprietárnou štandardizačnou organizáciou a bez právnych obmedzení na používanie. Napríklad, otvorený formát musí byť implementovateľný aj proprietárnym aj open source softvérom, za použitia im typických licencií. Na rozdiel od otvorených formátov, proprietárne formáty sa riadia a definujú súkromnými záujmami.</p> <p>Primárnym cieľom otvorených formátov je garantovať dlhodobý prístup k údajom bez súčasnej alebo budúcej neistoty ohľadom legálnych práv alebo technickej špecifikácie. Sekundárnym cieľom otvorených formátov je podporovať konkurenciu medzi ich implementáciami.</p>

Odborný termín	Definícia
Public Sector Information (Informácie verejného sektora)	<p>PSI sú definované v Directive 2013/37/EU, ktorá určuje jednotný trh pre informácie verejnej správy.</p> <p>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:</p> <ul style="list-style-type: none"> – dokumentov verejnoprávnych médií, – dokumentov vzdelávacích výskumných inštitúcií (školy, univerzity, knižnice a podobne), – dokumentov kultúrnych inštitúcií (múzea, knižnice, archívy, orchestre, opery, balety, divadlá).
Referencovateľný identifikátor	<p>Ide identifikátor dátového zdroja, ktorý referencuje samotné údaje, t.j. po zadaní identifikátora do prehliadača je používateľ presmerovaný na URL vyplnené v metaúdajoch dátového zdroja, napr.:</p> <p>http://data.gov.sk/data/uvo/zmluvy/zmluvy</p> <p>Referencovateľným identifikátorom je teda identifikátor, ktorý:</p> <ul style="list-style-type: none"> má formát Uniformed Resource Identifier (URI), je jednoznačný, je unikátny, je dlhodobo stabilný, je formátovo a štrukturálne konzistentný, je manažovateľný tak, aby umožňoval logicky rozširovať stanovenú štruktúru, je jasný, stručný a krátky, je pre fyzickú osobu jednoducho čitateľný, je bez súborových prípon, neobsahuje programátorské kľúčové slová, neobsahuje reťazec „www“, neobsahuje interpunkciu okrem znakov lomka, pomlčka a bodka, diakritiku a medzery, okrem identifikátora fyzickej osoby podľa osobitného predpisu, kde je možné použiť interpunkciu a diakritiku, obsahuje iba malé písmená, nahrádza špeciálne znaky, napríklad výkričník, úvodzovky, percento, hviezdička, zátvorka, dolár alebo mriežka, pomlčkami a podčiarkovníkmi,
Transformácia	<p>Predstavuje sled transformačných predpisov reprezentujúcich celý proces spracovania otvorených údajov. Predpisy umožňujú rozdelenie transformácie na spracovacie jednotky (data processing unit DPU) a ich reťazenie. Výsledkom sú spracované alebo obohatené otvorené údaje s príslušnými metadátami a prípadný záznam v katalógu.</p>

Odborný termín	Definícia
URI - Jednotný identifikátor zdroja	<p>Je kompaktný reťazec znakov používaný na identifikáciu alebo pomenovanie zdroja. Hlavný účel tejto identifikácie je umožniť interakciu s prezentáciami zdroja cez sieť, typicky cez World Wide Web, použitím špecifických protokolov.</p> <p>Príklady absolútnej URI: http://somehost/absolute/URI/with/absolute/path/to/resource.txt</p> <p>Príklady relatívnej URI: /relative/URI/with/absolute/path/to/resource.txt</p>
URL - Jednotná adresa zdroja	<p>Je reťazec znakov s definovanou štruktúrou, ktorý slúži k presnej špecifikácii umiestnenia zdroja informácií (v zmysle dokument alebo služba) na Internete. Definuje doménovú adresu servera, umiestnenie zdroja na serveri a protokol, ktorým je možné k zdroju pristupovať:</p> <p><i>protokol://server.doména_druhého_rádu.generická_doména:port/umiestnenie_na_serveri?formulárové_dáta#kotva</i></p> <p>Napríklad: http://cs.wikipedia.org:80/w/wiki.phtml?title=URL&action=edit</p> <p>protokol: http – odpovedajúci protokolu rovnakého mena server (počítač): cs. doména druhého rádu: wikipedia. generická doména (najvyššieho rádu): org(.) port: 80 – keďže pre http je port 80 implicitný, nie je potrebné ho v tomto konkrétnom prípade uvádzať umiestnenie na serveri: /w/wiki.phtml parametre (formulárové dáta): pre formulárové metódy POST a GET, ak sú špecifikované, sú uvedené znakom otáznika. Tu je prvý parameter s menom „title“ a hodnotou „URL“, druhý s menom „action“ a hodnotou „edit“. Parameter a hodnota sa oddeľujú znakom „rovná sa“, dvojica parameter–hodnota sa oddeľujú ampersantom.</p>

Skratky

Skratka	Popis
API	Application Program Interface (aplikačné rozhranie)
CDS	Centrálny dátový sklad verejnej správy
CKS	Centrálny konsolidačný systém
ETL	Extract Load Transform
MSP	Malí a strední podnikatelia
MUK	Modul úradnej komunikácie
OGD	Open Government Data (Otvorené údaje verejnej správy)

Skratka	Popis
OP EVS	Operačný program Efektívna verejná správa
OP II	Operačný program Integrovaná infraštruktúra
PSI	Public Sector Information (Informácie verejného sektora)

1.2 Definícia strategickej priority

Strategická priorita

Národná koncepcia informatizácie verejnej správy (2016) 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
- 6 Riadenie údajov a Big data
- 7 Otvorené údaje
- 8 Vládny cloud
- 9 Komunikačná infraštruktúra
- 10 Informačná a kybernetická bezpečnosť

NKIVS ku každej strategickej prioritě informatizácie verejnej správy vysvetľuje jej cieľ, prístup k riešeniu a tiež rámcový architektonický model.

Tento dokument predstavuje prvý návrh a high-level analýzu priority informatizácie verejnej správy Otvorené údaje **upravenú v kapitole 6.2.7 NKIVS**, pripravenú v rámci Architektonickej kancelárie verejnej správy. Dokument ešte môže byť z pozície architektonickej kancelárie verejnej správy upravovaný a dopĺňaný aj na základe pripomienok a komunikácie s gestorm tejto strategickej priority.

Zodpovednosť za ďalšie detailné riešenie konkrétnej priority, vypracovanie štúdie jej realizovateľnosti a následnú realizáciu formou zabezpečenia implementácie príslušného projektu, resp. projektov, má gestor podľa jemu prislúchajúcej kompetencie.

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 rok 2015. Navrhuje tiež nadstavbu nad nástrojmi pre zverejňovanie otvorených údajov, ktoré sa budujú v projekte eDemokracia v rámci Operačného programu Informatizácia spoločnosti. Väčšinu zo zaujímavých dát je však v súčasnosti 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 neexistujúca transpozícia európskej direktívy na opakované využívanie informácií verejného sektora (Directive 2003/98/EC, známa ako 'PSI Directive') do samostatného komplexného zákona.

V rámci priority sú rozpracované opatrenia v nasledujúcich oblastiach:

- Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov:
 - Jasne sa zdefinuje centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov v centrálnom katalógu data.gov.sk a opakovaným použitím otvorených údajov.

- Zavedie sa dátový model verejnej správy, z ktorého budú vychádzať spoločné a doménové ontológie (spravované centrálnymi ontológmi v Centrálnej kapacite) spĺňajúce aj medzinárodné štandardy otvorených prelinkovaných dát (ďalej Linked dáta) a pravidiel sémantickej a syntaktickej interoperability v európskom priestore.
 - Zavedú sa ďalšie automatizované metódy pre publikovanie otvorených údajov vďaka implementácii Centrálneho dátového skladu.
 - Zavedie sa dohľad nad kvalitou a štandardizáciou metadát a popisných štruktúr, ako aj nad kvalitou samotných dát (Riadenie kvality dát je riešené tiež v strategickej prioritě „Riadenie dát a Big Data“).
 - Do praxe sa uvedie platforma pre štandardizované vytváranie a spravovanie aplikačných rozhraní (API) pre strojový prístup k otvoreným údajom.
 - Začnú sa aplikovať nové postupy pre riadenie kvality datasetov a rizík ich publikovania, predovšetkým pre linked dáta.
 - 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 údajov cez Centrálny dátový sklad budú povinné osoby značne odbremenené;
 - Vyrieši sa problém efektívneho licencovania datasetov otvorených údajov, najmä v oblasti informácií verejného sektora (Public Sector Information). Povolené licencie pre otvorené údaje budú podporovať jednoduché opakované použitie pre komerčné a nekomerčné účely doma aj v zahraničí.
 - Zvýši sa kompetencia verejnej správy pre publikovanie linked dát.
 - Budú dostupné neustále sa zlepšujúce a interaktívne metodiky pre poskytovanie otvorených údajov.
- Poskytovanie otvorených údajov v samospráve:
- Aplikovanie štandardizovaných formátov otvorených údajov a API rozhraní do prostredia samosprávy;
 - Vytvorenie centrálného úložiska datasetov samosprávy v prostredí DECOMu s využitím centrálnych nástrojov pre publikovanie.
 - Vytvorí sa dátový model 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.
- Podpora kreatívnych komunit a inovatívnych podnikov:
- Na základe hodnotenia spätnej väzby sa budú publikovať tie otvorené údaje a API, ktoré dokážu podniky a komunity využiť pre lepšie služby spoločnosti, zvyšovanie transparentnosti a pre rast digitálnej ekonomiky.

