

Metodická príručka pre povinné osoby – Ako zverejňovať otvorené dáta

(Verzia 1 - 9)

Jún 2015

Tento dokument obsahuje 59 strán

1 Obsah

1	Obsah	i
2	Úvod	1
3	Oboznámenie sa s otvorenými dátami (OpenData)	3
3.1	Základné pojmy	3
3.2	Kontext otvorených dát	6
3.3	Stratégia nasadzovania otvorených dát	7
3.4	Ciele projektu eDemokracia	8
3.5	Výsledky projektu	9
3.5.1	Princípy Modulu otvorených dát	9
4	Povinnosti povinných osôb a ich plnenie	15
4.1	Legislatíva súvisiaca s publikovaním otvorených dát	15
4.2	Povinnosti vyplývajúce z legislatívy	16
5	Odporúčané organizačné zabezpečenie	18
6	Príprava plánu na publikovanie dát	21
6.1	Identifikovanie datasetov na zverejnenie	21
6.1.1	Analýza dostupných dátových zdrojov	22
6.1.2	Identifikovanie datasetov na zverejnenie	24
6.1.3	Stanovenie náročnosti zverejnenia jednotlivých datasetov	28
6.1.4	Stanovenie priorít na zverejnenie datasetov	32
6.2	Definovanie plánu zverejňovania	34
6.2.1	Stanovenie cieľového stavu	34
6.2.2	Definovanie plánu zverejnenia otvorených dát	35
7	Spracovanie dát na publikovanie	37
7.1	Príprava prístupu k dátovým zdrojom	37
7.2	Prevedenie transformačných procedúr	39
7.3	Popis datasetov (pridanie metadát)	40
7.4	Definovanie licencií	42
8	Publikovanie dát na portáli	47
8.1	Zverejnenie datasetov	47
8.2	Katalogizácia dát	48
9	Aktualizácia publikovaných dát	49
9.1	Správa datasetov	49
9.2	Aktualizácia dátových zdrojov	50

10	Archivácia publikovaných dát	51
10.1	Ukončenie správy datasetov	51
10.2	Ukončenie publikovania datasetov	51
11	Prílohy	53
11.1	Checklist na splnenie aktivít	53
11.2	Aktualizácia informačných systémov verejnej správy	55

2 Úvod

Účel metodiky

Metodika uvedená v tomto dokumente predstavuje ucelenú súpravu metodických odporúčaní na publikovanie otvorených dát vo verejnej správe. Metodika je určená všetkým organizáciám verejnej správy (povinným osobám), ktoré majú povinnosť sprístupňovať dáta vo forme otvorených dát vyplývajúcu z legislatívy, alebo sa rozhodli zverejniť niektoré dátové zdroje pred účinnosťou zákona o otvorených dátach.

Rámec aktivít na publikovanie otvorených dát

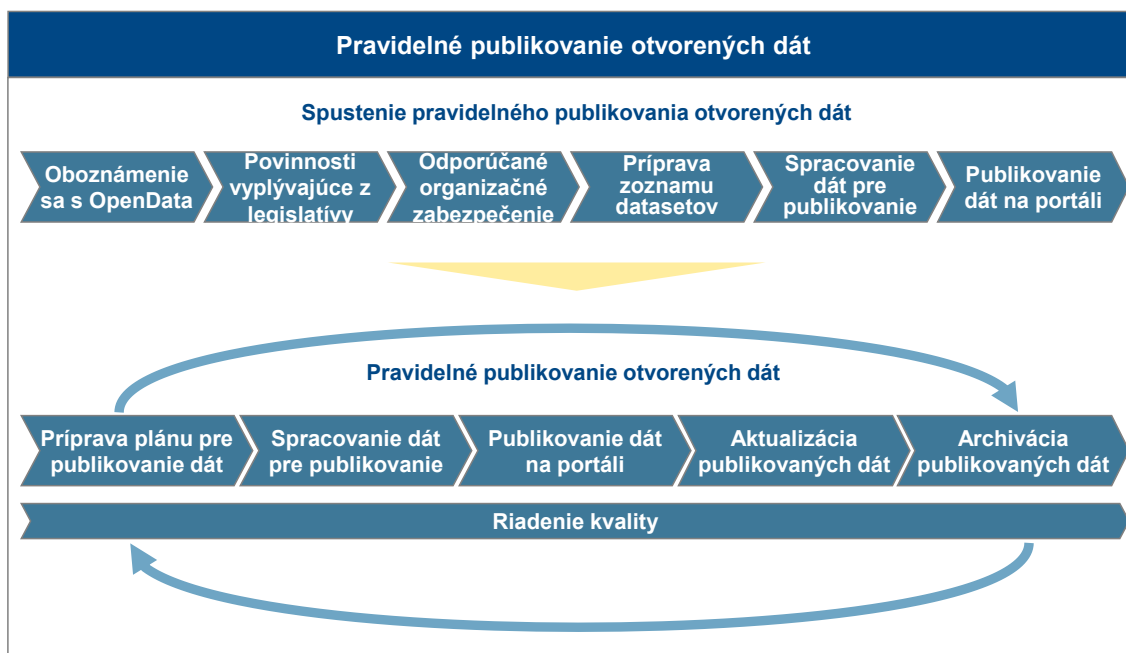
V úvodnej prípravnej fáze metodiky, nazývanej tiež Spustenie pravidelného publikovania otvorených dát, sa povinná osoba oboznámi s celým kontextom publikovania otvorených dát a pripraví sa na pravidelné a systematické publikovanie. Táto metodika ponúka tipy a nástroje nielen na splnenie si všetkých nevyhnutných povinností vyplývajúcich z legislatívy, ale aj na zjednodušenie a zefektívnenie práce spojenej so zverejňovaním. V rámci tejto prípravnej fázy, sa metodika venuje nasledujúcim oblastiam:

- Oboznámeniu sa s otvorenými dátami (OpenData), v ktorom sa povinná osoba dozvie o základných pojmoch, kontexte otvorených dát a stratégie ich nasadzovania do každodenného fungovania verejnej správy a o projekte eDemokracia (kapitola 3).
- Povinnostiam vyplývajúcim z platnej legislatívy (kapitola 4).
- Odporúčanému organizačnému zabezpečeniu (kapitola **Chyba! Nenašiel sa žiaden zdroj odkazov.**).
- Príprave zoznamu datasetov na publikovanie (na základe Akčného plánu Iniciatívy otvoreného vládnutia 2012 – 2013 inštitúcie štátnej správy vypracovali zoznam datasetov štátnej správy,¹ ktoré sa štátna správa zaviazala zverejniť, tento zoznam je však potrebné doplniť a aktualizovať, inšpiráciu možno tiež čerpať z výstupov projektu COMSODE)² - (kapitola 6).
- Spracovaniu dát na publikovanie – odporúča sa odskúšať si nástroje na spracovanie dát a oboznámiť sa s metodickými postupmi (kapitola 7).
- Publikovaniu dát na portáli a ich aktualizácii - odporúča sa odskúšať si nástroje na publikovanie dát a oboznámiť sa s metodickými postupmi na ich aktualizáciu (kapitola 8 a 9).

¹ <http://www.otvorenavlada.gov.sk/datasety-statnej-spravy/>

² <http://www.comsode.eu>

Obrázok 1: Rámec aktivít na pravidelné publikovanie otvorených dát



Následne po prípravnej fáze - Spustenie pravidelného publikovania otvorených dát - bude prebiehať periodicky sa opakujúci cyklus pravidelného publikovania otvorených dát. Kroky tohto cyklu sú detailne popísané v kapitolách 6 až 10.

3 Oboznámenie sa s otvorenými dátami (OpenData)

Informácie sú hodnotným národným zdrojom a strategickým majetkom verejnej správy, jej partnerov a verejnosti. Na účely maximálneho využitia potenciálu informačných zdrojov sa odporúča, aby si jednotlivé povinné osoby spravovali svoje dáta ako majetok počas celého životného cyklu s cieľom propagovať otvorenosť a interoperabilitu. Prístup k informáciám ako k verejnému majetku zvýši efektivitu fungovania verejnej správy, zníži náklady na chod organizácií, zlepši služby občanom a podnikateľom, podporí zámery vlády k vyššej miere transparentnosti, ochráni osobné údaje a výrazne zdokonalí prístup verejnosti k cenným dátam verejnej správy.³ Umožní sa tak budovanie inovatívneho štátu a výrazne sa podporí prebiehajúca reforma verejnej správy.

3.1 Základné pojmy

Definícia základných pojmov	
Pojem	Definícia
Prístup cez aplikačné rozhranie - API	Ide o aktívny prístup, kde používateľ zadáva serveru dotazy na konkrétne požadované údaje, ktoré server po spracovaní dotazu vyhledá a odošle.
Centrálny katalóg otvorených dát	data.gov.sk - Centrálny katalóg otvorených údajov, ďalej „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.
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 na účely 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 Data	Prelinkovaním umožňujú vytvoriť ekosystém poznania - webových služieb (aplikácií), ktoré publikujú, obohacujú a využívajú dáta o entitách v jednom globálnom zdieľanom dátovom priestore (Web of data).

³ US Open Government Directive <https://www.whitehouse.gov/open/documents/open-government-directive>

Definícia základných pojmov	
Pojem	Definícia
Metadáta pre dataset	Štruktúrované údaje obsahujúce informácie o primárnych údajoch – dátovom zdroji, pričom primárne údaje spravidla reprezentujú určitý hmotný objekt alebo nehmotný objekt. Metaúdaje sú určené najmä na vyhľadávanie, katalogizáciu a využívanie primárnych údajov.
Otvorené údaje (OpenData)	Otvorené údaje (OpenData) je názov pre paradigmu, v ktorej je možné s údajmi 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 nasledujúce: ▪ 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.
Podporná aplikácia	Softvér na zverejňovanie dát a metadát, dodávaný ako template v rámci MOD. Zabezpečuje synchronizáciu dát a metadát medzi MOD a systémom povinnej osoby. Umožňuje spracovávať a zverejňovať dokumenty, relačné a textové dáta zo súborového systému, z databáz alebo API.
Poskytovateľ údajov	Je povinná osoba zodpovedná za správnosť a aktuálnosť údajov zverejnených v datasete ako otvorené údaje. Je poverená vykonávať riadenie a koordinovanie registrácie a zverejnenia datasetu určitého úseku verejnej správy.
Používateľ údajov	Používateľom údajov je osoba, organizácia alebo informačný systém, ktorí používajú alebo požadujú poskytovanie datasetov otvorených údajov verejnej správy.

Definícia základných pojmov	
Pojem	Definícia
Referencovateľný identifikátor	Ide o 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.: http://data.gov.sk/data/uvo/zmluvy/zmluvy Referencovateľným identifikátorom je teda identifikátor, ktorý: 1. má formát Uniform Resource Identifier (URI), 2. je jednoznačný, 3. je unikátny, 4. je dlhodobý stabilný, 5. je formátovo a štrukturálne konzistentný, 6. je manažovateľný tak, aby umožňoval logicky rozširovať stanovenú štruktúru, 7. je jasný, stručný a krátky, 8. je pre fyzickú osobu jednoducho čitateľný, 9. je bez súborových prípon, 10. neobsahuje programátorské kľúčové slová, 11. neobsahuje reťazec „www“, 12. 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, 13. obsahuje iba malé písmená, 14. nahrádza špeciálne znaky, napríklad výkričník, úvodzovky, percento, hviezdička, zátvorka, dolár alebo mriežka, pomlčkami a podčiarkovníkmi,
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.
URI - Jednotný identifikátor zdroja	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. ▪ Príklady absolútnej URI: http://somehost/absolute/URI/with/absolute/path/to/resource.txt ▪ Príklady relatívnej URI: /relative/URI/with/absolute/path/to/resource.txt

Definícia základných pojmov	
Pojem	Definícia
URL - Jednotná adresa zdroja	Je reťazec znakov s definovanou štruktúrou, ktorý slúži na presnú špecifikáciu 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: <i>protokol://server.doména_druhého_rádu.generická_doména:port/umiestnenie_na_serveri?formulárové_dáta#kotva</i> Napríklad: http://cs.wikipedia.org:80/w/wiki.phtml?title=URL&action=edit ▪ 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.

3.2 Kontext otvorených dát

Zámerom iniciatívy otvorených dát je zverejnenie informačných zdrojov verejnej správy tak, aby sa dali jednoducho strojovo spracovať a použiť verejnosťou bez obmedzení, teda ide o zverejnenie vo formáte otvorených dát. Tento zámer je v súlade s cieľmi vládou schváleného Operačného programu Integrovaná infraštruktúra, prioritná os číslo 7 Informatizácia spoločnosti,⁴ ktorá obsahuje samostatný špecifický cieľ pre otvorené dáta. Operačný program vychádza zo Strategického dokumentu pre oblasť rastu digitálnych služieb a oblasť infraštruktúry prístupovej siete novej generácie (2014 – 2020), v ktorom sa Slovensko rozhodlo, že bude zverejňovať dáta verejnej správy ako otvorené dáta.

Aktivity spojené s publikovaním otvorených dát budú tiež podporené novou pripravovanou legislatívou o otvorených dátach, ktorá vzniká transpozíciou európskej smernice 2013/37/EÚ o opakovanom použití informácií verejného sektora.

Metodika sa ďalej odvoláva na prijímanú Stratégiu sprístupnenia a používania otvorených dát verejnej správy, ktorú pripravuje vedúci Úradu vlády Slovenskej republiky, a na schválený Akčný plán iniciatívy pre otvorené dáta na roky 2014 až 2015, a to konkrétne na tieto tri úlohy:

- Na základe plánu zverejňovania zverejňovať datasety na portáli otvorených dát;
- Vykonať prieskum dopytu verejnosti po najžiadanejších datasetoch;
- Na základe vyhodnotenia dopytu verejnosti po najžiadanejších datasetoch prednostne zverejniť príslušné datasety na portáli otvorených dát.

⁴ <http://www.informatizacia.sk/operacny-program-integrovana-infrastruktura/19080s>

3.3 Stratégia nasadzovania otvorených dát

Verejná správa produkuje a ukladá veľké množstvo cenných informácií, ktoré sú momentálne vo veľkej miere nesprístupnené na ďalšie využitie. Ako vyplýva zo Strategického dokumentu pre oblasť rastu digitálnych služieb a oblasť infraštruktúry prístupovej siete novej generácie (2014 – 2020),⁵ inštitúcie verejnej správy si nevymieňajú dostatok dát ani medzi sebou. Nedostatok informácií v informačnej spoločnosti bráni dospieť k optimálnym rozhodnutiam a vytvoriť si správnu predstavu o skutočnostiach. Potenciál otvorených dát tvoriť pridanú hodnotu tak ostáva nevyužitý, napríklad na zvýšenie stupňa transparentnosti. Ak sa zamyslíme, že trhovú hodnotu otvorených dát verejnej správy v Európe je odhadovaná na 32 miliárd eur,⁶ je pochopiteľné, že v analyzovaných stratégiách EÚ (Dánsko, Estónsko, Fínsko, Nemecko, Rakúsko, Švédsko, Veľká Británia)

táto téma rezonuje. V prípade Austrálie sú dokonca verejné informácie centrálnym cieľom.

Všetky dáta verejnej správy, ktoré nepodliehajú utajeniu alebo neobsahujú osobné údaje, musia byť zverejňované ako otvorené dáta, ideálne prostredníctvom verejne prístupných rozhraní (API), ktoré umožnia ich strojové spracovanie.

Stratégiu je možné vyjadriť v nasledujúcich krokoch:

1. Vypracovanie koncepcie otvorených dát.
2. Vybudovanie centrálneho portálu pre otvorené dáta.
3. Nasadenie nástrojov na vytváranie otvorených dát.
4. Vytvorenie možnosti na kontrolu používania dát.
5. Portál pre otvorené dáta bude orientovaný na vznik inovatívnych komunit.

Vypracovanie koncepcie otvorených dát

Prvotným krokom je presné definovanie koncepcie používania otvorených dát vo verejnej správe, prijatie legislatívy a výber štandardov a pravidiel a licenčných politík, na základe ktorých budú otvorené dáta zverejňované. Za cieľový stav podoby otvorených dát je možné považovať formát LinkedData.

Vybudovanie centrálneho portálu pre otvorené dáta

Ďalším krokom je rozvoj jednotnej centrálnej platformy nadväzujúci na skúsenosti s portálom data.gov.sk, ktorý bol prvotnou snahou v tejto oblasti ešte pred prijatím ucelenej stratégie, a tým neposkytuje funkcionality v uspokojivom rozsahu. Na platforme bude k dispozícii centrálny katalóg pre všetky otvorené dáta. Platforma tiež bude obsahovať implementované rozhrania na prístup k dátam a možnosť ukladania dát pre inštitúcie, ktoré nevyužijú svoje lokálne možnosti. Inštitúcie, ktoré budú dáta ukladať centrálnu, získajú metodickú podporu a transformačné nástroje pre prácu s dátami v rámci nasadzovania nástrojov na sprístupňovanie, ako je uvedené nižšie.

Nasadenie nástrojov na vytváranie otvorených dát

Dôležitou podmienkou na vytváranie otvorených dát je dôkladná analýza systémov na procesnej, aplikačnej a technologickej úrovni, aby boli známe možnosti, ktoré dáta bude možné vytvárať a ako efektívne toto vytváranie zabezpečiť. Prvým krokom je identifikácia zdrojov dát, ktoré je možné vytvárať alebo by bolo vhodné vytvárať. Zároveň sa určí aj kvalita, v akých sa tieto zdroje nachádzajú. Ďalej je potrebné zaoberať sa aj otázkou priority týchto dátových zdrojov z pohľadu ich využiteľnosti, pričom prednosť budú mať dáta, ktoré súvisia s výkonom

⁵ <http://www.informatizacia.sk/strategicky-dokument/16604s>

⁶ <http://ec.europa.eu/digital-agenda/en/pillar-i-digital-single-market/action-3-open-public-data-resources-re-use>

verejnej správy, či už referenčné údaje v podobne registrov, alebo údaje o vykonaných transakciách.

Z tejto analýzy budú vychádzať navrhnuté kroky pre konkrétnu inštitúciu. Výsledkom bude celková koncepčná, procesná a organizačná zmena fungovania organizácie tak, aby bola schopná vytvárať otvorené dáta a adekvátne nasadenie informačných technológií. Prechod k vytváraniu otvorených dát sa tak stáva dôležitou súčasťou celkovej reformy verejnej správy aj na procesnej úrovni.

Všetky nové informačné systémy budú budované tak, aby boli pripravené na otvorené dáta a boli tak schopné zverejniť informácie o svojich dátach do centrálnej platformy a zároveň poskytovať obsah (vrátane rozhraní - API). Ostatné systémy budú modifikované spôsobom, aby dokázali generovať vo vhodnom formáte a tvare požadované dáta a automaticky komunikovali s platformou. Tieto princípy budovania a rozvoja informačných systémov budú zakotvené v pripravovanej metodike „Open Data Ready“. Do procesu vytvárania otvorených dát budú nasadené aj korekčné a analytické nástroje, aby bolo zaručené správne publikovanie dát. Zároveň bude potrebné zabezpečiť dostatočnú mieru kvality existujúcich dát konsolidáciou dátových zdrojov, prepojovaním dát, čistením dát a podobne. Smerom navonok budú tieto informačné systémy verejnej správy obsahovať rozhrania, ktoré sprístupnia dáta v prehľadných procesoch.

Vytvorenie možnosti na kontrolu používania dát

Zverejňovanie nových typov dát vo veľkých množstvách na verejné použitie okrem pozitívnych efektov prináša aj nové druhy problémov, akými sú ochrana citlivých údajov (napríklad osobné údaje občana) či korektné používanie samotných dát. K narušeniu ochrany osobných údajov môže dôjsť napríklad v prípadoch, keď sa skombinujú rôzne datasety, ktoré samy osebe osobné údaje neprehrádzajú. Na riešenie takýchto a podobných problémov je potrebné, aby bolo používanie otvorených dát v budúcnosti čiastočne centrálné monitorované, a tým sa zabezpečila dôsledná kontrola citlivých údajov a samotný dohľad nad použitím dát.

Portál pre otvorené dáta bude orientovaný na vznik inovatívnych komunít

Publikovanie otvorených dát samo osebe nemá zmysel. Aby sa naplnilo ich poslanie, je nevyhnutné, aby vznikla inovatívna komunita, ktorá bude otvorené dáta aktívne používať, vytvárať aplikácie, diskutovať o inováciách a nových nápadoch. Takéto komunity budú podporované. Zároveň je mimoriadne zaujímavý prípadný výskum univerzít týkajúci sa otvorených dát.

Verejná správa sa bude môcť vznikajúcimi inováciami inšpirovať, využívať vznikajúce aplikácie vo svojich procesoch, tvoriť vlastné aplikácie pre svoje potreby rozhodovania i verejnosť ako nové typy služieb.

Portál otvorených dát v Module otvorených dát projektu eDemokracia predstavuje prvý krok pri naplňaní tejto stratégie.

3.4 Ciele projektu eDemokracia

Hlavným cieľom projektu je vytvorenie integrovaného informačného systému eDemokracie a otvorenej vlády (IIS eDOV) a portálu na poskytovanie transparentných informácií a otvorených dát širokej verejnosti, ako aj inštitúcií verejnej správy spolu s vytvorením nových elektronických služieb v zmysle iniciatívy otvoreného vládnutia.

Špecifické ciele:

- Vytvorenie štandardizovanej platformy na využívanie OpenData verejnej správy na Slovensku a novej služby Elektronická hromadná žiadosť.
- Vytvorenie podmienok otvorenej vlády prostredníctvom elektronizácie procesov služieb vlády SR a činností Úradu vlády SR.

- Vybudovanie komponentu dotačných schém, ktorý umožní v prehľadnej a interaktívnej forme vyhľadať informácie o dotačných mechanizmoch a ich využití v SR.
- Vytvorenie efektívneho aplikačného nástroja na poskytovanie výstupov z ITMS pre verejnosť a verejnú správu.

3.5 Výsledky projektu

Navrhovaný integrovaný informačný systém eDemokracie sa skladá z nasledujúcich modulov:

- **Modul otvorených dát:** Rozvojom riešenia portálu data.gov.sk vznikne priestor na poskytovanie kvalitných dát verejnej správy v otvorených formátoch spolu so zabezpečením ich prezentácie pre koncového používateľa údajov. Cieľom je výrazne zvýšiť dostupnosť údajov otvorenej vlády a znížiť čas potrebný na poskytnutie dát, ako aj administratívnu záťaž s tým spojenú. Zároveň sa vytvorí priestor na tvorbu nových pohľadov na údaje, s ktorými pracuje verejná správa. Tento výsledok sa dosiahne:
 - využitím a prepojením údajov poskytovaných rôznymi informačnými systémami verejnej správy, tzv. LinkedData;
 - poskytnutím rozhraní pre dáta na vytváranie špecializovaných aplikácií nad týmito dátami pre tretie strany;
 - tvorbou špecializovaných aplikácií na analýzu údajov;
 - dostupnosťou údajov platných pre rôzne časové obdobie.
- **Elektronická hromadná žiadosť:** Modul je určený na podávanie, tvorbu, hlasovanie a spracovanie elektronických hromadných žiadostí. Ďalej modul bude zabezpečovať prehľadné zobrazenie podaných žiadostí v stave riešenia, otvorených žiadostí na hlasovanie, uzavretých žiadostí. Vybudovaním tohto riešenia bude môcť vláda reagovať na podnety občanov skôr a efektívnejšie. Využitím nových technológií sa uľahčí komunikácia občanov a zjednoduší spôsob, ktorým sa budú môcť pripojiť k elektronickej hromadnej žiadosti.
- **Otvorená vláda:** Účelom elektronizácie oblasti otvorenej vlády je poskytovanie kompletného informačného servisu o činnosti a rozhodnutiach vlády Slovenskej republiky. Informácie, dokumenty a výmena dát bude prebiehať automaticky, nebude tak dochádzať k nepresnostiam a nekonzistentnosti dát na vstupe a výstupe procesov. Malo by dôjsť k skráteniu času, zníženiu administratívnej náročnosti a zvýšeniu informovanosti. Informácie budú prezentované transparentne a budú dostupné aktuálne a včas.
- **Dotačné schémy:** Modul predstavuje funkcionality zabezpečujúcu prehľadné zverejňovanie dotačných schém na účely informovania sa o možnostiach a podmienkach získania dotácií prostredníctvom verejnej správy. Ide o dotačné schémy v Slovenskej republike využívajúce verejné prostriedky zo štátneho rozpočtu, z rozpočtu samosprávy, dotácie z mimorozpočtových prostriedkov a dotácie/granty zo zahraničných zdrojov manažovaných, verejnými inštitúciami, ako aj ostatné dotačné schémy využívané v podmienkach Slovenskej republiky orgánmi verejnej moci.
- **ITMS OpenData:** Umiestnenie dát z jednotlivých programových období v jednom komponente uľahčí prístup občana, verejnosti a štátnej správy k predmetným dátam. Komponent kladie dôraz na jednoduché a intuitívne ovládanie, jednoduché vyhľadávanie relevantných dát a na jednoduché vytváranie, modelovanie analytických prehľadov a štatistických zisťovaní.

3.5.1 Princípy Modulu otvorených dát

Riešenie otvorených dát je koncipované ako otvorené, založené na najnovších štandardoch pre architektúru. Modul otvorených dát kladie dôraz na jednoduché a intuitívne ovládanie, efektívne vyhľadávanie relevantných dát a na zrozumiteľné vytváranie, modelovanie analytických

prehľadov a štatistických zisťovaní. Otvorené dáta a analytické výstupy z ich spracovania prispievajú k zlepšeniu rozhodovacích procesov a posunú kvalitu poskytovania služieb verejnej správy na novú úroveň. Analytické spracovanie údajov patrí k nástrojom reformy verejnej správy. Očakáva sa kvalitnejší návrh politík a regulácií a skúmanie ich dosahov na základe reálnych dát.

Portál otvorených dát bude pre povinné osoby poskytovať nasledujúcu funkcionálnosť:

- Zverejňovanie dát v strojovo spracovateľnej forme s použitím otvorených štandardov a verejných licencií s možnosťou:
 - zverejňovania metadát (ďalšie prístupné, v strojovo spracovateľnej forme s použitím otvorených štandardov a verejných licencií);
 - prepájania údajov poskytovaných rôznymi informačnými systémami verejnej správy (tzv. LinkedData);
 - konceptuálneho riešenia dostupnosti údajov platných pre rôzne časové obdobie.
- Nástroje pre extrahovanie dát z databáz, ich transformáciu a nahranie do centrálného dátového skladu (ETL nástroje – Extract, Transform, Load).
- Služby úložiska datasetov na centrálnom portáli.
- Sprístupnenie analytických aplikácií a nástrojov Business intelligence nad dátovým skladom otvorených dát.

Portál otvorených dát pre občanov, podnikateľov a neziskový sektor umožní:

- Poskytovanie a prehľadávanie metadát.
- Prístup k dátam vo forme spĺňajúcej otvorené a technologicky neutrálne štandardy, s použitím verejných licencií, ktoré umožnia ďalšie využívanie dát:
 - poskytovanie dát v csv formáte a v ďalších strojovo spracovateľných formátoch;
 - spracovanie žiadostí na nové dáta.
- Podávanie podnetov na zlepšenie kvality poskytovania otvorených dát.
- Vyhľadávanie dokumentov (neštruktúrovaných datasetov).
- Poskytnutie rozhraní pre dáta na vytváranie špecializovaných aplikácií nad otvorenými dátami pre tretie strany:
 - umožnenie tvorby špecializovaných aplikácií na analýzu údajov.
- Sprístupnenie analytických aplikácií a nástrojov Business intelligence nad dátovým skladom otvorených dát pre externých používateľov údajov.

3.5.1.1 Zverejňovanie dát na portáli

Otvorené dáta zverejňované na portáli musia spĺňať štandardy ISVS pre kvalitu datasetu.⁷ Bola prebraná schéma 5★, na základe ktorej je minimálna úroveň kvality pre OpenData definovaná ako 3★,⁸ čo znamená:

- Prístupnosť online pre bezodplatné nekomerčné aj komerčné využitie.
- Štruktúrované údaje spolu so zverejnenou schémou.
- Otvorený formát: CSV, JSON, XML, RDF.
- Aktuálnosť údajov.

⁷ Výnos č. 55/2014 Z. z., § 51 – Požiadavky na datasety sú popísané v kapitole 6.1.3 Stanovenie náročnosti zverejnenia jednotlivých datasetov.

⁸ Táto povinnosť je účinná od 15. 3. 2015.

- Licencia na voľné opakované použitie s možnosťou modifikovať, dopĺňať, kombinovať údaje.

Centrálne licencie budú pripravené v rámci MOD, pričom sú preferované všeobecné licencie (Creative Commons).

Na zverejňovanie dát na portáli otvorených dát bude použitý nasledujúci postup s vypracovanou metodikou pre povinné osoby:

- Evidencia datasetu v centrálnom katalógu otvorených dát.
- Vytvorenie požiadavky o nahratie dát datasetu na portál otvorených dát.
- Definovanie metadát k novému datasetu alebo dátovému zdroju.
- Centrálne zaevidovanie aktuálne platného datasetu zverejňovateľom na portál s možnosťou nahratia datasetu do úložiskových kapacít portálu (po schválení uloženia).
- Zápis a aktualizácia údajov z dátového zdroja.
- Overenie štruktúry a konzistencie datasetu a jeho metadát - v prípade správnosti publikovanie datasetu na portáli.

Požiadavka na zverejnenie otvorených dát môže pochádzať buď od samotnej povinnej osoby, ktorej túto povinnosť ukladá príslušná legislatíva a pracovné postupy, alebo od externého záujemcu o určité údaje a informácie. Požiadavka na zverejnenie datasetu sa vybaví zaevidovaním datasetu do katalógu a uložením údajov na portál otvorených dát alebo zverejnením URL linky na údaje do prostredia povinnej osoby, ktorá údaje zverejňuje prostredníctvom vlastných kapacít. V ideálnom prípade bude proces zverejnenia príslušných datasetov automatizovaný a budú ho podporovať samotné informačné systémy verejnej správy.