Vysvetlenie strategickej priority

Poskytovanie otvorených údajov pre opakované použitie predstavuje samostatný špecifický cieľ programového obdobia 2014 až 2020, ktorý sa zameriava hlavne na rozsah a kvalitu publikovaných dát (až 70 percent zverejňovaných otvorených údajov musí byť v najvyššej kvalite Linked dát), ako aj na intenzitu vyžívania týchto dát podnikateľmi a kreatívnymi komunitami (bude sa priamo monitorovať záujem a počet stiahnutí datasetov, pričom datasety sa musia stiahnuť minimálne 950 000 krát do roku 2020). Poskytovanie otvorených údajov však nie je jednorazový úkon, ale vyžaduje si pravidelnú správu otvorených údajov, ktorá začína dátovým auditom u povinnej osoby a pokračuje prípravou datasetov otvorených údajov, ich zverejnením, pravidelným aktualizovaním a prípadnou archiváciou. Počas celého cyklu musí byť tiež riadená kvalita zdrojových dát, ktorej sa venuje predovšetkým strategická priorita „Riadenie dát a Big Data“. Počas dátového auditu sa zmapujú u povinnej osoby existujúce dátové zdroje, ktoré budú podľa nastaveného plánu postupne zverejňované

Primárnym zámerom je využiť ekonomický potenciál otvorených údajov verejnej správy, ktorých odhadovaná trhovacia hodnota v EÚ je 55.3 miliárd EUR na rok 2016, do roku 2025 súhrnne až 325 miliárd, čo predstavuje približne 25 000 nových pracovných miest v oblasti otvorených údajov¹. Ďalším zámerom je podnietiť záujem občanov a komún o dianie vo verejnej správe a formou verejnej kontroly vytvárať tlak na efektívnejšie fungovanie verejnej správy v súlade s očakávaniami a prioritami verejnosti.

Strategická priorita je v súlade s princípom prístupovania 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 dopad na jej efektivitu;
- Na Slovensku vznikne dostatočne silné odvetvie pracujúce s otvorenými údajmi, pričom vzniknú kompetencie v progresívnych oblastiach informačných technológií, ako sú dátové analýzy;
- Inovatívne využitie otvorených údajov vygeneruje pridanú hodnotu pre používateľov, ktorá sa premietne do ekonomického dopadu.

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

- pre každú organizáciu je dostupný zverejnený zoznam jej dátových zdrojov, z ktorých možno publikovať otvorené údaje, a ich základný popis v katalógu,
- údaje sú zverejnené v datasete (zostave údajov), ktorý je popísaný metadátami 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 spracovanie, čo sa dosiahne dodržaním stanovených technologických štandardov, ktorých základom je používanie otvorených a technologicky neutrálnych riešení – **údaje sú teda zverejnené minimálne na úrovni 3 tzv. Berners-Lee modelu označovaného aj ako Five Star modelu** (viď nižšie),
- 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 (napr. ich spracovaniu, zverejneniu, spájaniu s inými údajmi) nie sú kladené žiadne legálne prekážky. Tento fakt je jednoznačne uvedený v licencii, ktorá musí byť priradená ku každému datasetu.

Úroveň kvality otvorených údajov poskytovaného povinnou osobou môže dosahovať jednu z nasledujúcich šiestich úrovní takzvaného Berners-Lee - päťhviezdičkového („Five Star“) modelu hodnotenia²:

Úroveň 0*

- Dataset nie je poskytovaný v elektronickej forme.

Úroveň 1*

- Dataset je dostupný vo webovom prostredí, avšak nemá štruktúrovaný obsah a nie je poskytovaný v otvorenom formáte.

¹ <http://www.europeandataportal.eu/en/content/creating-value-through-open-data>

² <http://5stardata.info/en/>

Úroveň 2*

- Dataset je dostupný vo webovom prostredí;
- Obsah datasetu je štruktúrovaný tak, že umožňuje automatizované spracovanie, avšak nie je poskytovaný v otvorenom formáte.
- Bežné sú formáty tabuľkových editorov: ODS, XLS(X) a podobne alebo HTML. V prípade textových dokumentov (ako napríklad zmluvy o dodávke externých služieb) sa väčšinou ponecháva formát textového editora (ODT, DOC(X) a podobne).

Úroveň 3* – minimálna úroveň pre publikovanie podľa Výnosu č. 55/2014 Ministerstva financií SR o štandardoch pre informačné systémy VS

- Dataset je dostupný vo webovom prostredí;
- Obsah datasetu je štruktúrovaný tak, že umožňuje automatizované spracovanie;
- Dataset je poskytovaný v otvorenom formáte, nezávislom na konkrétnom proprietárnom softvéri – to znamená, že jeho formát je technologicky neutrálny.
- Odporúčané formáty sú: CSV, XML alebo JSON.

Úroveň 4*

- Dataset je dostupný vo webovom prostredí;
- Obsah datasetu je štruktúrovaný tak, že umožňuje automatizované spracovanie;
- Dataset je poskytovaný v otvorenom formáte, nezávislom na konkrétnom proprietárnom softvéri;
- Na identifikáciu entít datasetu sa používajú referencovateľné identifikátory v podobe URI, čím je možné na tieto entity transparentne odkazovať z ostatných datasetov, je možné tak opakovane použiť len časť dát.
- Dáta sa dajú bezpečne kombinovať s inými dátami. URI predstavujú globálnu schému, takže ak majú dve entity rovnakú URI, tak je to zámerne.
- Rozdelením datasetu na entity možno optimalizovať prístup k nim (napríklad rozloženie záťaže databázy, cachovanie).
- Odporúčané formáty sú: RDF, CSV, XML alebo JSON.
- Spôsoby zaznamenania referencovateľného identifikátora:
 - Pre CSV sa vytvorí nový stĺpec pre URI entity. Tento stĺpec sa umiestni vedľa existujúceho stĺpca pre identifikátory entity.
 - Pre XML, sa využije na zaznamenanie URI rozšírenie HTML a XML dokumentov nazývané RDFa6 (XML atribút resource).
 - Pre JSON sa využije na zaznamenanie URI rozšírenie JSON nazývané JSON-LD7 (konštrukcia @id).

Úroveň 5*

-
- Pre datasety tejto úrovne platí všetko z predchádzajúcej úrovne 4.
- Navyše sú prelinkované referencovateľnými identifikátormi aj iné súvisiace entity v iných datasetoch – ide tak o formát otvorených údajov „linked data“, v ktorom sú URL zaznamenané pomocou štandardu TTL .
- Prelinkovanie datasetov umožňuje ich objavovanie a zvyšovanie ich hodnoty.

Metadáta, ktoré slúžia na popis datasetu, sú v maximálnej kvalite, ak sú správne vyplnené nasledujúce informácie podľa štandardu DCAT – W3C:

- Názov datasetu (či je v súlade s konvenciou pomenovania datasetov vydávanou v budúcnosti kapacitami Centrálnej kapacity pre koordináciu a dohľad nad publikovaním otvorených údajov);
- Jednoznačný identifikátor poskytovateľa datasetu (generovaný automaticky);
- Popis datasetu (či je dostatočne výstižný);

- Dátum prvého zverejnenia datasetu (generovaný automaticky);
- Katalóg, v ktorom je dataset evidovaný (či je správny);
- Semantické značky (tagy) súvisiace s datasetom – ide o kľúčové slová, ktoré vystihujú popis datasetu a ktoré je možné vybrať z centrálného zoznamu;
- Popis datasetu;
- Prístupnosť datasetu (zverejnený/nezverejnený).
- Identifikácia datasetu (generovaná automaticky);
- Informácia o poskytovaní datasetu otvoreným spôsobom použitia alebo o spôsobe vysporiadania právnych aspektov prístupu k údajom v datasete a ich používania;
- Priestorová informácia o územnej relevantnosti dát vo forme geodimenzie, prípadne zadanie priestorovej jednotky z číselníka;
- Jazyk datasetu;
- Periodicita zverejňovania XXX

Časť metaúdajov sa vyplní pri dátovom zdroji. Ak povinná osoba len pridáva dátový zdroj do existujúceho datasetu, pokračuje postupom vytvorenia nového dátového zdroja a vyplní nasledujúce metaúdaje:

- Názov dátového zdroja;
- Popis dátového zdroja;
- Formát dátového zdroja - formát bude doplňovaný pomocou autocomplete funkcie. V prípade, že sa formát nenachádza v databáze, bude nový formát do nej uložený;
- Licencia datasetu (zámerom je v budúcnosti zabezpečiť minimálny počet typov licencií, ideálne len Creative Commons 0);
- Referencia na miesto sprístupnenia dátového zdroja a jeho schémy údajov;
- Informácia o aktuálnosti datasetu podľa § 53 písm. g Výnosu č. 55/2014 Ministerstva financií SR o štandardoch pre informačné systémy VS);
- Informácia o začiatku a konci účinnosti datasetu alebo jeho údajov, najmä ak je účinnosť odlišná od platnosti;
- Prístupnosť dátového zdroja (zverejnený/nezverejnený).

Kontroluje sa tiež, či je ku každému datasetu je dostupná vyššie spomínaná schéma údajov, ktorá popisuje dátovú štruktúru a čiastočne tiež sémantické vzťahy medzi jednotlivými entitami datasetu. Dátová schéma znižuje riziko nesprávnej interpretácie datasetu a uľahčuje prácu vývojárom a ostatným používateľom otvorených údajov. Formát dátovej schémy závisí od formátu datasetu:

- Pre CSV formát sa používa Metadata Vocabulary for Tabular Data alebo JSON;
- Pre XML formát sa vyberie XML schéma ;
- Pre JSON formát sa vyberie JSON schéma .

Každá dátová schéma sa publikuje ako osobitný súbor, ktorý bude prelinkovaný v katalógu so zverejneným datasetom.

Vďaka vybudovaniu centrálného dátového skladu bude možné značnú časť datasetov publikovať centrálné vo vysokej kvalite (úrovne 4 alebo 5 podľa „Five Star“ hodnotenia), bez potreby intervencie povinných osôb. Pre jednotlivé kategórie datasetov otvorených údajov budú presne definované požiadavky na ich aktuálnosť. Úroveň 5 popísaná vyššie zodpovedá linked dátam, čo v praxi znamená posun od webu dokumentov k webu dát (takzvaný sémantický web), v ktorom bude väčšina dát dostupných pre strojové spracovanie, a to s využitím jednoduchšej architektúry a technológií samotného webu. To znamená, že v sémantickom webe (existujú už prvé pokusy na

Slovensku v podobe Slovpedia³⁴, ktorá vznikla podľa vzoru data.gov⁵, data.gov.uk⁶ a dbpedia.org⁷) webové stránky nebudú mať obsah len na čítanie pre ľudí, ale aj prelinkovaný obsah pre spracovanie počítačmi. Napríklad aplikácia bude schopná skombinovať dáta a o významných sektoroch podnikania, ukazovateľoch životného prostredia a kvalite vzdelávacích inštitúcií za účelom pomôcť rodine pri výbere miesta na život.