Zverejňovanie metadát

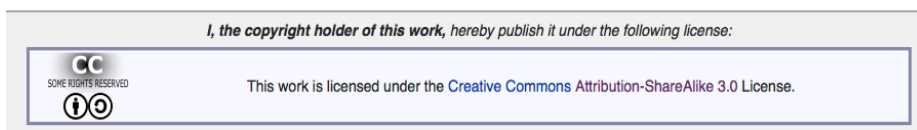
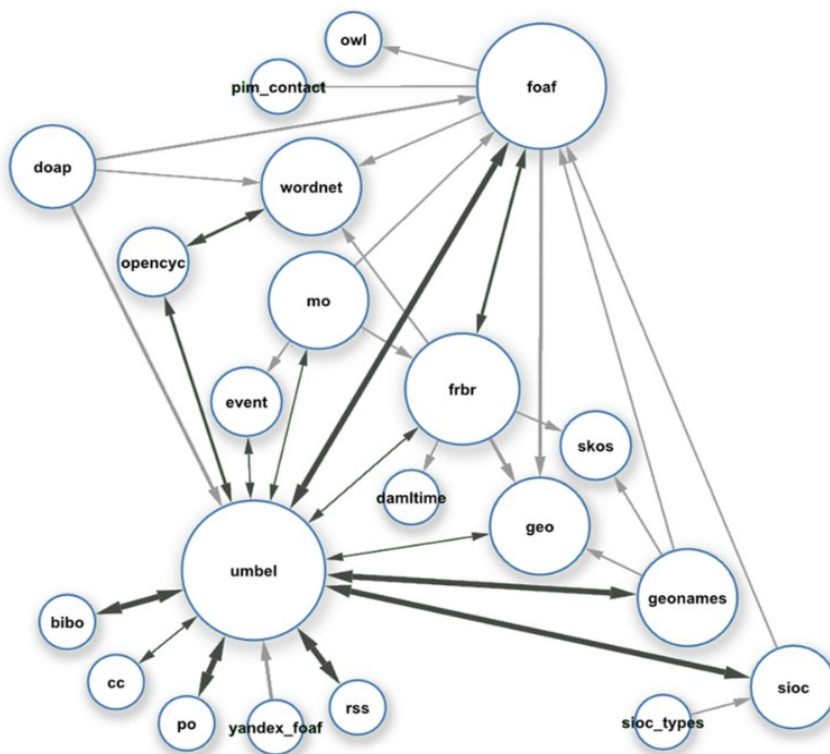
Pri zverejňovaní dát povinná osoba alebo informačný systém popíše dataset a dátový zdroj prostredníctvom metaúdajov, ktorú budú prístupné externým používateľom údajov v strojovo spracovateľnej forme s použitím otvorených štandardov a verejných licencií.

Metadáta o každom datasete budú evidované v centrálnom katalógu metadát, ktorý bude prístupný a prehľadateľný na portáli otvorených dát. K dispozícii bude komponent Správa katalógov ako kľúčové grafické rozhranie pre správu metadát o datasetoch. Umožňuje manuálne nahrávanie a aktualizáciu dát, nastavenie parametrov harvestovania. Umožňuje pridanie dátového zdroja k datasetu. Sprostredkúva poskytovateľom údajov správu údajov o prezentačných aplikáciách (aj tretích strán) a ich zverejňovanie. Vytvára na webe RSS kanál o zmenách metadát, umožňuje registráciu a zverejňovanie notifikácií o zmenách datasetov. Správa metadát je v zodpovednosti povinných osôb od 15. 3. 2015.

LinkedData

Koncept LinkedData umožňuje prepájanie údajov poskytovaných rôznymi informačnými systémami verejnej správy. LinkedData popisuje metódu zverejňovania štruktúrovaných dát, ktoré môžu byť prepájané pre zvýšenie užitočnosti. LinkedData sú založené na štandarde webových technológiách ako napríklad http, RDF a URIs, ale namiesto toho, aby slúžili ako webové stránky pre používateľov údajov, sú prispôsobené pre strojové automatické spracovanie. Tento prístup umožňuje prepájanie a spájanie dát z rôznych dátových zdrojov.

Obrázok 2: Koncept LinkedData



- Popis: Class diagram for the LOD datasets
- Pôvodný zdroj: https://upload.wikimedia.org/wikipedia/en/2/26/Linking-Open-Data-class-diagram_2008-10-05.png
- Originál obrázka: 934 × 856 pixelov, veľkosť súboru: 196 KB, MIME typ: image/png
- Dátum: 2008-10-10
- Autor: Michael K. Bergman
- Povolenie: https://commons.wikimedia.org/wiki/Commons:Reusing_content_outside_Wikimedia

Základné pravidlá pre LinkedData navrhol Tim Berners-Lee v knihe Design Issues: LinkedData. Pravidlá, ktoré sú implementované v MOD, sa dajú zhrnúť nasledujúcim spôsobom:

- Používanie URI pre označenie datasetov – napríklad [/relative/URI/with/absolute/path/to/resource.txt](#).
- Používanie HTTP URI, aby bolo možné dohľadať a referencovať datasety používateľmi údajov a rôznymi agentmi - `http://data.gov.sk/set/<názov datasetu>`.
- Poskytovanie užitočných informácií o datasete, v ktorých je URI referencovateľná, čo umožňuje rozvoj formátov ako RDF a SPARQL.
- Zahŕňanie odkazov do odvodených datasetov pomocou ich URI v prípade zverejňovania datasetov na internete.

Dostupnosť údajov platných pre rôzne časové obdobie

Komponent správy katalógov vytvára na webe RSS kanál o zmenách metadát, umožňuje registráciu a zverejňovanie notifikácií o zmenách datasetov.

API budú poskytovať informácie o aktuálnosti dát podľa Výnosu MF SR o štandardoch ISVS, aby bolo možné efektívne zisťovať, ktoré datasety a ich jednotlivé položky boli kedy:

- Pridané,
- Menené,
- vymazané.

Dostupné budú informácie o:

- dátume prvého zverejnenia datasetu,
- aktuálnosti datasetu podľa výnosu 55/2014 Z. z. § 53 písm. g),
- začiatku a konci účinnosti datasetu alebo jeho údajov, najmä ak je účinnosť odlišná od platnosti.

3.5.1.2 ETL nástroje – Extract, Transform, Load

Hlavnou úlohou transformácie je zvyšovanie kvality datasetov. K jednotlivým typom dát z dátového zdroja je možné priradovať konkrétne transformácie. Cieľom služby pre automatickú transformáciu je umožniť a uľahčiť vytváranie a publikovanie dát v štandardizovaných OpenData formátoch pre povinné osoby. Služba umožní importovať dáta z vybraných dátových zdrojov, definovať technické a organizačné metadáta a (automaticky alebo poloautomaticky) klasifikovať jednotlivé dokumenty podľa typu a obsahu (napr. zmluva, faktúra) a poskytnúť dáta pre katalogizačný komponent cez API na zaevidovanie metadát a dát.

Transformácie zohľadňujú definované ontológie a pracujú s ontologickým editorom. Ontologický editor umožňuje vytváranie a správu dátových slovníkov, pridelovanie jednoznačných identifikátorov a ontológií. Spravuje ontológie, tezaury, dátové prvky, jednoznačné identifikátory. Jeho primárnym používateľom je dátový kurátor alebo špecializovaný správca ontológií.

3.5.1.3 Služby úložiska datasetov na centrálnom portáli

Povinné osoby budú mať k dispozícii softvér na zverejňovanie dát a metadát, dodávaný v rámci MOD, aby sa im výrazne zjednodušili procesy na publikovanie otvorených dát. Tento nástroj zabezpečuje synchronizáciu dát a metadát medzi MOD a systémom povinnej osoby. Umožňuje spracovávať a zverejňovať dokumenty, relačné a textové dáta zo súborového systému, databáz alebo API.

Povinné osoby si budú môcť nahrať svoje dáta na portál buď do:

- Externého repozitára, ktorý obsahuje údaje o datasetoch a ich metadátach, ktoré sú zverejnené.
- Interného repozitára, ktorý obsahuje údaje o datasetoch a ich metadátach, ktoré nie sú zverejnené.

Pritom budú môcť využívať pre formáty:

- TXT, RTF a PDF: Súborový systém – primárna forma uloženia, nízka kvalita dát.
- CSV, JSON, XML: Relačnú databázu, ktorá slúži na ukladanie informácií o datasetoch, číselníkoch a ukladá dáta v relačnej podobe z vybraných dátových zdrojov.
- RDF (Linked Data): RDF úložisko, ktoré predstavuje špecifickú databázu pre ontológie a LinkedData. Využívajú ho všetky komponenty na prácu s LinkedData.

Jednotlivé typy úložísk budú prepojené. Navyše bude k dispozícii Data warehouse obsahujúci predspracované údaje pre potreby analytických nástrojov nad otvorenými dátami.

3.5.1.4 Sprístupnenie analytických aplikácií a nástrojov Business intelligence

Analytické aplikácie predstavujú aplikačnú vrstvu pre analytické a špecifické aplikácie na prácu s datasetmi, ktoré využívajú dátové služby. Analytické aplikácie poskytnú rôzne formy prezentácie údajov a špecializované aplikácie poskytnú spracovanie konkrétnych špecifických

datasetov, vykonajú presne definované výpočty, transformácie a poskytnú atraktívnu formu výstupov používateľom údajov.

4 Povinnosti povinných osôb a ich plnenie

Modul otvorených dát poskytne nástroje na efektívne a pohodlné plnenie povinností povinných osôb v oblasti publikovania otvorených dát.

4.1 Legislatíva súvisiaca s publikovaním otvorených dát

Slovenská legislatíva

- Ústava SR - čl. 26: „Sloboda prejavu a právo na informácie sú zaručené“.
- Zákon NR SR č. 211/2000 Z. z. o slobodnom prístupe k informáciám (zákon o slobode informácií).
- Zákon NR SR č. 275/2006 Z. z. o informačných systémoch verejnej správy a o zmene a doplnení niektorých zákonov v znení zákona č. 678/2006 Z. z. a o zmene a doplnení niektorých zákonov v znení neskorších predpisov.
- Výnos Ministerstva financií Slovenskej republiky č. 55/2014 Z. z. o štandardoch pre informačné systémy verejnej správy.
- Uznesenie vlády č. 50/2012 k návrhu Akčného plánu iniciatívy pre otvorené vládnutie v Slovenskej republike.
- Uznesenie vlády č. 59/2015 k návrhu Akčného plánu iniciatívy pre otvorené vládnutie v Slovenskej republike na rok 2015.
- Zákon NR SR č. 618/2003 Z. z. o autorskom práve a právach súvisiacich s autorským právom (autorský zákon).

Európske predpisy

- Smernica 2003/98/ES Európskeho parlamentu a Rady zo 17. novembra 2003 o opakovanom použití informácií verejného sektora.
- Smernice 2013/37/EÚ o opakovanom použití informácií verejného sektora.

Ďalšie relevantné dokumenty

- Akčný plán iniciatívy pre otvorené vládnutie na rok 2015.
- Akčný plán iniciatívy pre otvorené vládnutie v Slovenskej republike na roky 2012 – 2013.
- Operačný program Integrovaná infraštruktúra, prioritná os číslo 7 Informatizácia spoločnosti.
- Strategický dokument pre oblasť rastu digitálnych služieb a oblasť infraštruktúry prístupovej siete novej generácie (2014 – 2020).
- Vyhláška ministra zahraničných vecí č. 133/1980 Zb. o Bernskom dohovore o ochrane literárnych a umeleckých diel z 9. septembra 1886, doplnenom v Paríži 4. mája 1896, revidovanom v Berlíne 13. novembra 1908, doplnenom v Berne 20. marca 1914 a revidovanom v Ríme 2. júna 1928, v Bruseli 26. júna 1948, v Štokholme 14. júla 1967 a v Paríži 24. júla 1971.
- Vyhláška ministra zahraničných vecí č. 134/1980 Zb. o Všeobecnom dohovore o autorskom práve revidovanom v Paríži 24. júla 1971.
- Oznámenie Ministerstva zahraničných vecí Slovenskej republiky č. 189/2006 Z. z. Ministerstvo zahraničných vecí Slovenskej republiky oznamuje, že dňa 20. decembra 1996 bola v Ženeve uzavretá Zmluva WIPO o autorskom práve (WCT).
- Smernica Európskeho parlamentu a Rady 9/96/ES o právnej ochrane databáz.

- Smernica Európskeho parlamentu a Rady 2001/29/ES o zosúladiení niektorých aspektov autorských práv a s nimi súvisiacich práv v informačnej spoločnosti.

Legislatíva týkajúca sa IKT (Informačno-komunikačných technológií)

- Zákon NR SR č. 275/2006 Z. z. o informačných systémoch verejnej správy;
- Zákon NR SR č. 215/2002 Z. z. o elektronickom podpise;
- Zákon NR SR č. 610/2003 Z. z. o elektronických komunikáciách;
- Zákon NR SR č. 215/2004 Z. z. o ochrane utajovaných skutočností;
- Zákon NR SR č. 428/2002 Z. z. o ochrane osobných údajov;
- Zákon NR SR č. 22/2004 Z. z. o elektronickom obchode;
- Zákon NR SR č. 211/2000 Z. z. o slobodnom prístupe k informáciám;
- Zákon NR SR č. 618/2003 Z. z. - autorský zákon.

Legislatíva týkajúca sa eGovernmentu

- Zákon NR SR č. 575/2001 Z. z. o organizácii činnosti vlády a organizácii ústrednej štátnej správy (kompetenčný zákon);
- Zákon NR SR č. 215/2002 Z. z. o elektronickom podpise;
- Zákon NR SR č. 610/2003 Z. z. o elektronických komunikáciách;
- Zákon NR SR č. 215/2004 Z. z. o ochrane utajovaných skutočností;
- Zákon NR SR č. 428/2002 Z. z. o ochrane osobných údajov;
- Zákon NR SR č. 511/1992 Zb. o správe daní a poplatkov;
- Zákon NR SR č. 199/2004 Z. z. colný zákon;
- Zákon NR SR č. 25/2006 Z. z. o verejnom obstarávaní;
- Zákon NR SR č. 25/2007 Z. z. o elektronickom výbere mýta za užívanie vymedzených úsekov pozemných komunikácií;
- Vyhláška MF SR č. 661/2006 Z. z., ktorou sa mení a dopĺňa vyhláška MF SR č. 419/2006 Z. z., ktorou sa vykonávajú niektoré ustanovenia zákona č. 199/2004 Z. z. colný zákon a o zmene a doplnení niektorých zákonov v znení neskorších predpisov, v znení vyhlášky č. 530/2006 Z. z.

4.2 Povinnosti vyplývajúce z legislatívy

Povinnosti povinných osôb na publikovanie otvorených dát predovšetkým vychádzajú z:

- § 3 ods. 4 písm. d) zákona č. 275/2006 Z. z.: Povinné osoby sú povinné sprístupňovať verejnosti údaje z informačných systémov verejnej správy, ak osobitný predpis neustanovuje inak a e) povinné osoby, ktoré sú správcami, sú povinné sprístupňovať alebo na požiadanie poskytnúť bezplatne iným povinným osobám údaje z informačných systémov verejnej správy potrebné na ich činnosť, ak osobitný predpis neustanovuje inak.
- § 6 ods. 3 zákona č. 211/2000 Z. z.: Povinné osoby, ktoré prevádzkujú informačné systémy obsahujúce informácie, pri ktorých osobitný zákon nevylučuje verejnú prístupnosť, sú povinné informácie obsiahnuté v týchto registroch a zoznamoch uverejňovať na voľne prístupnej internetovej stránke. Také zverejnenie nie je porušením osobitných predpisov.
- § 6 ods. 4 druhá veta zákona č. 211/2000 Z. z.: Okrem informácií podľa predchádzajúcich odsekov môže povinná osoba zverejniť aj ďalšie informácie.