Pre dôkladné zmapovanie dátových zdrojov, efektívnu prácu s referenčnými údajmi a zjednocovanie formátov a schém údajov ako aj pre celkové riadenie kvality dát bude Dátovej vrstve v novom období venovaná výrazná pozornosť, aj v rámci strategickej priority: Riadenie dát a „Big Data“. Hlavným cieľom strategickej priority Riadenie dát a „Big Data“ bude dátová integrácia medzi jednotlivými systémami verejnej správy a zabezpečenie pravidelnej replikácie kvalitných a konsolidovaných transakčných dát do dátových skladov.

Vzťah k modelu architektonickej vízie

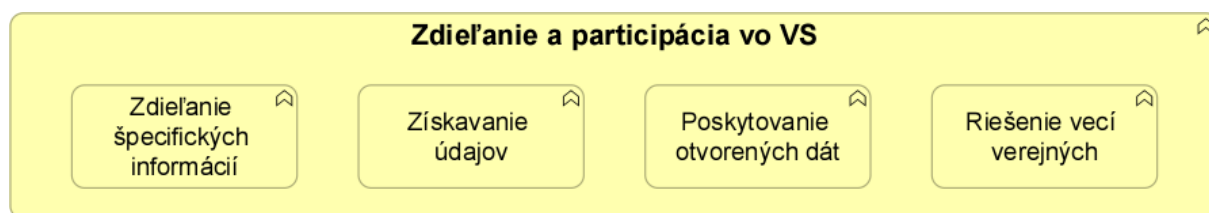
Strategická priorita Open Data je jedným z prvkov Architektonickej vízie verejnej správy 2020. Na úrovni aplikačnej vrstvy strategická priorita priamo súvisí:

- S aplikačnou funkciou Správa otvorených údajov (pričom funkcia ma istý presah aj na komponent Analytické nástroje) – v súčasnosti sa buduje aplikačný komponent IISeDOV (MOD) v rámci projektu eDemokracia.

Nepriamo táto strategická priorita súvisí s Centrálnou správou a poskytovaním referenčných údajov, v rámci ktorej sa štandardizujú dátové prvky a vznikajú prvé predpoklady pre komplexné ontológie verejnej správy.

Na úrovni biznis vrstvy strategická priorita súvisí s:

- S biznis funkciou **Zdieľanie a participácia vo VS**, v rámci ktorej sa nachádzajú biznis funkcie: Zdieľanie špecifických informácií, Získavanie údajov, Poskytovanie otvorených údajov, Riešenie vecí verejných.



Obrázok 1: Biznis architektúra – Riadenie a poskytovanie služieb – Zdieľanie a participácia vo VS

Strategická priorita Open Data by sa mala prakticky realizovať postupne, ako sa budú získavať skúsenosti z projektu eDemokracia, ako sa bude budovať dátová vrstva s ohľadom na kvalitu, centralizáciu a konsolidáciu dát a ako sa budú zvyšovať kompetencie povinných osôb pri zverejňovaní otvorených údajov.

Stavebné bloky

Za základné stavebné bloky, na základe ktorých je štruktúrovaný tento dokument sú považované:

- Správa otvorených údajov.

³ <http://slovpedia.com/>

⁴ <http://slovpedia.com/tripleskop/>

⁵ <http://data.gov>

⁶ <http://data.gov.uk>

⁷ <http://dbpedia.org>

2 Správa otvorených údajov

Rozšírenie koncepcie pre poskytovanie a správu otvorených údajov

Prvým krokom v oblasti otvorených údajov bude zavedenie presného konceptu pre ich poskytovanie a správu vo VS, ktorý bude pozostávať z klasifikácie dát verejnej správy, analýzy ich kvality a potenciálu pre opakované použitie a návrhu možností pre ich poskytovanie tak, aby sa postupovalo od dát s najvyšším potenciálom. Zvážia sa výhody a nevýhody rozširovania komplexného tripletstore (RDF) úložiska pre linked dáta, ktorý poskytuje aplikačné rozhrania pre analýzy a služby prostredníctvom dopytovacieho jazyka SPARQL. Zladia sa iniciatívy budovania Centrálného dátového skladu s tripletstom pre linked dáta a relačným storom pre datasets. Aktualizujú sa štandardy, metodiky a vylepší sa licenčná politika otvorených údajov (o problematike licencií sa nachádza viac v časti Legislatívne požiadavky).

Rozšírenie koncepcie pre využívanie otvorených údajov

Zadefinuje sa spôsob nasadenia využívania údajov a otvorených údajov do procesov a systémov inštitúcií verejnej správy. Určia sa prípady použitia pre sémantický web, tvorený linked dátami.

Podporí sa rozvoj komunity pracujúcej s otvorenými údajmi a zavedie sa efektívny dialóg ohľadom spôsobov poskytovania údajov a ich kvality.

Rozvoj centrálnej platformy pre otvorené údaje

Na budovanej jednotnej centrálnej platforme, na ktorej je k dispozícii katalóg pre všetky otvorené údaje ako aj SPARQL endpoint pre dopytovanie a prehľadávanie linked dát, sa vykonajú všetky patričné zmeny a rozšírenia, aby bola podporená:

- vylepšená koncepcia otvorených údajov,
- open API nad datasetmi otvorených údajov ako aj priamo nad transakčnými systémami pre vybrané prípady použitia,
- a aby sa zabezpečilo prepojenie platformy s Centrálnym dátovým skladom, z ktorého sa budú získavať vybrané datasets automaticky, bez potreby zásahu povinných osôb.

Implementácia nástrojov pre vytváranie otvorených údajov

V rámci inštitúcií VS aj samosprávy sa prevedie analýza systémov na procesnej, aplikačnej a technologickej úrovni, s cieľom identifikovať, ktoré údaje dátových zdrojov bude možné využiť pre automatické generovanie datasetov. Zároveň sa identifikuje kvalita a granularita, v akej sa dátové zdroje nachádzajú, a tiež ich priorita z pohľadu využiteľnosti na základe centrálnej metodiky. Informačné systémy VS budú budované a modifikované tak, aby boli pripravené na otvorené údaje a zverejňovali informácie o svojich dátach do centrálnej platformy, a zároveň poskytovali metadáta. Do procesu vytvárania otvorených údajov budú nasadené transformačné a analytické nástroje, s cieľom zaručiť správne publikovanie dát. Zabezpečí sa dostatočná miera kvality existujúcich dát konsolidáciou dátových zdrojov. Zároveň sa skvalitnia aplikačné rozhrania, ktoré sprístupnia dáta pre automatizované použitie.

Oblasti pre systematické zverejňovanie otvorených údajov

Za základné oblasti, v ktorých sa bude využívať centrálny systematický ukladanie a zverejňovanie otvorených údajov cez Centrálny dátový sklad, je možné považovať (bude však nevyhnutné vykonať dôslednú analýzu dátových zdrojov (dátový audit u povinných osôb) a ich licencovania):

- Geografické dáta podľa štandardu INSPIRE;
- Dáta o životnom prostredí;
- Dáta o vývoji podnikateľského prostredia
- Dáta o vývoji v sociálnej oblasti,
- Dáta o rozpočte a verejných investíciách,

- Dáta o výkonnosti štátnej správy a samosprávy, pričom existujú už riešenia na porovnávanie sa s ostatnými verejnými správami⁸:
 - Dáta o uzatvorených zmluvách;
- Dáta o účtovných závierkach;
- Dáta z legislatívnej oblasti;
- Dáta o kontaktných údajoch úradov, ich GPS súradniciach, ich službách a otváracích hodinách;
- Dáta zo súdnictva.

Typy analýz a aplikácií

Vďaka dostupnosti rôznych dát a ich vzájomnej previazanosti (Linked data) môže odborná verejnosť ale aj inštitúcie VS vykonávať rôzne analýzy nad dostupnými dátami. Príkladom analýz sú:

- Ex-ante vyhodnocovanie opatrení a alternatív a investícií,
- Ex-post vyhodnocovanie prijatých opatrení a investícií,
- Ex-post hodnotenie efektivity verejnej správy,
- Prognózy,
- Simulácie.

Je dôležité, aby sa všetky typy analýz vykonávali nad spoločnou množinou dát, ktorá má vysokú kvalitu, úroveň otvorenosti a konzistentnú ontológiu. Bude tak možné navrhovať a testovať predpoklady jednotlivých modelov a postupne ich zlepšovať.

Pre jednotlivé sektory verejnej správy a oblasti analýz bude možné definovať množinu analytických produktov. Tieto produkty budú pravidelne vytvárané a publikované pre potreby vlády SR, ministerstiev ale i verejnosti. Súhrnný prehľad všetkých produktov by mal byť ideálne dostupný na jednom webovom sídle.

Rovnako môžu byť nad týmito dátami vytvorené rôzne aplikácie. Ako príklad možno uviesť mapu pobočiek IOMO, ku ktorej je potrebné evidovať GPS súradnice každej pobočky a ďalšie údaje, napríklad Otváraciu dobu.

Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov

Navrhnutá Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov dokáže koncentrovať kompetenciu zverejňovania, licencovania a opakovaného používania otvorených údajov vo verejnej správe a v samospráve. Kapacita bude tiež metodicky usmerňovať povinné osoby v oblasti zverejňovania kvalitných datasetov, ktoré bude možné použiť aj na právne účely rovnako ako aj na zaujímavé komerčné služby tretích strán. Kapacita im k tomu bude poskytovať:

- najnovšie technológie (SW moduly transformácií, centrálny katalóg otvorených údajov napojený v budúcnosti aj na centrálny dátový sklad, vďaka čomu sa v budúcnosti zautomatizuje podstatná časť publikovania datasetov, integračnú platformu pre systémy povinných osôb),
- správu ontológií v spolupráci s Dátovou kanceláriou VS, ktorá okrem iného spravuje aj Dátový model VS (bližšie popísaná v strategickej prioritě „Riadenie dát a Big Data“),
- informovanie o najlepších praktikách pri práci s otvorenými údajmi a neustále zlepšované metodiky,
- poskytovanie sumárnych informácií o kvalite zverejňovaných dát vo verejnej správe a o úrovni otvorenosti jednotlivých inštitúcií ako aj o záujme verejnosti o datasety a počte ich stiahnutí – vytvorí sa interaktívny dashboard, ktorý bude sledovať napĺňanie cieľov otvorených údajov a motivovať zapojené inštitúcie k pokroku v tejto oblasti,
- dohľad nad kvalitou a bezpečnosťou zverejňovaných datasetov;

⁸ <http://opengov.com>

- právnu pomoc povinným osobám pri licencovaní;
- školenia,
- ako i špeciálne služby pre povinné osoby, akými sú technologické konzultácie, benchmarking, certifikácia aplikácií a datasetov pre zahraničnú interoperabilitu, testovanie použiteľnosti aplikačných rozhraní, a monitoring ich používania a spokojnosti klientov.

2.1 Ciele realizácie

2.1.1 Architektonické ciele

Táto kapitola obsahuje špecifikáciu architektonických cieľov⁹, ktoré by mali byť dosiahnuté prostredníctvom realizácie strategickej priority.

Zoznam cieľov

Cieľ	Ukazovateľ cieľa	Čas dosiahnutia cieľa
Zvýšenie inovačnej kapacity najmä malých a stredných podnikateľov v digitálnej ekonomike	<ul style="list-style-type: none"> ■ Počet inovatívnych riešení vytvorených podnikateľmi, ktoré využívajú otvorené údaje (1000) ■ Počet zapojených inovatívnych podnikov, ktorí využívajú otvorené údaje a open API (100) ■ Percento zverejňovaných otvorených údajov v kvalite minimálne 4* (? %) 	<ul style="list-style-type: none"> ■ 2020
Zlepšenie kvality, nezávislosti a efektívnosti súdnictva	<ul style="list-style-type: none"> ■ Počet anonymizovaných výsledkov súdnych konaní 	<ul style="list-style-type: none"> ■ 2020
Zlepšenie efektivity verejných investícií	<ul style="list-style-type: none"> ■ Počet datasetov spojených s verejnými investíciami (120) ■ Počet aplikácií pracujúcich s otvorenými údajmi v oblasti verejných investícií a sledovania fungovania verejnej správy (20) ■ Počet prípadov, kedy vďaka občianskej iniciatíve boli zablokované neefektívne investície (200) 	<ul style="list-style-type: none"> ■ 2020

⁹ Vid' dokument - Biznis kontext a motivačný aspekt EA

Naplnenie cieľov OP II

Cieľ OP II	Ukazovateľ cieľa
7.5 Zlepšenie celkovej dostupnosti dát verejnej správy vo forme otvorených údajov	<p>Ukazovateľ výstupov:</p> <ul style="list-style-type: none"> ■ Dodatočný podiel inštitúcií verejnej správy prepojených s centrálnou platformou pre otvorené údaje (cieľový stav sa očakáva 99.9 % inštitúcií verejnej správy, ktoré vytvárajú otvorené údaje) ■ Počet nových datasetov publikovaných vo formáte s vysokým potenciálom na znovupoužitie (úroveň 5 podľa five star modelu) (cieľový stav sa očakáva 70 % datasetov) <p>Ukazovateľ výsledkov:</p> <ul style="list-style-type: none"> ■ Počet stiahnutí otvorených údajov (cieľový stav sa očakáva 950 000 stiahnutí)

Naplnenie cieľov OP EVS

(Tabuľka sa vyplní v prípade, ak je relevantná)

Cieľ OP EVS	Opatrenia a ukazovatele cieľa
Zamedzenie korupcie a podpora transparentnosti	<p>Projektové ukazovatele:</p> <ul style="list-style-type: none"> ■ Počet aplikácií, ktoré využívajú otvorené údaje a slúžia na zvýšenie verejnej kontroly a transparentnosti (60), ■ Pomer rozhodovacích procesov vo verejnej správe, ktoré sú dokladované otvorenými údajmi (80%).
Podpora partnerstva a sociálneho dialógu prostredníctvom budovania kapacít a spolupráce VS so sociálnymi a ekonomickými partnermi a subjektmi MNO	<p>Projektové ukazovatele:</p> <ul style="list-style-type: none"> ■ Veľkosť komunity (počet organizácií mimovládneho sektora), ktoré systematicky pracujú s otvorenými údajmi (50).

2.1.2 Stakeholderi a ich záujmy

V tejto časti jsou identifikovaní stakeholderi, ktorých sa daná strategická priorita (alebo jej časť) týka a ich relevantné záujmy (hnačie prvky9).

Štúdia otvorených údajov (The Open Data Study (2010)) zistila, že na úspešnú implementáciu programu otvorených údajov 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 odporozoval 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 stakeholderov.

Ministerstvá, samospráva a ďalšie štátne organizácie

- Zaviesť princípy štátu, fungujúceho na základe využívania dát v praxi,
- Vykonávať dohľad nad efektívnosťou opatrení,
- Mať kontrolu nad efektívnosťou vynakladaných zdrojov a investícií,
- Získavať konštruktívnu spätnú väzbu od verejnosti a neziskového sektora,
- Legitimizovať pred verejnosťou svoje rozhodnutia a schopnosť zavádzať pozitívne opatrenia do praxe,
- Efektívne plniť povinnosti vyplývajúce zo Zákona o slobodnom prístupe k informáciám.

Úrad vlády SR

- Mať prehľad o stave naplňania strategických priorít,
- Efektívne implementovať Akčný plán Iniciatívy pre otvorené vládnutie na rok 2015 a viesť centrálny katalóg otvorených údajov prostredníctvom agentúry NASES.
- Komunikovať s tretím sektorom.

Ministerstvo hospodárstva SR

- Vytvárať podmienky pre vznik inovatívnych podnikov v digitálnej ekonomike,
- Zvyšovať transparentnosť v procese verejného obstarávania pre konkurencieschopnú ekonomiku.

Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov

- Zabezpečiť vysoký počet užitočných datasetov otvorených údajov vo vysokej kvalite, ktorú budú aktuálne a intenzívne využívané inovatívnymi podnikmi a komunitami,
- Efektívne vykonávať dohľad nad zverejňovaním otvorených údajov a dodržiavaním štandardov a najlepších praktík,
- 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.

Verejnosť

- Zvyšovať transparentnosť vo verejnej správe a v procesoch verejného obstarávania,
- Mať prehľad o predpokladaných dopadoch politik a regulácií,
- Mať prehľad o zdôvodnení investícií vo verejnom záujme,
- Mať prehľad o výkone verejnej správy v jednotlivých sektoroch (až na úroveň jednotlivých inštitúcií) – vrátane ratingov,
- Mať jednoduchý, zrozumiteľný a centrálny prístup ku všetkým otvoreným údajom verejnej správy a samosprávy,
- Efektívne aplikovať práva vyplývajúce zo Zákona o slobodnom prístupe k informáciám.
- Možnosť participatívnej účasti.

Občianske komunity a inovatívne komunity

- Okrem vyššie uvedeného, chcú mať k dispozícii kvalitné údaje v strojovo 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 komunit a k ovplyvňovaniu politik 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 stakeholderov sa považuje vždy za pionierov v oblasti rozvoja programov OGD, hlavne vďaka implementáciám pilotných projektov preukazujúcich úžitok zo zverejňovania otvorených údajov. Odporúča sa tiež orientovať na vytváranie takzvaných verejno-komunitných partnerstiev, ktoré budú experimentovať s novými typmi spolupráce medzi týmito veľmi rozdielne fungujúcimi stakeholdermi¹⁰.

Podnikateľské prostredie

- Mať prístup ku kvalitným analýzám a prognózam o vývoji ekonomiky a o plánovaných opatreniach a obstarávaní verejnej správy,
- Zvyšovať transparentnosť verejnej správy pre zlepšovanie podnikateľských príležitostí,
- Nastavenie regulačných rámcov tak, aby bola znížená administratívna záťaž a vylepšené podnikateľské prostredie v súlade s agendou Lepšie regulácie,
- Možnosť participatívnej účasti na príprave regulácií.
- Možnosť realizovať rizikové inovácie financované štátom,
- Mať dostupné nástroje, API a otvorené údaje a pre poskytovanie inovatívnych komerčných služieb.

2.2 Organizácia

2.2.1 Zodpovednosť

Zodpovedný gestor

- Gestorom strategickej priority Open Data je **Úrad vlády** – vedie Centrálny katalóg otvorených údajov, ďalej „centrálny katalóg“, na Ústrednom portáli verejnej správy. Katalóg využíva medzinárodne uznávané štandardy vyžadované Európskou komisiou. Centrálny katalóg je informačný systém verejnej správy, ktorý obsahuje údaje o informáciách sprístupňovaných na opakované použitie a otvorených údajoch, najmä ich obsah, metaúdaje, podmienky použitia a obmedzenia použitia. Centrálny katalóg je prevádzkovaný agentúrou NASES:
- Za manažment kvality verejnej správy a koordináciu merania výkonnosti jednotlivých sektorov verejnej správy zodpovedá Analyticko-metodická jednotka MV SR,
- Za koordináciu analytických jednotiek zodpovedá Ministerstvo vnútra SR,
- Je tiež nevyhnutné vytvoriť Centrálnu kapacitu pre koordináciu a dohľad nad publikovaním otvorených údajov vo verejnej správe a v samospráve, ktorá bude plniť funkcie uvedené v úvode kapitoly 2 Správa otvorených údajov.