- § 21 d ods. 1: Povinná osoba je povinná sprístupniť informácie na účel opakovaného použitia na základe žiadosti alebo ich môže sprístupniť vlastným rozhodnutím.
- Uznesenie vlády SR č. 59/2015 z 11. 2. 2015 – na základe ktorého je povinné vypracovať aktualizovaný zoznam všetkých datasetov orgánov uvedených ako zodpovedných vrátane rozpočtových a príspevkových organizácií v ich pôsobnosti spolu s plánom ich postupného zverejňovania na portáli otvorených dát. Na základe plánu zverejňovania je potrebné zverejňovať datasety na portáli otvorených dát a prioritizovať ich na základe vyhodnotenia dopytu verejnosti po najžiadanejších datasetoch

Zo spomínaných bodov vyplýva, že na základe existujúceho právneho stavu majú byť sprístupnené všetky údaje, okrem tých, pre ktoré to špecificky určitý zákon vylučuje.

Príkladom kategórie údajov, ktoré majú plošne obmedzený režim sprístupňovania a nemali by tak byť zverejnené na portáli otvorených dát, sú osobné údaje. Podmienky ich sprístupňovania upravuje zákon č. 428/2002 Z. z., najmä v § 7. Podľa tohto zákona je však možno sprístupniť, alebo zverejniť osobné údaje na základe súhlasu dotknutej osoby alebo napríklad ak to ustanovuje osobitný zákon.

Pri zverejňovaní otvorených údajov je ďalej potrebné dbať na ochranu utajovaných skutočností. Právnym základom ochrany utajovaných skutočností v Slovenskej republike je zákon č. 215/2004 Z. z. o ochrane utajovaných skutočností a o zmene a doplnení niektorých zákonov v znení neskorších predpisov (viac v časti 6.1.2.1).

Poslednou kategóriou problematických dát sú tie, na ktoré sa vzťahuje autorský zákon - zákon NR SR č. 618/2003 Z. z. o autorskom práve a právach súvisiacich s autorským právom. Tu treba získať súhlas autora na udelenie licencie na dataset podľa kapitoly 7.4.

Ak dátový zdroj obsahuje údaje, ktoré nie je možné sprístupniť, povinná osoba má povinnosť dátový zdroj sprístupniť po vylúčení týchto informácií. Vylúčenie týchto údajov môže byť dosiahnuté napríklad pri osobných údajoch ich anonymizáciou. Tento prístup je uplatňovaný aj pri iných spôsoboch sprístupňovania informácií, pozri napr. § 12 zákona č. 211/2000 Z. z. Informácie sa okrem zverejňovania na data.gov.sk v strojovo spracovateľnom formáte sprístupňujú na základe žiadosti aj nasledujúcimi spôsobmi:

- ústne,
- nahliadnutím do spisu vrátane možnosti vyhotoviť si odpis alebo výpis,
- odkopírovaním informácií na technický nosič dát,
- sprístupnením kópií predlôh s požadovanými informáciami,
- telefonicky, faxom, poštou alebo elektronickou poštou.

Ak informáciu nemožno sprístupniť spôsobom určeným žiadateľom, dohodne povinná osoba so žiadateľom iný spôsob sprístupnenia informácie. Dlhodobým zámerom iniciatívy otvorených údajov je však sprístupňovať informácie namiesto uvedených spôsobov len zverejňovaním na portáli otvorených dát.

Na základe pripravovanej novely zákona č. 211/2000 Z. z. dát je povinná osoba povinná sprístupniť údaje ako otvorené dáta v strojovo spracovateľnom formáte na portáli otvorených dát, a to i v prípade, ak sú napríklad údaje už dostupné cez webové stránky. Natento účel musí bez zbytočného odkladu zaevidovať informácie sprístupnené na opakované použitie v centrálnom katalógu otvorených údajov. Centrálny katalóg otvorených údajov, ďalej „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. Povinnosť zverejňovať sa však nevzťahuje na informácie, ku ktorým múzeá, knižnice, akademické knižnice, a archívy vykonávajú majetkové práva duševného vlastníctva, a s ktorými môžu nakladať v prospech iných osôb podľa osobitného predpisu.

Všetky štandardy týkajúce sa technického spôsobu sprístupňovania informácií vo forme otvorených údajov sú zachytené vo výnose 55/2014, ktorom sú definované pravidlá ako kvalita údajov, formáty a vytváranie referencovateľných identifikátorov.

5 Odporúčané organizačné zabezpečenie

Na publikovanie otvorených dát je nevyhnutné, aby povinná osoba zaistila dostatočné organizačné zabezpečenie. Minimálne požiadavky na organizačné zabezpečenie (roly a ich definícia) sú uvedené v nasledujúcej tabuľke.

Tabuľka 1: Odporúčaná definícia rolí a ich zodpovedností u povinnej osoby

Definícia rolí a ich zodpovedností v organizácii	
Rola	Definícia/zodpovednosť
Povinná osoba	Osoba, ktorá zodpovedá za poskytnutie datasetov verejnosti (vytvára datasety a publikuje ich na portáli otvorených dát data.gov.sk).
Vlastník	Osoba alebo subjekt, ktorý vlastní dataset - je držiteľom práv k datasetu alebo má právo s ním nakladať, alebo o ňom rozhodovať. Vlastník má právo rozhodnúť, či určitý súbor údajov je v súlade s legislatívou na publikovanie v podobe otvorených dát a má tiež právomoc poskytnúť licenciu k datasetu. Vlastník je gestorom zodpovedným za správnosť a aktuálnosť atribútov dátového prvku; gestor však nezodpovedá za obsah prenášaný dátovým prvkom. Vlastník spolupracuje na konzistencii dát so správcom obsahu zodpovedným za správu obsahu informačného systému verejnej správy a v ňom uložených informácií. Technickým prevádzkovateľom je prevádzkovateľ informačného systému verejnej správy podľa zákona, ktorý vykonáva činnosti určené správcom obsahu v súvislosti s technickou prevádzkou informačného systému.
IT profesionál	IT profesionál je osoba so zručnosťami a znalosťami v oblasti informačných a komunikačných technológií. Poskytuje podporu pri spracovaní dát, vyvíja a testuje ETL postupy a vykonáva transformáciu dát do formátov cieľových dát. V prípade zverejňovania otvorených dát v podobe LinkedData by mal mať potrebné vedomosti a zručnosti na zverejnenie (RDF, ontológie, URI dizajn, softvér nástroje na tvorbu LinkedData).
Manažér kvality dát	Manažér kvality dát je osoba so zručnosťami a znalosťami v oblasti kvality dát. Je zodpovedný za dohľad nad všetkými prvkami kvality dát podľa štandardov a nad životným cyklom kvality dát.
Právny expert	Právny expert je osoba so zručnosťami a znalosťami v oblasti práva, licencií a legislatívy. Poskytuje svoju odbornosť v oblasti licencovania otvorených datasetov a podieľa sa na analýze publikovaných datasetov, pri ktorých sa vyžaduje legislatívna expertíza, aby sa zabezpečilo zverejnenie datasetov v súlade s právnymi predpismi (predovšetkým v súlade s ochranou osobných údajov).

V prípade potreby sa môže povinná osoba obrátiť na expertov z Úradu vlády SR, uvedených v tabuľke nižšie (predovšetkým ide o dátového kurátora, správcu transformácií a hlavného právneho experta), ktorí v rámci svojej kompetencie poskytnú patričnú súčinnosť.

Tabuľka 2: Definícia rolí a ich zodpovedností na Úrade vlády SR

Definícia rolí a ich zodpovedností v organizácii	
Rola	Definícia/zodpovednosť

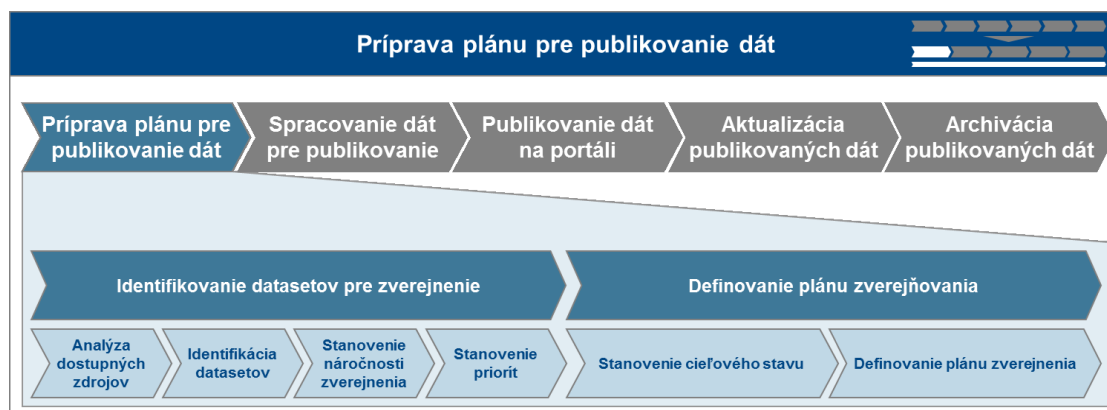
Definícia rolí a ich zodpovedností v organizácii	
Rola	Definícia/zodpovednosť
Hlavný dátový kurátor	Hlavný dátový kurátor je zodpovednou osobou pre otvorené údaje s celoslovenskou pôsobnosťou, ktorý vykonáva dozor nad sprístupňovaním informácií na opakované použitie a otvorených údajov.
Dátový kurátor	Osoba alebo subjekt, ktorý spravuje a udržiava datasety zverejnené v centrálnom katalógu a dohliada nad správnosťou metadát. Kurátor udržiava publikované datasety verejnej správy presné, aktuálne a v súlade so štandardmi. Ďalej sa venuje centrálnej správe a evidencii dostupných tagov na označovanie datasetov.
Ontologický kurátor (plánovaný v programovom období 2014 - 2020)	Vytvára, spravuje a pripomienkuje ontológie platné pre všetky povinné osoby (povinné osoby môžu podávať návrhy na centrálnu zaevidovanie platnej ontológie). Usmerňuje povinné osoby v oblasti vytvárania LinkedData a udržiava publikované datasety vo formáte LinkedData presné, aktuálne a v súlade so štandardmi.
Správca katalógu aplikácií	Schvaľuje žiadosti o zaregistrovanie novej aplikácie nad otvorenými dátami, dohliada nad správnosťou informácií zverejnených o aplikácii, pridáva nové aplikácie a vymazáva nerelevantné aplikácie.
Moderátor	Vykonáva dohľad nad diskusiami a komentármi na portáli data.gov.sk. Má oprávnenie vymazávať nevhodné komentáre a zbiera spätnú väzbu používateľov údajov z diskusií pre povinné osoby.
Správca transformácií (plánovaný v programovom období 2014 - 2020)	Správca transformácií spravuje transformačné predpisy - vytvára nové a aktualizuje existujúce. Informuje povinné osoby o dostupnosti nových transformácií a o spôsobe ich použitia.
IT profesionál	IT profesionál je osoba so zručnosťami a znalosťami v oblasti informačných a komunikačných technológií. Poskytuje podporu pri spracovaní dát, vyvíja a testuje ETL postupy a vykonáva transformáciu dát do formátov cieľových dát. V prípade zverejňovania otvorených dát v podobe LinkedData by mal mať potrebné vedomosti a zručnosti na zverejnenie (RDF, ontológie, URI dizajn, softvér nástroje na tvorbu LinkedData).
Manažér kvality dát (plánovaný v programovom období 2014 - 2020)	Manažér kvality dát je osoba so zručnosťami a znalosťami v oblasti kvality dát. Je zodpovedný za dohľad nad všetkými prvkami kvality dát podľa štandardov a nad životným cyklom kvality dát.
Expert na kvalitu dát (plánovaný v programovom období 2014 - 2020)	Expert na kvalitu dát je osoba so zručnosťami a znalosťami v oblasti fáz a nástrojov životného cyklu kvality dát. Je zodpovedný za analýzy, aplikovanie a vytváranie ETL komponentov v oblasti kvality dát. Vykonáva analýzy a poskytuje poradenstvo v oblasti identifikovania možných unikov osobných údajov alebo utajovaných skutočností, ktoré môžu nastať publikovaním a kombinovaním viacerých datasetov.

Definícia rolí a ich zodpovedností v organizácií	
Rola	Definícia/zodpovednosť
Hlavný právny expert (plánovaný v programovom období 2014 - 2020)	Právny expert je osoba so zručnosťami a znalosťami v oblasti práva, licencovania a legislatívy. Poskytuje svoju odbornosť v oblasti licencovania otvorených datasetov pre verejnú správu. Poskytuje právne poradenstvo, aby sa zabezpečilo zverejnenie datasetov v súlade s právnymi predpismi (predovšetkým v súlade s ochranou osobných údajov a citlivých skutočností).

6 Príprava plánu na publikovanie dát

Príprava plánu na publikovanie dát predstavuje strategickú činnosť, ktorej predchádza analytická práca s dostupnými dátovými zdrojmi.

Obrázok 3: Príprava plánu na publikovanie dát



Povinná osoba vykoná prípravu plánu na publikovanie dát v dvoch krokoch, ktoré sú detailne popísané v nasledujúcich podkapitolách:

- Identifikovanie datasetov na zverejnenie.
- Definovanie plánu zverejňovania.

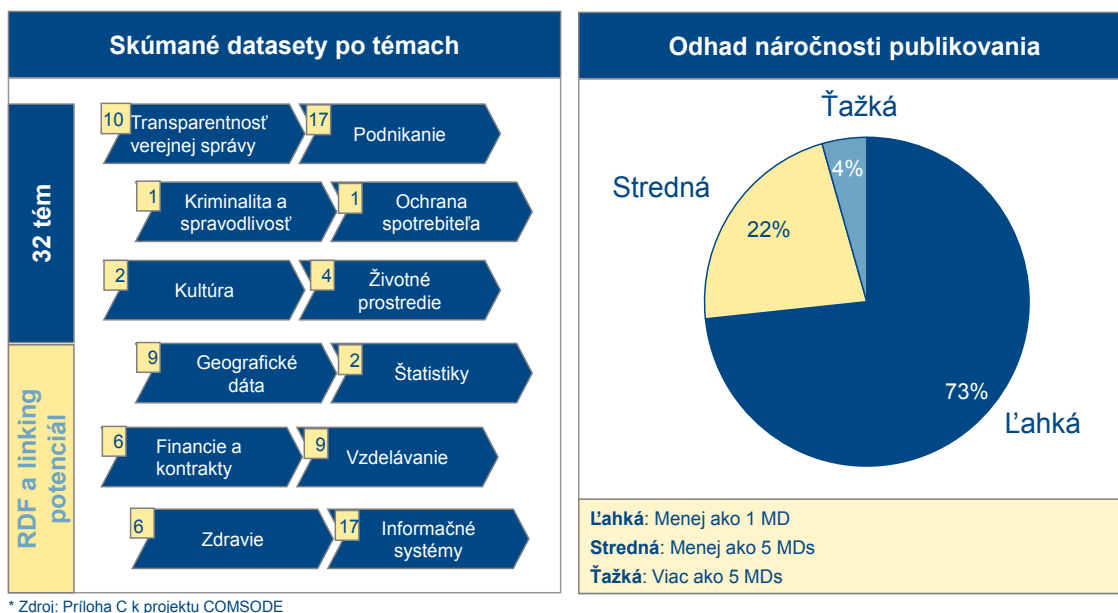
6.1 Identifikovanie datasetov na zverejnenie

Keďže nemožno očakávať (z finančných a personálnych dôvodov), že všetky dátové zdroje bude možné zverejniť ihneď, cieľom je v úvode identifikovať, ktoré datasety by mala povinná osoba prioritne zverejniť v podobe otvorených dát a ktoré by mali byť zverejnené v ďalších iteráciách, alebo by zverejnené nemali byť (napríklad z dôvodu dodržania zákona o utajovaných informáciách alebo zákona o osobných údajoch).

Inšpiráciou je výstup medzinárodného projektu COMSODE,⁹ ktorý sa snaží odstrániť bariéry pri zverejňovaní otvorených dát. V rámci Slovenskej republiky sa zaoberal 90 zaujímavými datasetmi v oblastiach znázornených na nasledujúcom obrázku. Pre jednotlivé datasety bola odhadnutá aj náročnosť ich publikovania.

⁹ Zdroj: Príloha C k projektu COMSODE.

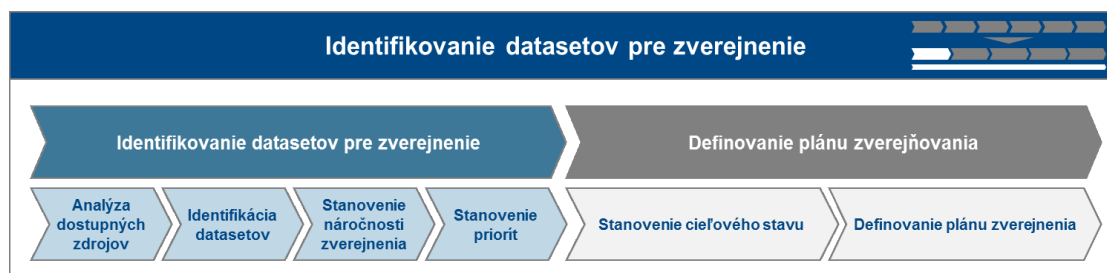
Obrázok 4: Inšpirácia na identifikovanie datasetov



Identifikovanie datasetov na zverejnenie bude prebiehať v štyroch krokoch:

1. Analýza dostupných zdrojov.
2. Identifikácia vhodných datasetov na zverejnenie.
3. Stanovenie náročnosti zverejnenia jednotlivých datasetov.
4. Stanovenie priorít na zverejnenie datasetov.

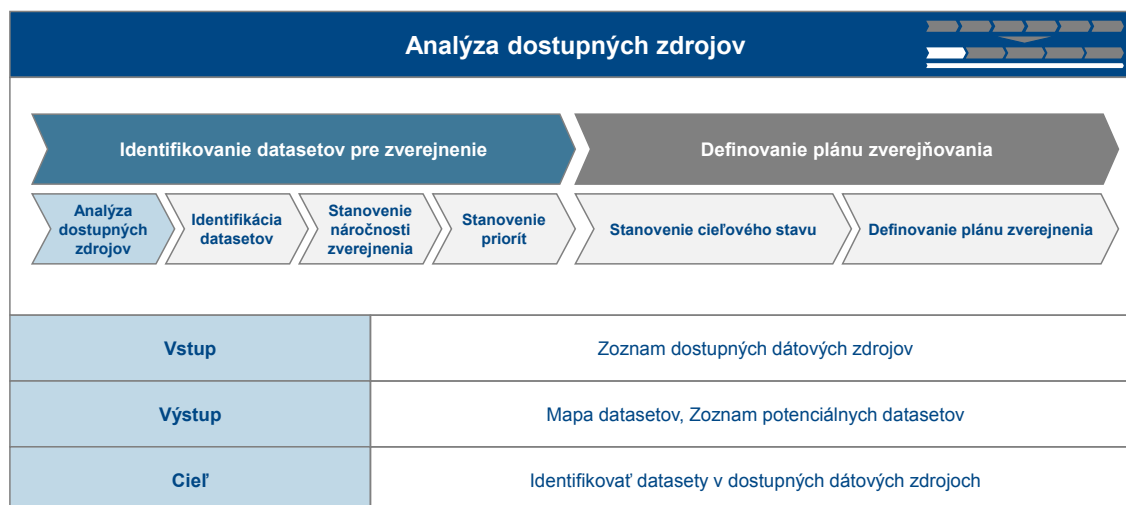
Obrázok 5: Identifikovanie datasetov na zverejnenie



6.1.1 Analýza dostupných dátových zdrojov

Vytvorenie plánu na publikovanie dát sa začína analýzou dátových zdrojov spravovaných povinnou osobou a identifikovaním potenciálnych datasetov na základe obsahu dátových zdrojov.

Obrázok 6: Analýza dostupných zdrojov



Povinná osoba v rámci identifikovania dostupných zdrojov uskutoční nasledujúce kroky:

- Analýzu organizačnej štruktúry povinnej osoby a jej bežných agend:
 - Identifikovanie aktivít jednotlivých organizačných jednotiek, ktoré sa zaoberajú zberom, tvorením alebo správou dát, ktoré majú potenciál na tvorbu datasetu.
 - Vytvorenie zoznamu potenciálnych datasetov, ktoré vznikajú pri identifikovaných aktivitách.
- Analýzu výročných správ povinnej osoby a ďalších verejne publikovaných dokumentov (vrátane informácií publikovaných na webovom portáli povinnej osoby), ktoré informujú o aktivitách a výsledkoch fungovania povinnej osoby:
 - Identifikovanie tabuliek a grafov v dokumentoch, ktoré informujú o potenciálnych datasetoch.
 - Zistenie, ktoré organizačné jednotky pripravujú tieto dokumenty.
 - Pridanie identifikovaných potenciálnych datasetov do zoznamu.
- Identifikovanie informačných systémov v organizácii:
 - Identifikovanie potenciálnych datasetov spravovaných informačnými systémami.
 - Pridanie týchto potenciálnych datasetov do zoznamu.
- Analýza požiadaviek verejnosti na zverejnenie informácií povinnou osobou:
 - Identifikovanie datasetov, ktoré by mohli byť zaujímavé pre verejnosť a porovnanie ich s datasetmi identifikovanými v predošlých krokoch.
 - Pridanie nových potenciálnych datasetov do zoznamu, pričom povinná osoba zohľadní aj výzvy dátového kurátora, ktorý hodnotí žiadosti verejnosti na zverejnenie nového datasetu.
- **Vytvorenie mapy datasetov**, ktorá zobrazuje organizačné jednotky a datasety, za ktoré sú zodpovedné. Mapa datasetov by mala byť zobrazená graficky a sprístupnená ako dataset.
- Diskusia nad mapou datasetov a identifikovanie datasetov so zodpovednými kontaktnými osobami.

Pre každý identifikovaný dataset zachytený v zozname by si povinná osoba mala zaznamenať nasledujúce informácie:

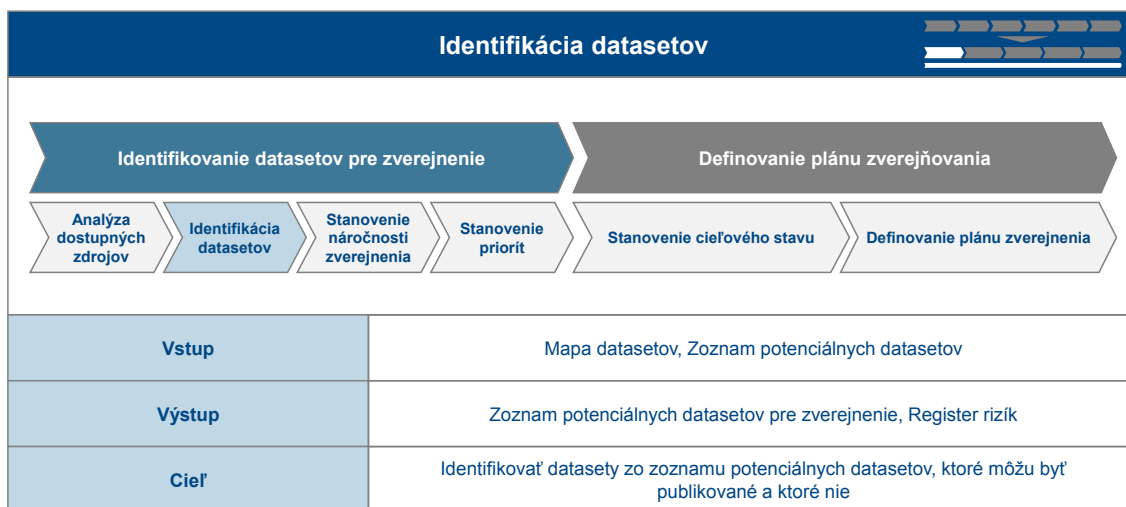
- Názov datasetu a jeho popis.
- Zodpovednú organizačnú jednotku povinnej osoby.

- Kontaktnú osobu na konzultácie o datasete.
- Súčasný formát dát (spôsob a formát uloženia datasetu) a krátky popis jeho formátu.

6.1.2 Identifikovanie datasetov na zverejnenie

Po analýze dostupných zdrojov prejde povinná osoba na tvorbu zoznamu potenciálnych datasetov na zverejnenie. Mapa datasetov, ktorá bola vytvorená v predchádzajúcom kroku a ktorá znázorňuje graficky prepojenie dátových zdrojov, datasetov a zodpovedných organizačných jednotiek, a zoznam všetkých potenciálnych datasetov. Slúžia ako vstup do procesu identifikácie datasetov vhodných na zverejnenie.

Obrázok 7: Identifikácia datasetov



Obsah datasetov, ktoré by mali byť zverejnené, musí byť v súlade so:

- Strategickou víziou pre postup zverejňovania datasetov – strategické rozhodnutie vlastníka datasetu môže definovať, ktoré sú primárne datasety na zverejnenie.
- Legislatívou – viac kapitola 4.2.
- Akčným plánom iniciatívy na otvorené vládnutie v Slovenskej republike.

Príkladom zaujímavých datasetov pre vybrané inštitúcie možno vidieť na obrázku nižšie.

Obrázok 8: Príklady datasetov vybraných inštitúcií

Vybrané inštitúcie a ich datasety	
NASES	■ Informácie o používaní služieb verejnej správy ■ Katalóg služieb verejnej správy a platby za služby verejnej správy (IS PEP) ■ Katalóg rolí vo verejnej správe (ÚPVS)
MV SR	■ Informácie z registrov a evidencií MV SR ■ Informácie o nehodovosti a priestupkoch
FR SR	■ Informácie z registra colných a daňových subjektov
ÚVO	■ Údaje o verejných zákazkách
ŠÚ SR	■ Informácie zo sčítania obyvateľov, domov a bytov ■ Ekonomické datasety
MD	■ Datasety o stave dopravných komunikácií (zmeny značiek, zákazy vjazdu a iné) ■ Cestovné poriadky
MS SR	■ Obchodný register
Samospráva	■ Zmluvy

Pre každý dataset v zozname potenciálnych datasetov je nutné uviesť možnosť zverejnenia alebo nezverejnenia jeho údajov v podobe otvorených dát.

V prípade možného zverejnenia je do Zoznamu potenciálnych datasetov nevyhnutné uviesť dôvody na uprednostnenie sprístupnenia daného datasetu alebo dôvody. Dôvody na uprednostnenie sprístupnenia datasetu sa budú pravdepodobne líšiť prípad od prípadu, ale vo väčšine prípadov možno uviesť ako hlavnú motiváciu na zverejnenie datasetov nasledujúce dôvody:¹⁰

- Zvýšenie transparentnosti.
- Stimulovanie ekonomického rastu.
- Zlepšenie vládnych služieb.
- Podpora opakovaného použitia dát a komúní.
- Zlepšenie verejných vzťahov a postoja k štátnej správe.
- Zlepšenie vládnych dát a procesov.

Datasety, ktoré sú v súlade s motiváciou na zverejnenie dát, sú požadované od komunity používateľov údajov a ktoré pomáhajú na dosiahnutie strategických cieľov, sú dobrými kandidátmi na **prioritné** zverejnenie. Napríklad informácie o rozpočtoch, obstarávaní a zmluvách môžu prispieť k zvýšeniu transparentnosti. Informácie o výkonoch orgánov verejnej správy môžu pomôcť k zvýšeniu dôvery a zlepšiť pohľad na verejnú správu. Obrázok nižšie ponúka príklady 3 najštáhovanejších datasetov v troch krajinách (USA, UK a Francúzsko) na základe benchmarku Capgemini.

¹⁰ Logica Business Consulting (2012).

Obrázok 9: Top 3 najst'ahovanejších datasetov v troch krajinách (USA, UK a Francúzsko) na základe benchmarku Capgemini

Štát	Názov datasetu	Popis	Inštitúcia	Kategória
USA	NOAA National Weather Service - National Mosaic of Weather Radar	Výstup z radaru umožňujúci interaktívne zobrazovanie počasia	National Oceanic and Atmospheric Administration, Department of Commerce	Agrikultúra
	Job Openings and Labor Turnover Survey	Informácie o otvorených pracovných pozíciách a novo prijatých zamestnancoch	Department of Labor	Spotrebiteľ, financie
	Consumer Complaint Database	Informácie o doručených sťažnostiach na poskytnuté o finančné produkty a služby	Consumer Financial Protection Bureau	Financie
UK	Live traffic information from the Highways Agency	Aktuálne dopravné informácie pre strategickú cestnú sieť v Anglicku, operovanú Highways Agency	Highways Agency	Doprava
	Learning Aim Reference Service	Dataset ponúka nástroj pre rýchle hľadanie, ktorý umožňuje užívateľom vyhľadávať podľa najčastejšie hľadaných výrazov	Skills Funding Agency	Vzdelávanie
	Bona Vacantia Unclaimed Estates and Adverts	Dataset inzeruje majetok zosnulých osôb pre identifikáciu príbuznými na uplatnenie vlastníckeho nároku	Treasury Solicitor's Department	Spoločnosť
Francúzsko	Population	Dataset poskytuje prístup k výsledkom sčítania ľudu	Institut National de la Statistique et des Etudes Economiques (INSEE)	Populácia
	Liste des Immeubles protégés au titre des Monuments Historiques	Zoznam chránených historických pamiatok	Ministère de la Culture et de la Communication	Spoločnosť
	Recensement des équipements sportifs, espaces et sites de pratiques	Národné sčítanie všetkých športových zariadení, priestorov a tréningových miest	Ministère des droits des femmes, de la ville, de la jeunesse et des sports	Spoločnosť

V prípade nezverejnenia datasetu musí byť toto rozhodnutie učené na základe analýzy rizík a bariér na zverejnenie. Môže ísť napríklad o analýzu možného úniku osobných údajov alebo utajovaných skutočností, ktoré môžu byť odkryté po skombinovaní viacerých datasetov, pričom z každého datasetu boli takéto údaje odstránené. Keďže ide o komplexný rozhodovací proces o možnosti publikovania, odporúča sa uskutočniť formou workshopu s prítomnosťou vlastníkov datasetov, právnikov, dátového kurátora a ďalších stakeholderov.

Príkladom dôvodu na nepublikovanie môže byť:

- Zverejnenie osobných údajov alebo utajovaných skutočností (platí aj pre kombináciu viacerých datasetov podľa kapitoly 6.1.2.1).
- Náklady na zverejnenie násobne prevyšujú prínosy.
- Nízky potenciál na opakované použitie.