Zapojené organizácie

Jedná sa o všetky organizácie verejnej správy a samosprávy, pretože na základe existujúceho právneho stavu majú byť sprístupnené povinnými osobami všetky údaje, okrem tých, pre ktoré to špecificky určitý zákon vylučuje.

2.2.2 Organizačné zmeny

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

- Zriadenie Centrálny kapacity pre koordináciu a dohľad nad publikovaním otvorených údajov vo verejnej správe a samospráve, ktorá bude mať na starosti audit nad rozsahom zverejňovaných informácií, kvalitou a dodržiavaním štandardov, správu ontológií a centrálnu koordináciu zverejňovania otvorených údajov v správnom formáte. Auditor má tiež právo dohliadať na vybavovanie požiadaviek verejnosti na zverejňovanie otvorených údajov. Vhodné by bolo tiež disponovať právnyimi odborníkmi v oblasti dát a licencovania ako aj

¹⁰ Acknowledgement to the European Public Sector Information (PSI) Platform

centrálnou kapacitou dátového kurátora pre riešenie a koordinovanie nadrezortných úloh v oblasti transformácie a publikovania kvalitných otvorených dát. Moderátor by sa mal starať o kvalitu diskusie na fóre data.gov.sk, odpovedať na otázky a požiadavky na zvýšenie kvality zverejňovaných údajov a prípadne aj odstraňovať nevhodné príspevky.

- Vytvorenie alebo posilnenie analytickej jednotky v oblasti otvorených údajov na každom ministerstve a ostatných dôležitých inštitúciách verejnej správy a samosprávy. Analytické jednotky budú realizovať analýzy a produkovať analytické produkty, ktoré budú dohodnuté. Budú tiež disponovať kapacitou dátového kurátora.

2.3 Stratégia a riešenie

2.3.1 Popis súčasného stavu

Na základe štúdie¹¹ európskej komisie sa Slovensko v oblasti vyspelosti poskytovania otvorených údajov nachádza v strednej skupine takzvaných nasledovateľov „Followers“, spolu s krajinami ako Poľsko, Holandsko, Slovinsko a ďalšími. Zámerom je posunúť sa do skupiny, ktorá udáva trendy.

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>.

Použité komponenty	Skratka	Funkcionalita	Licencia
CKAN	CKAN	Katalóg otvorených údajov	Open-source AGPL 3.0
Pool Party	PPT	Ontologický subsystém, údržba riadených slovníkov, extrakcia konceptov	Proprietárna
Virtuoso Universal DB	Virtuoso	Databáza pre triplety, tzv. "triplestore" a spracovanie dotazov v SPARQL	Proprietárna
Unified Views	UV	Transformácie dát	Open-source GPL
XWiki		Crowdsourcing, komunikácia metodík	Open-source LGPL
Open data node	ODN	Podporná aplikácia pre povinné osoby	Open-source AGPL, LGPL

Použité komponenty možno mapovať na nasledujúce komponenty MOD.

Použitý komponent	Komponent MOD
CKAN	čistočne Správa katalógov, Repozitár datasetov, čiastočne Prezentačná aplikácia
PPT	Ontologický editor, Správa ontológií a identifikátorov

¹¹ <http://www.europeandataportal.eu/en/content/open-data-maturity-europe>

Virtuoso	RDF úložisko, čiastočne "SPARQL" Query engine
UV	Transformačný komponent, čiastočne Správa transformácií
ODN	Podporná aplikácia pre povinné osoby

V nasledujúcej časti sú popísané kľúčové komponenty MOD, spomenuté vyššie.

Ontologický editor pre linked dáta

Ontologický editor umožňuje správu dátových slovníkov (ontológií).

Definície:

- **Koncept** – slovo alebo slovné spojenie s definovaným významom. Jednoznačne identifikované pomocou URI.
Príklady: Osoba, Povinná Osoba, Štát, IT Projekt, Bratislava (mesto), Prešov (mesto), Prešov (okres)
- **Ontológia** – množina konceptov (tripletovej), ktorá opisuje vybranú doménu. Medzi konceptami sú definované rôznorodé vzťahy. Každý koncept môže mať rôzne vlastnosti (atribúty).

Príklady: Geografická ontológia (štáty, regióny, mestá, okresy, ...), Friend of a friend (rozšírená ontológia používaná pre potrebu popisu osôb a vzťahov)

Ontológia môže povoľovať jednoduché hierarchické vzťahy (nadriadený koncept – podriadený koncept), všeobecné symetrické vzťahy (koncept A "súvisí s" s konceptom B), prípadne iné typy vzťahov ("je podobný ako", "je taký istý ako", "je opak") alebo úplne arbitrárne vzťahy ("je veliteľom misie Apollo").

Ontologický editor spolu s publikačným nástrojom pre XML schémy teda zabezpečuje nasledujúce prísluhy Modulu Open Data:

- register ontológií na báze linked data
- register konceptov definovaných v linked data ontológiách
- register URI identifikátorov konceptov
- register ontológií na báze XML schém
- register dátových prvkov definovaných v XML schémach
- register URI identifikátorov dátových prvkov (tak ako budú manuálne zadané v XML schémach).

Okrem toho, pomocou štandardných nástrojov (transformačný modul, priamy import do katalógu) môže každá povinná osoba.

Transformačný modul

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

Register referencovateľných identifikátorov

Každý dátový slovník je zároveň registrom referencovateľných identifikátorov (každý koncept je totiž identifikovaný svojím URI).

Generovanie URI

Bude možné použiť jednu z nasledovných konvencií na generovanie URI pre koncepty v rámci ontológie.

1. URI: <http://data.gov.sk/.../menovka-konceptu>
2. URI: <http://data.gov.sk/.../<Increment>>

3. URI: `http://data.gov.sk/.../<UUID>`

Príklad: Ontológia "Osoby a Vzťahy" s dátovým prvkom "Povinná osoba":

1. URI: `http://data.gov.sk/def/osoby-a-vztahy/povinna-osoba`

2. URI: `http://data.gov.sk/def/osoby-a-vztahy/123`

3. URI: `http://data.gov.sk/def/osoby-a-vztahy/123e4567-e89b-12d3-a456-426655440000`

Predpokladá sa, že bude preferované používať prvý spôsob.

Podľa potreby bude možné pre rôzne slovníky zvoliť rôzne konvencie generovania URI.

Pomocou uvedených funkcionalít bude napr. možné vytvoriť ontológiu/ontológie v zmysle Výnosu a spravovať ich v ontologickom editore.

Podporná aplikácia

Bude poskytovať 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 Modulu otvorených pdajov.

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

Riziká súčasného stavu

Zoznam rizík, ktoré vyplývajú zo skutočnosti, ak by sa súčasný stav nezmenil:

- Verejná správa nevyužije potenciál, ktorý prinášajú otvorené údaje a nové technológie spracovania údajov – z dát nevznikne vzácný zdroj a Slovenska republika nevytvorí jedinečné inovačné prostredie v oblasti dátových analýz,
- V centrálnom katalógu nebude dostatok zaujímavých datasetov, a to z nasledujúcich dôvodov:
 - Zverejňovanie otvorených údajov bude príliš komplikované a povinné osoby si nevytvoria dostatočné kapacity a kompetencie na zverejňovanie otvorených údajov;
 - Budú pretrvávajúť problémy s licencovaním dátových zdrojov pre opakované použitie na komerčné aj nekomerčné účely;
 - Budú pretrvávajúť obavy, aby nedošlo k úniku osobných údajov, obchodného tajomstva alebo utajovaných skutočností.
- Datasetsy nebudú poskytované v uspokojivej kvalite;
- Nepodarí sa dostatočne zapojiť občianske tech komunity,
- Nevzniknú inovatívne firmy, ktoré budú schopné vyťažiť z potenciálu otvorených údajov,
- Nevznikne dostatočne motivovaná a funkčná komunita, ktorá prostredníctvom verejnej kontroly dokáže vytvoriť adekvátny tlak na úspešné presadenie zmien fungovania verejnej správy

2.3.2 Najlepšie skúsenosti

Zoznam najlepších skúseností (best-practice) zo zahraničia, ktoré sú relevantné pre strategickú prioritu, respektíve sa z nich vychádza:

- Open data portály v EU, UK a USA,
- Medzinárodný projekt COMSODE,

- European Public Sector Information (PSI) Platform¹²¹³;
- Štandardy W3C¹⁴;
- Government Linked Data (GLD) Working Group¹⁵;
- Európske slovníky pre sémantickú interoperabilitu¹⁶;
- DBpedia, ktorá umožňuje sofistikované dopyty nad Wikipédiou a prelinkovávať rôzne datasety na Webe s informáciami na Wikipédii¹⁷ vo forme sémantického webu;
- Sociálna sieť Flickr, ktorá využíva na prelinkovanie svojich fotografií tiež sémantický web (linked data)¹⁸

Linked dáta

RDF dokumenty a schémy možno popisovať na troch rôznych úrovniach abstrakcie:

- Na syntaktickej úrovni sa jedná o XML dokumenty.
- Na štruktúrálnej úrovni ide o takzvaný triplet ako znalosť vo forme „subjekt predikát objekt“:
- Na sémantickej úrovni vytvárajú RDF dokumenty jeden alebo viacero grafov s čiastočne preddefinovanou sémantikou.

Na základe zahraničnej literatúry je najefektívnejšie prehľadávať RDF dokumenty na sémantickej úrovni.