Do diskusií o zverejnení je nevyhnutné zapojiť aj názory komunity využívajúcej otvorené dáta, aby sa znížilo riziko, že zverejnené datasety nebudú využívané. Na zber spätnej väzby slúži predovšetkým portál data.gov.sk a využíva sa v procese neustáleho zlepšenia kvality a ponuky datasetov. Spätňú väzbu na používanie otvorených dát a dopyt po nových datasetoch možno získať prostredníctvom:

- Prieskumov a ankiet.
- Online hlasovania za datasety zo zoznamu predvybraných kandidátov.
- Workshopov, napríklad v priestoroch, kde pôsobia start-upy.
- Verejných diskusií s odbornou komunitou.
- Konferencií.

Odporúča sa tiež sledovať výstupy medzinárodných štúdií ako the Open Data Index¹¹ alebo the Open Data Barometer.¹²

Odfiltrovaním datasetov nevhodných na publikovanie zo Zoznamu potenciálnych datasetov dostaneme Zoznam potenciálnych datasetov na zverejnenie.

¹¹ <https://index.okfn.org>

¹² <http://www.opendataresearch.org/project/2013/odb>

6.1.2.1 Údaje vylúčené z povinnosti zverejňovania

Základná kategorizácia údajov pri rozhodovaní o vhodnosti ich zverejnenia rozlišuje dve úrovne:

- **Zverejniteľné údaje (Z)** – údaje, ktorých zverejnenie neohrozuje fungovanie štátu a jeho systémov a neprezrádza utajované skutočnosti, obchodné tajomstvo alebo informácie o osobnosti a súkromí fyzickej osoby. Tieto údaje je preto možné kedykoľvek zverejniť.
- **Nezverejniteľné údaje (NZ)** – údaje, ktoré nie je vhodné zverejniť v žiadnom prípade, pretože zverejnenie nesie riziko okamžitého alebo neskoršieho pokusu o narušenie informačnej bezpečnosti, súkromia alebo obchodného tajomstva. Pre nezverejniteľné údaje môže existovať podmienka, za splnenia ktorej sa stanú zverejniteľnými. Napríklad osobné údaje možno zverejniť so súhlasom dotknutej osoby alebo ak to ustanovuje osobitný zákon. V opačnom prípade musí prebehnúť anonymizácia údajov, ako je uvedené v časti 8.1. Obchodné tajomstvo sa napríklad netýka informácií o znečisťovaní životného prostredia alebo o poskytnutí štátnej pomoci.

O spomenutých typoch údajov hovoria paragrafy 8, 9, 10 a 11 zákona č. 211/2000. Táto kategorizácia, samozrejme, môže byť vo vnútornom prostredí organizácie ďalej rozšírená, napr. podľa bezpečnostnej politiky organizácie (napr. na chránené, služobné).

Ochrana osobných údajov je problematikou, ktorou je potrebné sa zaoberať aj v rámci zverejňovania otvorených údajov. Zákon č. 428/2002 Z. z. neupravuje konkrétny zoznam údajov, ktoré sú považované za osobné údaje; v súlade s ustanovením § 3 poskytuje demonštratívny výpočet charakteristík určujúcich fyzickú osobu. V zmysle tohto ustanovenia je potrebné individuálne a v každom prípade jednotlivito taktiež rozlišovať, či rôzne údaje, napríklad o IKT, majú zároveň aj charakter osobného údaju v zmysle uvedeného § 3 zákona č. 428/2002 Z. z. Rovnako je dôležité aj to, že po priradení určitej informácie k osobnému údaju, ktorá sama osebe nemá charakter osobného údaju, sa takáto informácia môže stať tiež osobným údajom, ak vedie k lepšej identifikácii konkrétnej fyzickej osoby, napr. údaje o používanom softvéri sa priradia ku konkrétnemu zamestnancovi orgánu verejnej správy identifikovanému menom, priezviskom a zamestnaneckou príslušnosťou k danému orgánu ako svojmu zamestnávateľovi. Podobná situácia môže nastať v prípade kombinovania viacerých datasetov.

Osobné údaje a zároveň údaje o IKT sú, resp. môžu byť napr.: prihlasovacie meno užívateľa informačného systému, logy (prihlasovanie/odhlasovanie užívateľa), rôzne súbory na pamäťových médiách, vlastné údaje databáz a registrov (t. j. používateľské údaje), nastavenie prístupových práv používateľov informačných systémov, ako aj ďalšie IKT, ktoré môžu mať určitú vypovedaciu schopnosť vedúcu k identifikácii konkrétnej fyzickej osoby. Konkrétne posúdenie citlivosti údajov možno nájsť v Metodickom pokyne pre kategorizáciu citlivosti údajov z dôvodu bezpečnosti.

Podmienky na zverejniteľnosť údajov môžu byť rôzne a po odbornom zvážení by o nich mala rozhodovať príslušná organizácia. V prípade, že má organizácia zavedenú adekvátnu a funkčnú organizáciu bezpečnosti (vyplývajúcu z bezpečnostnej politiky), o posudzovaní zverejniteľnosti by mala rozhodovať niektorá z bezpečnostných rolí celého systému (napr. bezpečnostný manažér alebo metodik bezpečnosti). Podmienky zverejniteľnosti môžu byť najmä nasledujúce:

- uplynul určitý čas (uvedené kritérium je potrebné používať s rozvahou, pretože nemusí znamenať, že riziko ohrozenia pominulo),
- údaj už nie je relevantný a existujúce skutočnosti a systémy neohrozí,
- údaj sa týka nefunkčných („neživých“) alebo nepoužívaných systémov atď.

Pri zverejňovaní otvorených údajov je ďalej potrebné dbať na ochranu utajovaných skutočností. Právnym základom ochrany utajovaných skutočností v Slovenskej republike je zákon č. 215/2004 Z. z. o ochrane utajovaných skutočností a o zmene a doplnení niektorých zákonov v znení neskorších predpisov. Uvedená platná právna úprava definuje konkrétny podmienky na ochranu utajovaných skutočností, práva a povinnosti právnických osôb a fyzických osôb pri tejto ochrane, pôsobnosť Národného bezpečnostného úradu a pôsobnosť ďalších štátnych orgánov a obcí vo vzťahu k utajovaným skutočnostiam vrátane uplatňovania administratívnej zodpovednosti za porušenie povinností vo vzťahu k fyzickým, ako aj k právnickým osobám.

Utajované skutočnosti je potrebné chrániť pred nepovolanou osobou a pred cudzou mocou v súlade s uvedenou platnou právnou úpravou. Preto je nevyhnutné vykonať analýzu zverejnených datasetov a ich kombinácií, či nedôjde k zverejneniu utajovaných skutočností. Odporúča sa rozpoznať datasety, ktoré sa týkajú príbuzných tém, ktoré sú predmetom utajovaných skutočností organizácie (napríklad umiestnenie kritickej infraštruktúry štátu) a definovať možné riziká spojené s ich zverejňovaním v registri rizík popísanom v nasledujúcej časti.

6.1.2.2 Odporúčaný register rizík

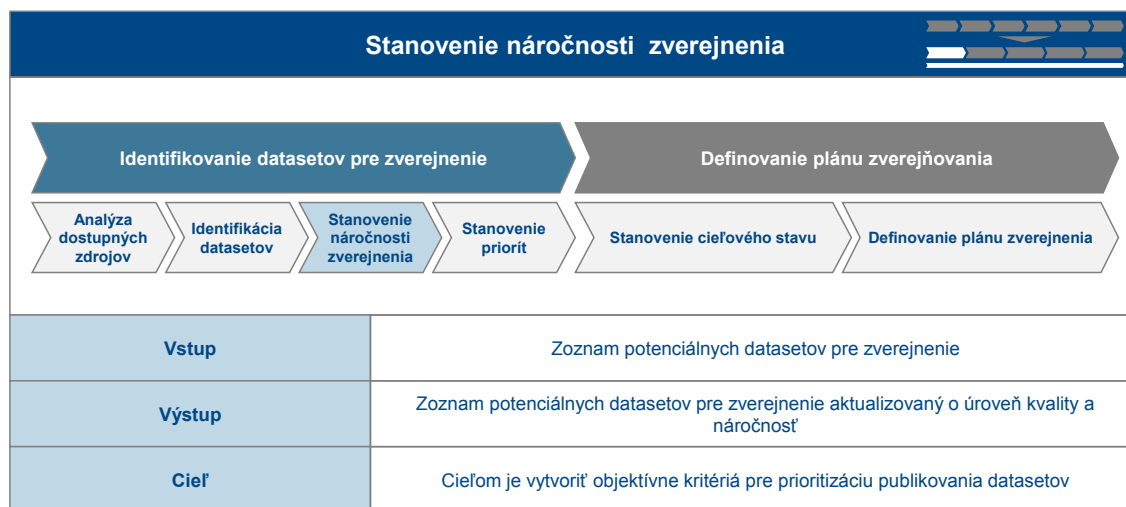
Odporúča sa vykonať analýzu rizík zverejnenia datasetu a každé riziko zaznamenať do odporúčaného **registra rizík**, ktorý obsahuje:

- ID riziko.
- Meno a popis rizika.
- Publikované datasety, na ktoré sa vzťahuje.
- Popis možného dosahu rizika.
- Informáciu o pravdepodobnosti výskytu.
- Mitigačnú stratégiu.

6.1.3 Stanovenie náročnosti zverejnenia jednotlivých datasetov

Stanovenie náročnosti zverejnenia slúži na vyhodnotenie, či sú identifikované datasety vhodné na zverejnenie a aká bude náročnosť ich zverejnenia.

Obrázok 10: Stanovenie náročnosti zverejnenia



Prvým krokom pri posudzovaní náročnosti zverejnenia je využitie nástroja Five Star hodnotenia vyplývajúce z výnosu o štandardoch pre informačné systémy verejnej správy.¹³ Tento nástroj slúži na určenie výslednej úrovne kvality pri zverejnení.

¹³ Výnos č. 55/2014 Z. z. Ministerstva financií Slovenskej republiky o štandardoch pre informačné systémy verejnej správy.

Obrázok 11: Five Star hodnotenie

Five Star hodnotenie					
Úroveň	Dostupnosť na internete	Štruktúrovaný obsah	Otvorený formát	Referencovateľné identifikátory	LinkedData
0	X	X	X	X	X
1	✓	X	X	X	X
2	✓	✓	X	X	X
3	✓	✓	✓	X	X
4	✓	✓	✓	✓	X
5	✓	✓	✓	✓	✓

Úroveň kvality datasetu poskytovaného povinnou osobou môže dosahovať jednu zo šesť úrovní, ktorými sú:

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.

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 na publikovanie podľa štandardov pre ISVS

- Dataset je dostupný vo webovom prostredí.
- Obsah datasetu je štruktúrovaný tak, že umožňuje automatizované spracovanie.
- Dataset je poskytovaný v otvorenom formáte, nezávislom od konkrétneho proprietárneho softvéru.
- Povinné 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 od konkrétnom proprietárneho softvéru.

- Na identifikáciu entít datasetu sa používajú referencovateľné identifikátory v podobe URL, čím je možné na tieto entity transparentne odkazovať z ostatných datasetov.
- Odporúčané formáty sú: RDF, CSV, XML alebo JSON.
- Spôsoby zaznamenania referencovateľného identifikátora:
 - Pre CSV sa vytvorí nový stĺpec pre URL entity. Tento stĺpec sa umiestni vedľa existujúceho stĺpca pre identifikátory entity.
 - Pre XML sa odporúča využiť na zaznamenanie URL rozšírenie HTML a XML dokumentov nazývané RDFa6¹⁴ (XML atribút *resource*).
 - Pre JSON sa odporúča využiť na zaznamenanie URL rozšírenie JSON nazývané JSON-LD7¹⁵ (konštrukcia @id).

Úroveň 5*

- Dataset je dostupný vo webovom prostredí.
- Obsah datasetu je štruktúrovaný tak, že umožňuje automatizované spracovanie.
- Dataset je poskytovaný v otvorenom formáte, nezávislom od konkrétneho proprietárneho softvéru.
- Na identifikáciu entít datasetu sa používajú referencovateľné identifikátory.
- Navyše sú prelinkované referencovateľnými identifikátormi aj iné súvisiace entity v iných datasetoch – ide tak o formát otvorených dát „LinkedData“, v ktorom sú URL zaznamenané ideálne pomocou štandardu TTL.¹⁶
- Každá URL publikovaných entít je dereferencovateľná, to znamená, že aplikácia získa prístupom na danú URL strojovo čitateľnú reprezentáciu entity v RDF modeli.

Minimálna úroveň kvality datasetu otvorených dát, ktoré musí povinná osoba publikovať, je úroveň 3*.

Náročnosť zverejnenia sa počíta pre každý potenciálny dataset, aby ju bolo možné porovnávať medzi jednotlivými datasetmi. Otvplyvňujú ju nasledujúce parametre:

- **Zložitosť datasetu**, ktorá vyjadruje, ako bude zložitá transformovať dataset zo súčasného formátu do cieľového formátu na zverejnenie. Z toho vyplýva, že čím menej zmien je potrebných, tým nižšia bude náročnosť zverejnenia.
- **Anonymizácia** – ak je nevyhnutné dataset pred publikovaním anonymizovať, zvyšuje sa náročnosť zverejnenia. Ak aj anonymizácia prebieha automatizovane, musí sa jej výsledok skontrolovať, aby sa predišlo riziku narušenia súkromia alebo úniku obchodného tajomstva či utajovaných skutočností, čo môže vyžadovať viacero manuálnych operácií.
- **Rozsah manuálnych operácií** určuje nutný čas odborných zamestnancov venovaný ich výkonu – manuálne operácie môžu byť spojené s anonymizáciou alebo transformáciou datasetov či kontrolou výsledkov automatizovaných nástrojov.
- **Veľkosť datasetu** – čím väčší dataset, tým časovo náročnejšie sú všetky manuálne operácie nad ním a zdrojovo náročnejšie je jeho publikovanie, ukladanie a aktualizovanie.
- **Cieľová periodičita zverejnenia** ovplyvňuje celkový odhad náročnosti na dané časové obdobie, pretože určuje, koľkokrát sa úkony nevyhnutné na prípravu a publikovanie datasetu budú opakovať. Odhadovaná náročnosť úkonov sa preto prenášobí počtom opakovaní pre dané časové obdobie. Avšak v niektorých prípadoch sa náročnosť opakovaných úkonov počas spravovania datasetu znižuje vďaka zvyšovaniu efektivity

¹⁴ <http://www.w3.org/TR/rdfa-primer/>

¹⁵ <http://www.w3.org/TR/json-ld/>

¹⁶ <http://www.w3.org/TR/turtle/>

úkonov. Na zohľadnenie tejto zvýšenej efektivity sa môžu do vzorca pridať príslušné koeficienty.

Výsledná náročnosť zverejnenia sa v nástroji určí z nasledujúceho rozsahu:

0. Náklady prevyšujúce prínos zverejnenia.

1. Veľmi vysoké náklady.

2. Vysoké náklady.

3. Stredné náklady.

4. Nízke náklady.

5. Veľmi nízke náklady.

6.1.3.1 Vytváranie referencovateľných identifikátorov

Datasety obsahujú entity vo forme:

- Riadkov v tabuľke vo formáte CSV.
- Prvkov v dokumente XML.
- Objektov v dokumente JSON.