2.3.3 Strategický prístup k riešeniu

Prístup k systematickému riešeniu Open Data je možné zhrnúť ako:

- Implementovať lessons learned z projektu eDemokracia,
- Ako súčasť reformy verejnej správy, vytvorí sa centrálna kapacita schopná centralizovane a systematicky riadiť, dohliadať a koordinovať zverejňovanie a používanie otvorených údajov vo verejnej správe a v samospráve:
 - Správa centrálného katalógu,
 - 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 údaje,
 - Koordinácia rôznych iniciatív, napríklad INSPIRE,
 - Koordinácia propagácie témy Open Data, organizácia workshopov a hackathonov.
- Na jednotlivých ministerstvách budú zriadené analytické jednotky zodpovedné za konkrétne oblasti verejnej správy (politiky, regulácie, výkonnosť, investície) a so schopnosťami zverejňovať otvorené údaje – teda aj s kompetenciami dátového kurátora (obdobné kapacity budú vybudované aj pre samosprávu).
- Podporované budú segmentové iniciatívy s cieľom komplexného zabezpečenia publikovania v segmente a v samospráve:

¹² <http://www.europeandataportal.eu>

¹³ <https://open-data.europa.eu/en/data/>

¹⁴ <https://www.w3.org/standards/semanticweb/>

¹⁵ https://www.w3.org/2011/gld/wiki/Main_Page

¹⁶ https://joinup.ec.europa.eu/community/semic/og_page/core-vocabularies

¹⁷ <http://wiki.dbpedia.org>

¹⁸ <https://www.flickr.com/photos/ukodi>

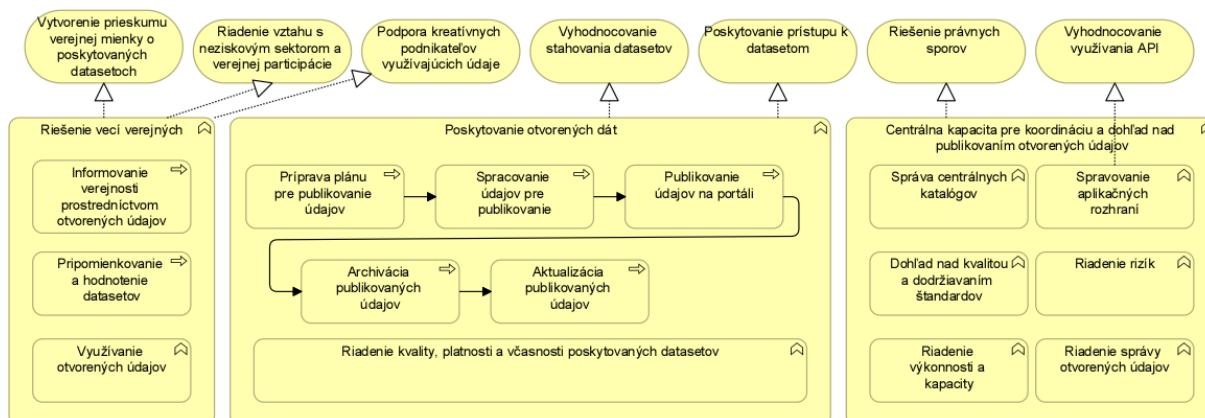
- Identifikácia datasetov, ktoré sa budú zverejňovať v rámci segmentu,
 - Nastavenie procesov zverejňovania dát,
 - Zabezpečenie, že datasety sú pravidelne publikované cez portál „data.gov.sk“.
- Implementuje sa informačný systém Centrálny dátový sklad, ktorý bude zhŕňať dáta z verejnej správy a umožní centrálné zverejňovanie otvorených údajov verejnej správy. CDS bude tvorený súborom nástrojov pre správu dátových zdrojov, umiestnený v eGovernment cloude. Pre potreby samosprávy bude obdobne aktualizovaný DCOM.
 - Jednotlivé informačné systémy verejnej správy budú pripájané k Centrálnemu dátovému skladu cez dátovú integráciu a ETL procesy.

Princípy

Zoznam architektonických princípov⁹, na ktorých je riešenie postavené:

- Údaje sú aktívne,
- Údaje sú zdieľané,
- Údaje sú dostupné,
- Údaje majú vlastníka a správcu,
- Spoločné údajové ontológie a definícia konceptov.

2.3.4 Biznis vrstva



Obrázok 2: Pohľad na biznis vrstvu v oblasti Open Data

Požiadavky

Zoznam požiadaviek:

- Publikovanie dát bude efektívne, jednoduché a transparentné, a čo v najvyššej možnej miere automatizované,
- Otvorené údaje budú publikované vo vysokej kvalite spĺňajúcej úroveň 4* alebo 5*,
- Riešenie umožňuje jednoducho prideliť licenciu dátovému zdroju,
- Bude efektívne monitorovaný dopad opatrení otvorených údajov a v prípade nedosahovania cieľov budú dostupné nápravné opatrenia,
- Existuje centrálny dátový model verejnej správy vrátane ontológií pre otvorené údaje,
- Riešenie obsahuje centrálné automatické transformácie dát,
- Datasety je možné referencovať, zdieľať, vyhľadávať, komentovať a sú dostupné cez jednotný prístup,
- Budú dostupné neustále sa zlepšujúce a interaktívne metodiky pre publikovanie otvorených údajov,

- Podnikatelia, verejnosť a kreatívne komunity sa budú môcť jednoducho zapojiť do procesu otvoreného vládnutia.

Služby

Zoznam služieb:

- Poskytovanie prístupu k datasetom
- Vyhodnocovanie využívania API
- Vyhodnocovanie sťahovania datasetov
- Vytvorenie prieskumu verejnej mienky o poskytovaných datasetoch
- Riadenie vzťahu s neziskovým sektorom a verejnej participácie
- Podpora kreatívnych podnikateľov využívajúcich údaje
- Riešenie právnych sporov

Procesy a funkcie

Základné funkcie:

- Poskytovanie otvorených údajov

Súvisiace funkcie, ktoré používajú služby:

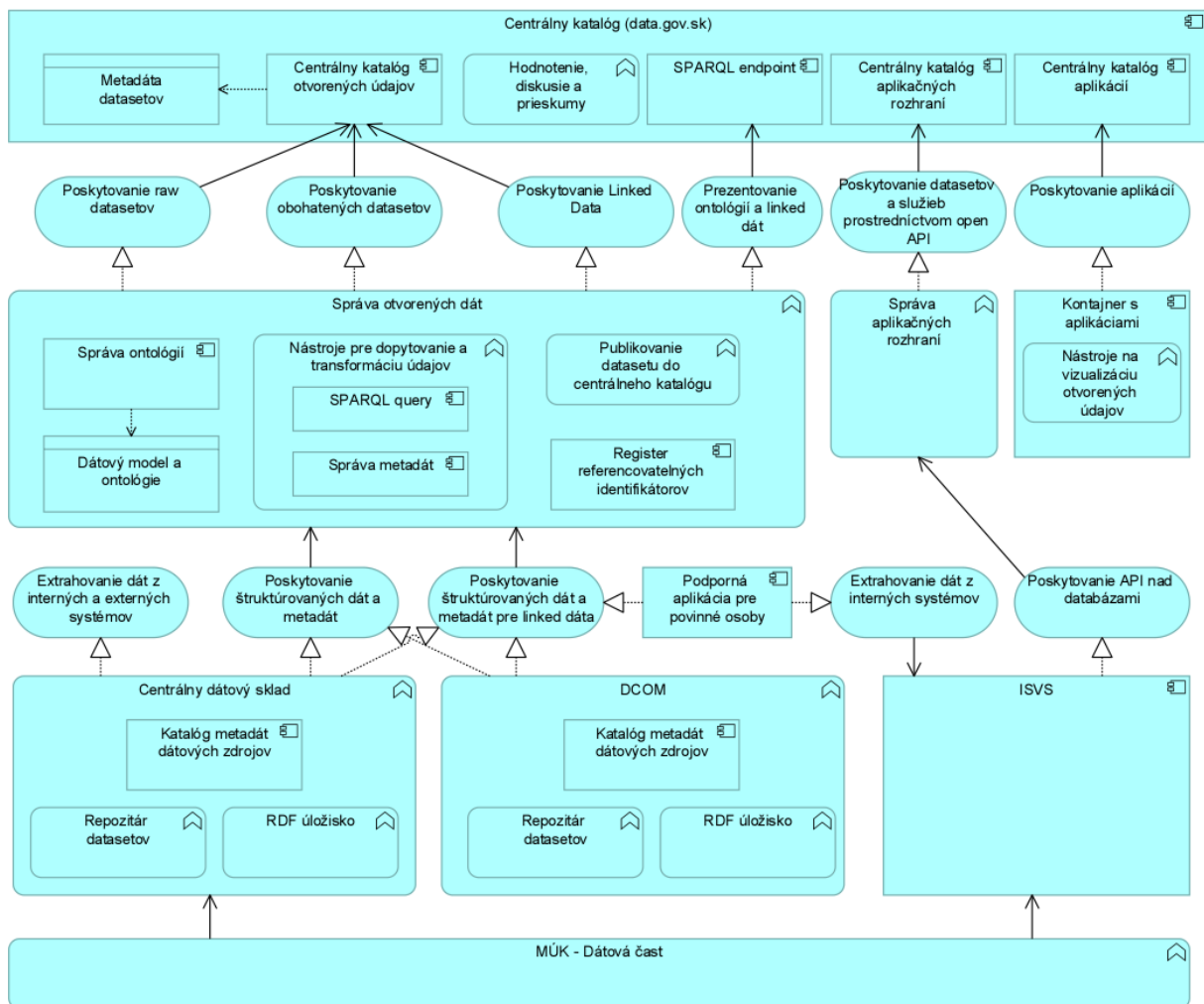
- Riešenie vecí verejných:
 - Využívanie otvorených údajov.

Základné procesy:

- Príprava plánu pre publikovanie dát: Je nevyhnutné vykonávať prieskumy v rámci kreatívnych komunit a inovatívnych podnikov, o ktoré dáta by mali najväčší záujem, a podľa toho prioritizovať v pláne publikovanie. Odporúča sa tiež sledovať Globálny index otvorených údajov <http://index.okfn.org/place/>, ktorý formou crowd-sourcingu vyhodnocuje rozsah, kvalitu a užitočnosť datasetov zverejňovaných naprieč verejnými správami.
- Spracovanie dát pre publikovanie:
 - Príprava prístupu k dátovým zdrojom, extrakcia dát,
 - Transformácia dát na otvorené datasety, prípadne linked dáta,
 - Popis datasetov a dátových zdrojov metadátami,
 - Pridelenie licencie,
- Publikovanie dát na portáli:
 - Zverejnenie datasetov a ich sprístupnenie cez API a centrálny katalóg,
- Aktualizácia publikovaných dát;
- Archivácia publikovaných dát;
- Riadenie kvality, platnosti a včasnosti poskytovaných datasetov
- Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov:
 - Správa centrálnych katalógov:
 - Otvorených údajov;
 - API;
 - Aplikácií nad otvorenými údajmi.
 - Návrh a spravovanie aplikačných rozhraní
 - Dohľad nad kvalitou a dodržiavaním štandardov: pre validovanie schém údajov samosprávy možno použiť <http://validator.opendata.esd.org.uk/contracts>. 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,
- Riadenie výkonnosti a kapacity,
- Riadenie správy otvorených údajov.

2.3.5 Aplikačná vrstva



Obrázok 3: Pohľad na aplikačnú vrstvu pre Analýzu dát

Požiadavky

Zoznam požiadaviek

- 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,

Linked dáta:

- 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 uľahčí prelinkovanie akéhokoľvek súvisiaceho objektu alebo zdroja.

Aplikačné služby

Zoznam služieb

- Poskytovanie raw datasetov
- Poskytovanie obohatených datasetov
- Poskytovanie Linked Data
- Extrahovanie dát z interných a externých systémov
- Spracovanie dát na otvorené údaje (konverzie a transformácie, čistenie, anonymizácia, obohacovanie, prelinkovanie a mnohé ďalšie)
- Ukladanie publikovaných datasetov a metadát
- Poskytovanie datasetov a služieb prostredníctvom open API
- Vizualizovanie dát

Aplikačné komponenty

Zoznam základných komponentov:

- Centrálny dátový sklad a DCOM, ktoré po extrakcii a transformácii dát z distribuovaných úložísk poskytujú:
 - Katalóg metadát dátových zdrojov,
 - Repozitár datasetov,
 - RDF úložisko,
- Modul otvorených údajov z projektu eDemokracia;
- Centrálny katalóg, obsahujúci štandardizované metadáta o položkách na data.gov.sk:
 - Otvorených údajov;
 - API (nový komponent, ako inšpirácia slúži <https://www.data.gov/developers/apis>);
 - Aplikácií nad otvorenými údajmi,
 - Funkcionalita pre hodnotenie datasetov, diskusie a prieskumy,
 - SPARQL endpoint.
- Správa otvorených dát a Podporná aplikácia pre povinné osoby, obsahujúce hlavne:
 - Nástroje na dopytovanie a transformáciu dát (ETL nástroje),
 - Publikovanie datasetu do centrálného katalógu,

- Register referencovateľných identifikátorov,
- Správa ontológií,
- Správa aplikačných rozhraní pre tvorbu, publikovanie a spravovanie open API,
- Kontajner aplikácii obsahujúci aplikácie vytvorené kreatívnymi komunitami a podnikateľmi:
 - Nástroje na vizualizáciu otvorených údajov

2.3.6 Technologická vrstva

Informačný systém bude umiestnený v eGovernment cloude, v ktorom budú vytvorené platformové služby na ukladanie dát (Big Data), extrakciu dát z dátových zdrojov a ich prvotnú transformáciu a získavanie pre analytické účely a publikovanie otvorených údajov. Časť riešenia pre samosprávu bude umiestnená v DCOM.

2.4 Problémy a riziká

Dôsledky

Zoznam dôsledkov, ktoré vyplývajú z realizácie strategickej priority pre jednotlivé subjekty:

- Informačné systémy verejnej správy je potrebné integrovať s CDS (cez ETL nástroje),
- Zmení sa proces pri poskytovaní informácií na základe Zákona č. 211,
- Vo verejnej správe bude potrebné zamestnať väčší počet dátových, technologických a právnych expertov,
- Bude potrebné zvýšiť kapacitu a efektivitu komunikácie s verejnosťou, školami, univerzitami, podnikateľmi a tretím sektorom.

Problémy

Zoznam problémov, ktoré bude potrebné vyriešiť počas realizácie strategickej priority:

- Ako navrhnuť prioritizáciu vzniku dátových skladov, pripájanie informačných systémov do dátového skladu,
- Ako zabezpečiť ochranu osobných údajov a údajov týkajúcich sa národnej bezpečnosti,
- Ako získať plnú podporu zo strany autorov informačných systémov pre opakované používanie PSI pre komerčné aj nekomerčné účely,
- Ako vytvoriť efektívnu platformu s uspokojivou SLA pre zapojenie kreatívnych komunit a inovatívnych podnikateľov.
- Ako získať dostatočné kapacity, schopné zverejňovať a spravovať otvorené údaje a riadiť inovačné snahy.

Riziká

Zoznam rizík, ktoré je dôležité mitigovať pri realizácii strategickej priority:

- Návrh dátového modelu bude príliš komplexný a nepoužiteľný v praxi,
- Nepodari sa získať otvorenú licenciu pre kľúčové datasety zaujímavé pre inovatívne podniky ako meteorologické a geografické informácie – v tomto smere je dôležité vytvoriť a aplikovať Open Data Ready Metodiku, popísanú v časti Legislatívne požiadavky,
- O datasety a open API nebude záujem (verejnosti, občianskych komunit a podnikateľov),
- Kvalita dát bude nízka.

2.5 Legislatívne požiadavky

Legislatívne požiadavky

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

Legislatívny predpis	Navrhované opatrenie
Metodika Open Data Ready	Ide o súbor kritérií podstatných pre možnosť publikácie otvorených údajov 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.
Zákon o otvorených údajoch	Nový komplexný zákon, ktorý prehľadne transponuje európsku direktívu (Directive 2003/98/EC, známa ako 'PSI Directive').
Aktualizácia Výnosu 55/2014 o štandardoch ISVS	Upravuje výnos na základe trendov rozvoja otvorených údajov, predovšetkým v oblasti linked dáta a ontológií.
Nové štandardy	Súvisiace aj s kvalitou dát ako takých (formáty), štandardizácia metadát a ontológií
Aktualizácia Autorského zákona	Bude obsahovať presné vymedzenie používania licencií pre PSI, pričom primárne sa odporúča používať CC0 a následne CC-BY.
Organizačné opatrenie	Zriadenie Dátovej kancelárie verejnej správy
Organizačné opatrenie	Zriadenie Centrálnej kapacity pre koordináciu a dohľad nad publikovaním otvorených údajov
Metodické opatrenie	Zavedenie štandardov pre analytickú a dátovú vrstvu vo verejnej správe

Požiadavky na Operačný program Efektívna verejná správa

- Návrh a nastavenie nových procesov pre manažment dát vo verejnej správe a samospráve pre potreby otvorených údajov.
- Podpora zriadenia a vybudovanie kapacít Dátovej kancelárie verejnej správy.
- Podpora zriadenia a vybudovanie kapacít Centrálnej kapacity pre koordináciu publikovania otvorených údajov.
- Vybudovanie analytických jednotiek v relevantných inštitúciách
- Zefektívnenie komunikácie s občianskymi kreatívnymi komunitami a inovatívnymi podnikateľmi.
- Návrh opatrení pre podporu občianskych komunít a inovatívnych podnikateľov.

2.6 Plánovanie a migrácia

Identifikácia projektov, ktoré je vhodné realizovať:

- Centrálny dátový sklad s možnosťou centrálného publikovania otvorených údajov (projekt je definovaný v rámci strategickej priority „Riadenie dát a Big Data“);
- Otvorené údaje pre samosprávu;
- Open API;
- Dopytové projekty pre inovatívne podniky a podnikateľov,

- Centrálna kapacita pre koordináciu publikovania otvorených údajov (projekt z OP EVS).

Navyše je v každom projekte, ktorý sa venuje modernizácii agendových systémov, nutné uviesť spôsob integrácie na platformu otvorených údajov.

2.6.1 Centrálny dátový sklad

Popis a cieľ

Cieľom projektu je vybudovanie technologickej infraštruktúry založenej na technológiách Big Data (ako HADOOP), integrácia zdrojov dát, naplnenie relevantnými dátami, návrh analytických a dátových modelov a zavedenie používania analytických nástrojov do reálnej praxe.

Aktivity

- Analýza a návrh riešenia,
- Implementácia CDS,
- Nasadenie Query a analytických nástrojov,
- Implementácia dátového modelu verejnej správy,
- Implementácia ontológií pre otvorené údaje,
- Integrácia zdrojov dát,
- Integrácia dátových skladov MOD z projektu eDemokracia
- Testovanie.

Navrhovaný realizátor

- Štatistický úrad SR

Navrhovaný partneri

- Ministerstvo financií SR
- Ministerstvo vnútra SR

Odhadovaná dĺžka realizácie projektu (projektov)

- 3 roky (viacero fáz, výstupy postupne)

Závislosti na iných projektoch a prioritách (OPII a OPIS)

- Centrálny konsolidačný systém
- Modul úradnej komunikácie – dátová časť

Závislosti na iných projektoch a prioritách (OP EVS)

- Vybudovanie dátovej kancelárie verejnej správy

2.6.2 Otvorené údaje pre samosprávu

Popis a cieľ

Väčšina samospráv je na začiatku procesu publikovania otvorených údajov. V súčasnosti sa publikujú prevažne len povinné údaje. Tlak občanov a potreby participácie vytvárajú prostredie, ktoré stimuluje proaktívne publikovanie otvorených údajov v každej samospráve.

Cieľom projektu je priniesť dáta samospráv občanovi nielen ako OpenData, ale aj ako človeku zrozumiteľnú interpretácia konkrétnych faktov v konkrétnych situáciách. Proaktívne zverejňované a interpretované dáta z doménovej oblasti samospráv je možné zapojiť do procesov občianskej participácie. Tieto témy plánuje projekt rozvíjať spoločne s ďalšími aktérmi ako ZMOS, DEUS, USVROS v spolupráci s tretím sektorom a s medzinárodnými aktivitami v oblasti výskumu.