Je veľmi dôležité definovať identifikátor pre každú entitu. Tieto identifikátory sú dôležité pre vývojárov aplikácií nad otvorenými dátami. Používajú ich na identifikovanie entít datasetu v zdrojovom kóde a na zlúčenie informácií o entitách z rôznych dátových zdrojov. Štandardom pre referencovateľný identifikátor je používanie nasledujúcej schémy: {základná Unified Resource Identifier (URI) } / {zdrojová cesta} / {typ} / {trieda} / {podtrieda 1 / podtrieda 2 ...} / {referencia}, pričom:

- základnú Unified Resource Identifier (URI) tvorí `http://`,
- zdrojovú cestu tvorí `data.gov.sk`,
- {typ} je reťazec, ktorý je založený na názve triedy, do ktorej entita patrí. Ak entita patrí do viacerých tried, odporúča sa vybrať tú najšpecifickejšiu. Typ sa odporúča písať malými písmenami, pričom jednotlivé slová sú oddelené pomlčkou -:
 - `id` – ak sa označuje neinformačný zdroj, ktorým je entita skutočného sveta,
 - `doc` – ak sa identifikuje dokument,
 - `def` – ak sa identifikuje definícia konkrétneho konceptu,
 - `set` – ak sa identifikuje súbor údajov.
- {triedu} tvorí slovo alebo reťazec, ktorý zachytáva podstatu identifikovanej entity skutočného sveta, napríklad škola, a to podľa číselníka taxonómií vydávaných ministerstvom financií,
- {podtriedu} tvorí slovo alebo reťazec sekundárnej klasifikácie triedy, ak je to potrebné,
- referenciu tvorí reťazec, ktorý sa používa na identifikáciu jednotlivých inštancií konceptu, obvykle v tvare kódu z číselníka.

Na zadefinovanie identifikátora je potrebné riadiť sa nasledujúcimi pravidlami:

- Každá entita by mala mať identifikátor, ktorý sa skladá len z jedného atribútu entity.
- Ako referencia by sa nemala používať umelo vygenerovaná hodnota bez významu uložená ako primárny kľúč v databáze. Malo by ísť o hodnotu, ktorá sa používa na výmenu informácií o entite v reálnom svete a v reálnych systémoch. Napríklad podnik v danej krajine môže byť identifikovaný cez IČO, pretože ide o číslo, ktoré sa tiež používa v informačných systémoch verejnej správy na identifikáciu a výmenu informácií.
- Iba ak neexistuje žiadna referencia v reálnom svete, možno vygenerovať novú aj bez významu.

- Každá referencia, trieda a podtrieda musia byť popísané v dátovej schéme datasetu. Spôsob popisu závisí od skladby jazyka danej schémy.
- Na vytváranie triedy alebo podtriedy či referencie sa používa anglický jazyk, iba v nevyhnutných prípadoch slovenský.

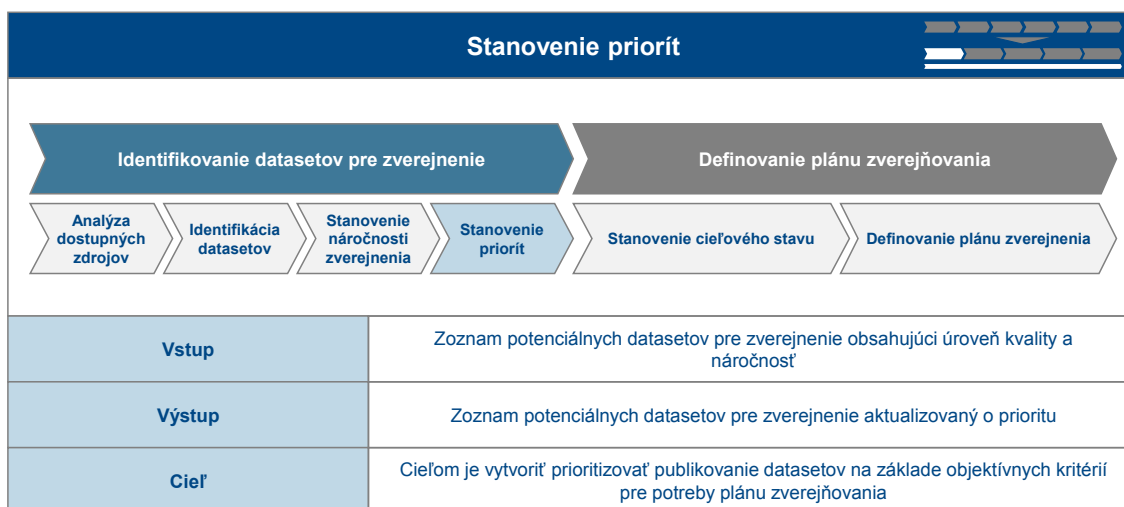
Ak sú datasety publikované na úrovni 4* alebo 5*, musia mať jednotlivé entity referencovateľné identifikátory v podobe URL, na ktorých vytvorenie sa používa uvedený identifikátor a ktorý spĺňa nasledujúce pravidlá:

1. Používa sa v strojovo čitateľnom formáte entity ako jej referencovateľný identifikátor.
2. Nemení sa počas celého životného cyklu entity.
3. Používa sa na prelinkovanie entity z iných súvisiacich entít v rovnakom alebo inom datasete tej istej alebo inej organizácie. Je podobná URL webovej stránky, ktorá sa tiež používa na prelinkovanie iných webových stránok na danú stránku.
4. V prípade úrovne 5* je URL dereferencovateľná. To znamená, že ak sa do webového klienta zadá URL cez http protokol, vráti sa zo serveru strojovo čitateľná reprezentácia entity v RDF formáte serializovaná do vhodného formátu (napríklad TTL alebo JSON-LD).

6.1.4 Stanovenie priorít na zverejnenie datasetov

Po zmapovaní datasetov a stanovení náročnosti zverejnenie si povinná osoba stanoví priority na zverejnenie datasetov. Postup uvedený v tejto kapitole je odporúčaný a povinná osoba si ho môže prispôbiť svojim potrebám a prípadne aj zjednodušiť.

Obrázok 12: Stanovenie priorít



Datasety s vysokou pridanou hodnotou, maximálne strednými nákladmi na zverejnenie, a s vysokou hodnotou na opakované použitie by mali byť prioritné datasety na zverejnenie. Príkladom takého datasetu sú napríklad dáta, ktoré môžu byť opakovane použité prostredníctvom aplikácií na cestovanie verejnou dopravou. Dáta, ktoré pomáhajú obyvateľom lepšie lokalizovať a využívať verejné služby, by mali byť zverejnené na účely zlepšenia fungovania verejnej správy. Dáta popisujúce vyžívanie služieb verejnej správy a poskytujúce informácie o poskytovateľoch údajov a podmienkach na využitie služieb sú príkladom z tejto kategórie.

Priority publikovania otvorených dát

Priority publikovania otvorených dát môžu byť stanovené na základe týchto krokov:

1. Stanovenie kľúčových kritérií pre prioritizáciu datasetov.

2. Stanovenie váh prioritizačných kritérií.
3. Odstránenie datasetov, ktoré nemôžu byť zverejnené v podobe otvorených dát z katalógu potenciálnych datasetov.
4. Kalkulácia priorít na základe kritérií a ich váh.
5. Zoradenie potenciálnych datasetov na základe kalkulácií.
6. Upravenie poradí datasetov na základe reflektovania cieľov na publikovanie otvorených dát a skutočnej hodnoty datasetu.
7. Prípadná úprava váh jednotlivých kritérií a rekalkulácia na základe nových výsledkov.
8. Finálna selekcia datasetov na publikovanie a stanovenie času releasu datasetu.

Prioritizačné kritériá

Návrh¹⁷ prioritizačných kritérií pre povinnú osobu:

- Dopyt po dátach, ktorý možno určiť na základe prieskumu alebo vyhodnotením žiadostí o sprístupnenie súvisiacich datasetov v danej téme.
- Súčasný formát dát – ak sa dáta už nachádzajú v strojovo čitateľnom formáte, bude oveľa jednoduchšie ich zverejniť.
- Dostupná dokumentácia dátovej schémy, ktorá zvýši potenciál datasetu na jeho opakované použitie.
- Prínosy datasetu – datasety s vyšším odhadovaným prínosom by mali byť zverejnené prioritne.
- Riziko publikovania datasetu – menej rizikové datasety sú preferované.
- Odhadovaná náročnosť na zverejnenie - datasety, ktoré sú jednoducho publikovateľné a zároveň majú vysokú hodnotu, musia byť publikované ako prvé.

Publikovanie datasetov vo formáte LinkedData

Ak sa povinná osoba rozhodne publikovať datasety vo formáte LinkedData, musia byť navyše využité nasledujúce prioritizačné kritériá:

- Referencovateľné identifikátory – Datasety s prirodzenými referenciami v identifikátoroch sú najlepšími kandidátmi na publikovanie vo formáte LinkedData, pretože nie je nevyhnutné sa dohodnúť medzi rôznymi stranami na tom, ako identifikovať entity.
- Potenciál pre Linked Data – Odhad počtu datasetov, ktoré môžu byť prepojené s daným datasetom.

Ak sú dôvodom na nízku prioritu publikovania cena alebo náročnosť publikovania, možno zvážiť nasledujúce:

- Znížiť úroveň kvality pre Open Data na minimálnu možnú úroveň 3★.
- V prípade, že je dataset príliš komplexný alebo nákladný na zverejnenie aj na úrovni 3★ alebo vyššej, povinná osoba zváži zverejnenie datasetu v ďalších cykloch a nastaví nápravné kroky (napríklad v oblasti rozvoja príslušného agendového systému), aby bolo možné dataset poskytovať neskôr menej nákladne na uspokojivej úrovni.
- Zvážiť, či sa dataset nezverejňuje už inou povinnou osobou, alebo či nie je súčasťou iného väčšieho datasetu, na ktorý sa možno len prelinkovať.
- Ak je zverejnená inou povinnou osobou len časť datasetu, možno rozdeliť dataset na dve časti a publikovať len primárne dáta prelinkované s externým datasetom. Nevyhnutné linky

¹⁷ Prioritizačné kritériá môžu byť upravené podľa potrieb a špecifik jednotlivých povinných osôb.

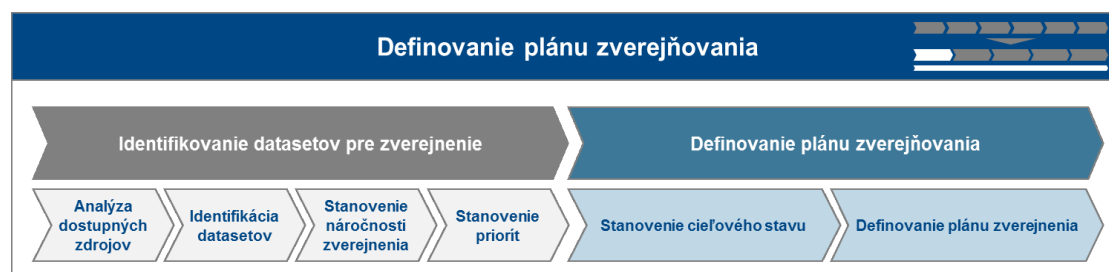
musia byť súčasťou záznamu v katalógu alebo, ak sa už uplatňujú princípy LinkedData (úroveň kvality 5★), je možné prelinkovať v datasetoch napriamo.

V niektorých prípadoch sa môže stať, že dataset je už publikovaný v rámci väčšieho datasetu. V tomto prípade nie je nutné znovu zverejňovať spomínaný dataset, ale namiesto toho stačí na neho odkázať. Z tohto dôvodu by mala povinná osoba skontrolovať, či nie je dataset niekde inde zverejnený, aby nedochádzalo k duplikovaniu práce. V prípade, ak je dataset zverejnený iba čiastočne, povinná osoba zanalyzuje, či by nebolo vhodné rozdeliť dataset a publikovať iba primárne dáta prepojené na externú databázu. Požadované odkazy na prepojenie by mali byť uvedené v katalógu, ale v prípade aplikácie LinkedData pravidiel je možné prepojenie objektov priamo v datasete.

6.2 Definovanie plánu zverejňovania

Definovanie plánu zverejňovania sa skladá z dvoch krokov: stanovenie cieľového stavu a definovanie plánu zverejnenia otvorených dát na dosiahnutie tohto cieľového stavu.

Obrázok 13: Definovanie plánu zverejňovania



Plán zverejňovania otvorených dát predstavuje finálny výstup prvej analytickej fázy publikácie otvorených dát. Tento plán dokumentuje aj ciele zverejňovania, ktoré musia byť v súlade so strategickými cieľmi vlastníka dát a s platnou legislatívou. Príklady na možné odôvodnenie prioritného zverejnenia alebo nezverejnenia sa nachádzajú v kapitole 6.1.2. Motivácia môže byť daná národnou stratégiou ako Akčným plánom iniciatívy pre otvorené vládnutie v Slovenskej republike a platnou legislatívou, ktorá smeruje k zverejňovaniu všetkých dát verejnej správy, ktoré neobsahujú:

- osobné údaje a ochranu osobnosti,
- utajované skutočnosti,
- obchodné tajomstvo,
- intelektuálne vlastníctvo tretích strán.

6.2.1 Stanovenie cieľového stavu

Stanovenie cieľového stavu je založené na predchádzajúcich aktivitách, v ktorých sa získané informácie skonsolidujú do finálneho katalógu potenciálnych datasetov na publikovanie v podobe otvorených dát. Možno vykonať poslednú revíziu priorit datasetov pri komplexnej kontrole, či daná množina datasetov v katalógu napĺňa strategické ciele a legislatívne požiadavky definované v § 3 ods. 4 písm. d) zákona č. 275/2006 Z. z. a v § 6 ods. 3 zákona č. 211/2000 Z. z.

Obrázok 14: Stanovenie cieľového stavu



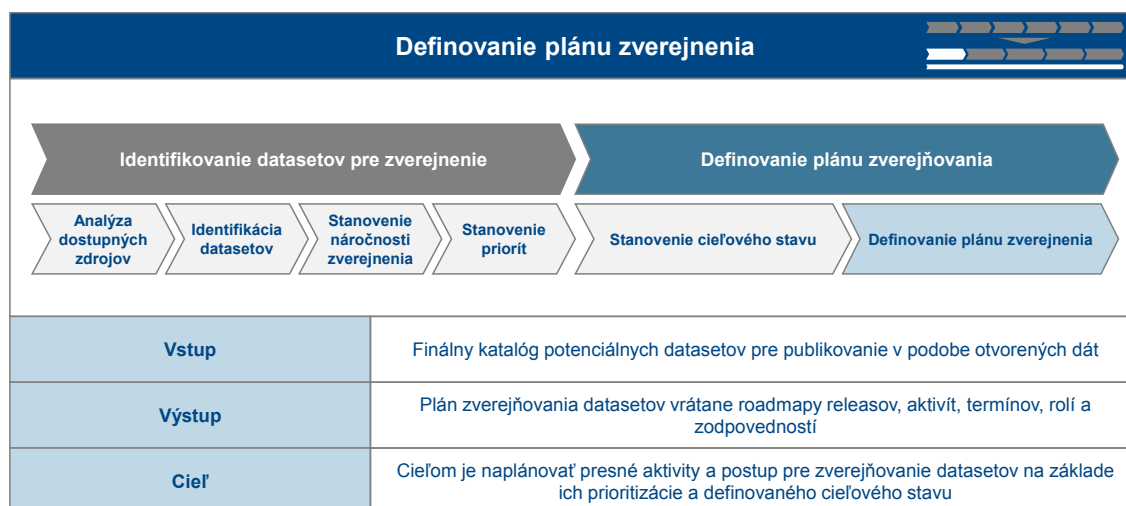
V cieľovom stave by mali byť publikované všetky datasety, pri ktorých nie je dôvod na ich nezverejnenie. Pri každom datasete by taktiež mala byť uvedená náročnosť zverejnenia a kvalita datasetu odvodená od Five Star hodnotenia.