V rámci projektu sa nasadia nástroje a postupy pre zverejňovanie otvorených údajov samosprávy. Väčšina inštitúcií verejnej správy patrí do samosprávy, a preto pre naplňovanie cieľov v oblasti otvorených údajov je nutné sa samospráve systematicky venovať.

Aktivity

- Návrh organizačného zabezpečenia Centrálnych koordinačných aktivít zverejňovania otvorených údajov v samospráve
- Analýza a návrh riešenia,
- Štandardizácia formátov a API rozhraní,
- Štandardizácia samotnej štruktúry prioritných datasetov (pre budúce spracovanie je nevyhnutné aby boli datasety ako výstupy jednotlivých agend samosprávy publikované v obdobných štruktúrach. Inak bude nemožné efektívne spracovávať obsah datasetov z viacerých samospráv
- Koordinovaná tvorba odporúčaní a metodík pre fungovanie samospráv rôznych veľkostí tak, aby proces publikovania otvorených údajov prebiehal efektívne a výsledné dáta boli prínosom pre spoločnosť
- Spolupráca a koordinácia projektov vrátane tvorby spoločných metodík.
- Zjednotenie pravidiel využívania referencovateľných identifikátorov a ontológií v doméne samospráv tak, aby bolo možné automatizovane publikovať dáta aj vo formátoch linked data.
- Implementácia,
- Vytvorenie ontológie pre samosprávu,
- Integrácia zdrojov dát,
- Vytvorenie platformy na publikovanie otvorených údajov v Dátovom centre obcí a miest,
- Testovanie.

Prínosy pre občanov

- Zvýšená možnosť participácie občana
- Vyššia transparentnosť
- Jednotný prístupový bod k dátam samosprávy – dátový katalóg samospráv automaticky replikovaný aj na úroveň data.gov.sk
- Jednotný formát publikovania prioritných datasetov – dáta jednotlivých agend publikované v rovnakých štruktúrach
- Jednotné aplikácie prezentujúce dáta samospráv – občan zvyknutý na prezentáciu dát jednej samosprávy dokáže pomocou rovnakej aplikácie pracovať s dátami inej samosprávy
- Zníženie nákladov na vyhľadanie dát
- Zníženie neprehľadnosti v porovnaní so samostatne nekoordinovane publikovanými dátami

Prínosy pre biznis

- Jednotný prístupový bod k dátam
- Jednotná štruktúra publikovaných datasetov
- Zjednodušenie potenciálneho spájania datasetov
- Využitie doménového predspracovania datasetov
- Výrazne zníženie nákladov získania dát z viacerých samospráv

Prínosy pre samosprávu

- Možnosť využitia jednotných metodík a nástrojov špecializovaných pre potreby samospráv
- Zníženie nákladov na proces publikovania
- Zjednodušenie infraštruktúry

- Možnosť využitia jednotných aplikácií nad dátami samospráv
- Možnosť spolupráce a koordinácie pri publikovaní, navrhovaní aplikácií a tvorby metódik
- Možnosť využívať interný katalóg pre účely harmonizácie interných dát ich transformácia, licencovanie a nastavenie pravidiel publikovania
- Služba externého katalógu embedovateľná do web stránky samosprávy
- Automatizácia publikovania dát
- API rozhrania automaticky publikujúce dáta pre potreby občanov, biznisu a verejnej správy
- Možnosť bezpečne publikovať aj dáta ktoré nie sú Open – pre G2G potreby
- Zvýšenie legitimity činnosti samosprávy

Prínosy pre verejnú správu

- Jednotné prístupové miesto k dátam samosprávy
- Jednotné formáty datasetov, ktoré sa dajú ľahko spracovávať (v porovnaní so stavom, že si každá samospráva publikuje ako chce je to asi hlavná pridaná hodnota projektu)
- Možnosť využívať špecializované API rozhrania
- Zjednodušenie komunikácie so samosprávami, ktoré budú proaktívne publikovať svoje dáta
- G2G komunikácia

Navrhovaný realizátor

- DataCentrum elektronizácie územnej samosprávy Slovenska

Navrhovaný partneri

- Združenie miest a obcí Slovenska
- Ministerstvo financií SR

Odhadovaná dĺžka realizácie projektu (projektov)

- 4 roky

Závislosti na iných projektoch a prioritách (OPII a OPIS)

- Centrálny dátový sklad
- eDemokracia

Závislosti na iných projektoch a prioritách (OP EVS)

- KOMPETENČNÉ CENTRUM ZMOS
- DEUS - koordinačné centrum informatizácie miestnej územnej samosprávy
- USVROS - Podpora partnerstva a dialógu medzi verejnou správou, občanmi a mimovládnyimi neziskovými organizáciami na národnej, regionálnej a lokálnej úrovni v oblasti participatívnej tvorby verejných politík

2.6.3 Open API

Popis a cieľ

Cieľom projektu bude vybudovať platformu pre tvorbu, publikovanie a spravovanie otvorených API, ktoré budú dostupné cez centrálny katalóg podnikateľom a kreatívnym komunitám. Tieto API budú poskytované nad repozitármi datasetov, RDF úložiskom ako aj nad vybranými informačnými systémami verejnej správy. Pomocou týchto API bude možné vybudovať ľubovoľné komerčné aj nekomerčné aplikáciu kombináciou dát a služieb („mashup“).

Aktivity

- Analýza a návrh riešenia,

- Implementácia platformy pre open API,
- Vytváranie a publikovanie open API nad otvorenými údajmi a systémami verejnej správy,
- Zavedenie dohľadu nad bezpečnosťou a využívaním open API,
- Testovanie,
- Propagácia možností Open API a spolupráca s univerzitami,

Navrhovaný realizátor

- Ministerstvo financií SR (ako realizátor MUK)

Navrhovaný partneri

- -.

Odhadovaná dĺžka realizácie projektu (projektov)

- 2 roky (viacero fáz, výstupy postupne)

Závislosti na iných projektoch a prioritách (OPII a OPIS)

- Modernizácia agendových informačných systémov,
- eDemokracia,
- Otvorené údaje pre samosprávu,
- Centrálny dátový sklad.

Závislosti na iných projektoch a prioritách (OP EVS)

- Momentálne žiadne

2.6.4 Dopytový projekt pre inovatívne komunity a podnikateľov

Popis a cieľ

Cieľom projektu je nachádzať inovatívne riešenia využívajúce otvorené údaje, ktoré by boli prospešné pre verejnú správu a pre občanov, ako aj zvýšiť povedomie o tom, na čo otvorené údaje môžu slúžiť. Dopytový projekt môže byť realizovaný napríklad súťažou nápadov podľa vzoru challenge.gov.

Aktivity

- Návrh konkrétneho problému, témy alebo výzvy, ktorá by sa mala riešiť,
- Dodávka algoritmu, aplikácie nad otvorenými dátami alebo vizualizácie otvorených údajov,
- Workshopy a hackathony, ktorých účelom bude propagovať a vzdelávať verejnosť v oblasti otvorených údajov a občianskej participácie,
- Spolupráca s univerzitami a školami.

Navrhovaný realizátor

- Ministerstvo hospodárstva SR

Navrhovaný partneri

- SBA
- Ministerstvo financií SR
- NASES

Odhadovaná dĺžka realizácie projektu (projektov)

- Pôjde o maximálne ročné malé projekty komunit a podnikateľov

Závislosti na iných projektoch a prioritách (OPII a OPIS)

- Centrálny dátový sklad
- Otvorené údaje pre samosprávu;
- Open API;

Závislosti na iných projektoch a prioritách (OP EVS)

- Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov

2.6.5 Centrálna kapacita pre koordináciu a dohľad nad publikovaním otvorených údajov

Popis a cieľ

Hlavným cieľom národného projektu je plošné zvýšenie dostupnosti a kvality otvorených datasetov ako aj zavádzanie nových štandardov v oblasti Linked Data a sémantického webu prostredníctvom vytvorenia Centrálnej kapacity. Toto centrum bude mať na starosti podporu kontinuálneho zlepšovania poskytovania otvorených údajov povinnými osobami a nasadzovanie inovatívnych nástrojov, nových štandardov, metód a technológií na centrálny katalóg otvorených údajov ako aj získavanie a spracovanie spätnej väzby od komunity využívajúcej otvorené údaje. Dopadom projektu bude vytvorenie kreatívnej ekonomiky nad otvorenými údajmi ako aj zvýšenie schopnosti verejnej správy využívať svoje dáta ako vzácny zdroj.

Aktivity

- Zvýšenie dostupnosti verejných služieb – zlepšenie prístupu k dátam a informáciám (zavádzanie princípov otvoreného vládnutia),
- Budovanie a posilňovanie inštitucionálnych kapacít: vybudovanie centrálnej kapacity, ktorá sa bude podieľať na reforme verejnej správy v oblasti transparentnosti, poskytovania údajov a práce s dátovými zdrojmi.
- Vzdelávanie zamestnancov v oblasti licencovania datasetov, práce s dátovými štruktúrami a nástrojmi pre transformáciu a anonymizáciu datasetov, Linked Data, ochrany osobných údajov a utajovaných skutočností a ďalších témach spojených s transparentnosťou a poskytovaním otvorených údajov (vrátane prípravy obsahu pre centrálny eLearning).
- Znižovanie korupcie vo VS sprístupnením dokumentov a dát širokej verejnosti (v rámci podpory otvoreného vládnutia)

Navrhovaný realizátor

- Bude určené na základe rokovania s MV SR

Navrhovaný partneri

- -.

Odhadovaná dĺžka realizácie projektu (projektov)

- 4 roky

Závislosti na iných projektoch a prioritách (OPII a OPIS)

- Centrálny dátový sklad
- Otvorené údaje pre samosprávu
- Dopytový projekt pre inovatívne komunity a podnikateľov

Závislosti na iných projektoch a prioritách (OP EVS)

- Vybudovanie dátovej kancelárie verejnej správy