Cieľový stav zachytáva ciele publikovania otvorených dát povinnej osoby a mali by byť v súlade s cieľmi vlastníka dát. Ciele publikovania otvorených dát reflektujú potreby kreatívnych komún na vznik nových foriem podnikania, zámery na zvyšovanie transparentnosti a na podporu rozhodovania vo verejnej správe na základe dát. Možné motivácie na publikovanie otvorených dát sú popísané v kapitole 6.1.2 Identifikovanie datasetov na zverejnenie.

6.2.2 Definovanie plánu zverejnenia otvorených dát

Ak nie je možné publikovať všetky potenciálne datasety v danom okamihu, publikovanie datasetov by malo byť rozdelené do niekoľko releasov v rámci viacerých iterácií periodického publikovania v súlade so stanovenými prioritami.

Obrázok 15: Definovanie plánu zverejnenia



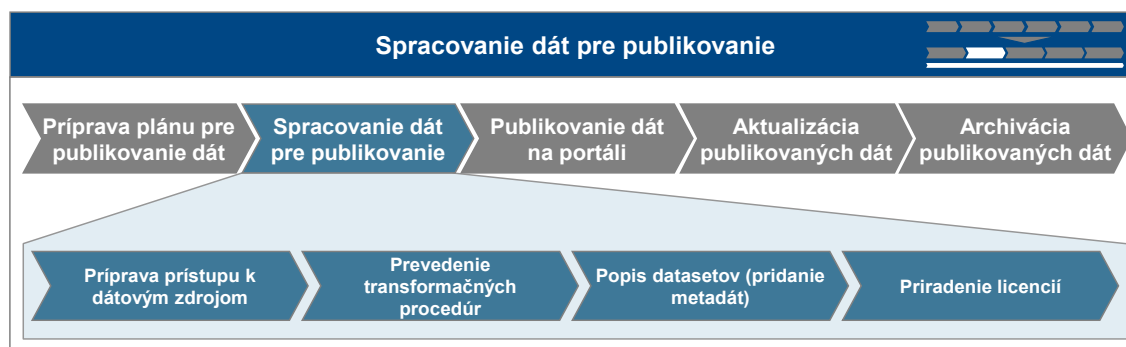
Na základe finálnej selekcie a priority publikovania datasetov si povinná osoba zostaví roadmapu zverejňovania otvorených dát. Táto roadmapa musí tiež obsahovať aktivity nevyhnutné na publikovanie otvorených dát spolu s termínmi ich vykonania. Plán publikovania zahŕňa tiež popis rolí a zodpovedností stakeholderov zapojených do publikovania otvorených dát (kapitola 5 Odporúčané organizačné zabezpečenie). Definovanie plánu na publikovanie

otvorených dát by malo byť koordinované na základe plánu riadenia prínosov a s ohľadom na plán mitigácie rizík. Plán identifikuje aj ďalšie úlohy, ktoré musí byť vykonané na publikovanie datasetov.

7 Spracovanie dát na publikovanie

Po stanovení plánu na publikovanie dát a cieľového stavu je nutné zdrojové dáta spracovať do vhodného formátu na publikovanie.

Obrázok 16: Spracovanie dát na publikovanie



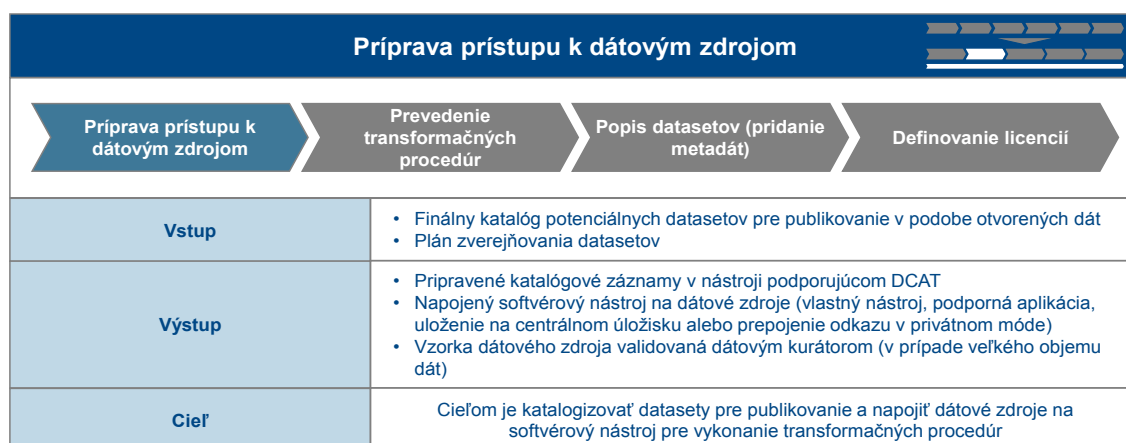
Spracovanie dát na publikovanie je vykonané v štyroch krokoch, ktoré sú detailne popísané v nasledujúcich podkapitolách:

- Príprava prístupu k dátovým zdrojom.
- Vyhodenie transformačných procedúr.
- Popis datasetov (pridanie metadát).
- Priradenie licencií.

7.1 Príprava prístupu k dátovým zdrojom

Príprava prístupu k dátovým zdrojom je prvým krokom spracovania dát na publikovanie. Dlhodobým zámerom je celý proces publikovania automatizovať.

Obrázok 17: Príprava prístupu k dátovým zdrojom



Dátové zdroje povinnej osoby môžu byť nasledujúceho typu:

- Databázový server prislúchajúci k danému informačnému systému, ktorý vytvára dáta určené na publikovanie.
- Dátový súbor uložený na lokálnom filesystéme - podporované formáty: CSV, XML, JSON, TXT, HTML, TIFF, JPG, PDF.

- Tabuľkové súbory a dokumenty nachádzajúce sa vo vzdialenom úložisku súborov povinnej osoby – úložisko týchto dát musí byť softvérovému nástroju prístupné cez SCP/SFTP protokol.

Vzhľadom na to, že je nutné dáta len extrahovať a publikovať, read-only prístup pre akýkoľvek nástroj pracujúci s databázou je dostatočný a odporúčený z bezpečnostných dôvodov.

Pre databázový server budú existovať tri možnosti prístupu cez vlastný softvérový nástroj, pomocou transformačného modulu po nahratí dát na centrálné úložisko (detailný postup je uvedený v používateľskej príručke) alebo pomocou podpornej aplikácie:

- Cez konto používateľa s právami len na čítanie, ktorým bude možné dáta priamo extrahovať.
- Pomocou databázových skriptov pre špecifické dátové zdroje, ktoré periodicky vytiahnu dáta z databáz do strojovo čitateľných formátov (CSV, XML alebo JSON) - SQL dotaz cez JDBC/ODBC priamo do databázy.

Pre dátové súbory nachádzajúce sa na lokálnom filesystéme alebo na vzdialenom úložisku bude možné dáta nahráť buď do podpornej aplikácie, alebo na centrálné úložisko centrálného katalógu (kde ich možno ďalej spracovať transformačným modulom podľa kapitoly nižšie), prípadne ich s centrálnym katalógom len prelinkovať (ak sú dáta zverejnené už u povinnej osoby). V tejto fáze sú nahrané alebo prelinkované dáta v stave private, to znamená, že ešte nie sú zverejnené.

Povinná osoba spustí nahrávanie údajov. Ak ide o import tabuľkových formátov ako je CSV, systém údaje uloží do relačného úložiska. Ak ide o súbory ako napr. PDF (naskenovaná faktúra), systém uloží údaje na file systém.

Ak veľkosť datasetu, ktorý má byť uložený na centrálnom úložisku, presiahne istú hranicu (závisiacu od aktuálneho vyťaženia úložisk), musí povinná osoba najskôr požiadať o nahratie cez formulár ÚPVS a priložiť vzorku nasledujúcim postupom:

1. Navštívi webové sídlo UPVS.
2. Zadá prihlasovacie údaje do systému. Systém na základe vyplnených prihlasovacích údajov autentifikuje a autorizuje poskytovateľa údajov.
3. Vyhľadá si stránku s elektronickým formulárom žiadosti a vyplní elektronický formulár metadátami dátového zdroja:
 - Názov datasetu, ktorému dátový zdroj patrí.
 - Názov dátového zdroja.
 - Periodicitu aktualizácie dátového zdroja.
 - Formát dátového zdroja.
 - Dôvod na uloženie dátového zdroja.
 - Licenciu (viac o licenciách v kapitole 7.4).
4. Vopred si pripraví dátovú vzorku z dát dátového zdroja, ktorú priloží k podaniu - Povinná osoba v zobrazenom formulári klikne na tlačidlo pri poli na nahrávanie súborov. Systém zobrazí prehliadač súborov lokálneho počítača povinnej osoby.
5. Naviguje sa v prehliadači k lokalite dátovej vzorky a vyberie požadované údaje. Povinná osoba v zobrazenom formulári klikne na tlačidlo pri poli na nahrávanie súborov a vyberie podľa krokov 2 a 3 súbor so schémou údajov.
6. Odošle formulár.
7. Systém validuje vyplnené údaje. V prípade správne vyplneného formulára žiadosť prepošle do agendového systému. V prípade nesprávne vyplnených údajov zobrazí poskytovateľovi údajov formulár a nesprávne vyplnené údaje, ktoré je potrebné opraviť.

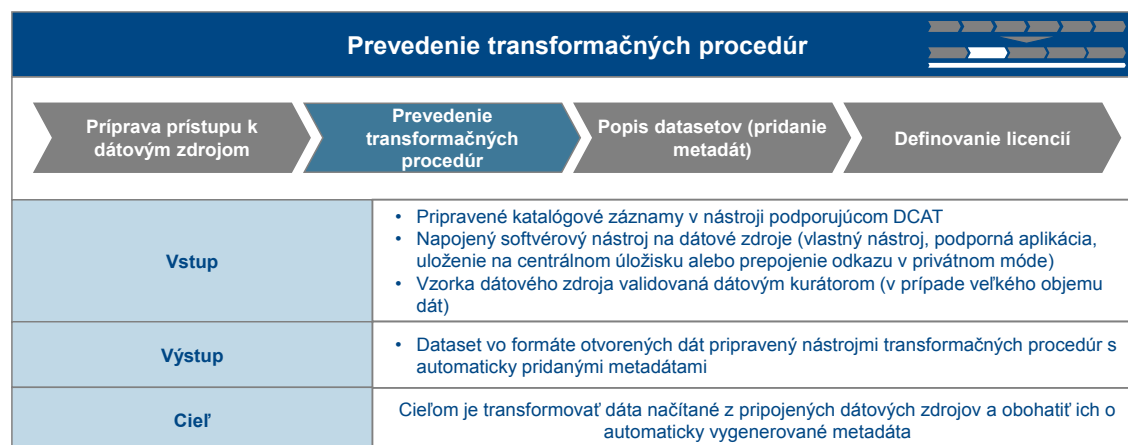
8. Povinná osoba počká na vyjadrenie dátového kurátora, ktorý schváli alebo zamietne žiadosť o nahranie dátového zdroja. Vzorka dát je postačujúca na posúdenie vhodnosti dátového zdroja a zároveň nevyžaduje náročné dátové prenosy pri podávaní žiadosti (z technických príčin nie je možné prostredníctvom podávania žiadosti cez ÚPVS pripojiť prílohu s ľubovoľnou veľkosťou).

Ak sú dáta nahrané v podpornej aplikácii alebo na centrálnom úložisku, je možné v tomto kroku tiež pripraviť v internom katalógu aplikácie alebo v centrálnom katalógu podporujúcom štandard DCAT aj katalógové záznamy v súlade s danou schémou katalógu. To znamená, že sa vyplnia tie atribúty, ktoré sú známe v čase prípravy na publikovanie. Nevypĺňajú sa tie údaje, ktoré budú známe až pri publikovaní, ako napríklad čas a dátum publikovania, a ktoré budú automaticky vyplnené počas transformačnej procedúry.

7.2 Prevedenie transformačných procedúr

Na základe prípravy prístupu k dátovým zdrojom sú nastavené a vykonané transformačné procedúry. Tieto procedúry možno previesť povinnou osobou v podpornej aplikácii alebo v samotnom transformačnom module centrálne, keď sú dáta označené ako private a možno nad nimi robiť transformačné procedúry. Ďalšou alternatívou je vytvorenie vlastného nástroja na transformačné procedúry, ako aj na proces nahrania dát z databáz, prípadne z ďalších dátových zdrojov.

Obrázok 18: Prevedenie transformačných procedúr



Táto metodológia odporúča automatizáciu procesov na publikovanie identifikovaných datasetov, ktorých súčasťou je aj prevedenie transformačných procedúr dedikovaným nástrojom (vlastným, centrálnym, dostupným ako transformačný modul MOD alebo podpornou aplikáciou). V centrálnom nástroji je možné nastaviť periodické behy transformačných procedúr. Pre osobitné datasety, ktoré sú veľmi komplexné, publikované len raz s veľmi dlhou lehotou aktualizácie, môže byť nevyhnutné vykonať transformačné procedúry manuálne, bez softvérového nástroja. Nástroj po načítaní dát z dátového zdroja implementuje súbor transformácií vrátane:

- Konverzie dátových formátov.
- Čistenia dát.
- Anonymizácie.
- Prelinkovania na ďalšie relevantné dáta a informácie.
- Obohacovania.
- OC.
- Sémantickej analýzy.

- Rozdielovej analýzy.

Po transformácii je nevyhnutné výsledný dataset skontrolovať, či naozaj spĺňa všetky náležitosti.

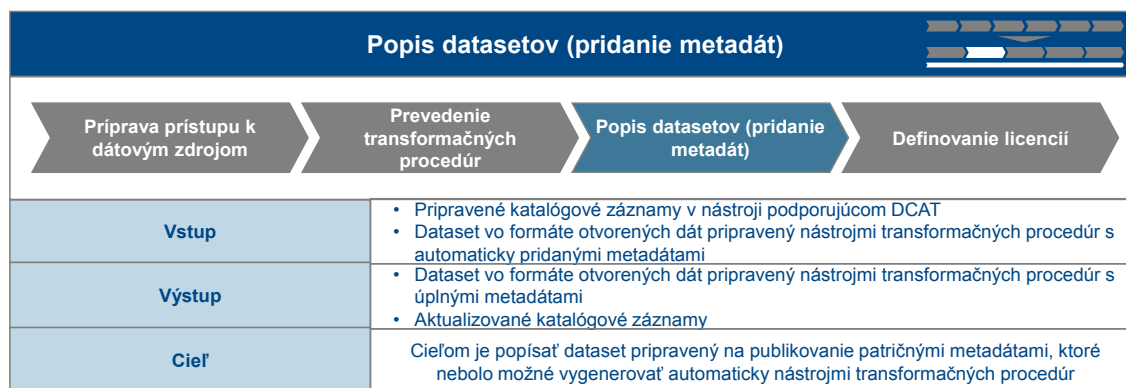
Samotný proces transformácie pozostáva z nasledujúcich krokov:

- Nakonfigurovanie transformačného predpisu (ak želaný predpis ešte nie je nakonfigurovaný podľa potrieb v systéme):
 - Výber transformačného predpisu zo zoznamu predpisov vytvorených a sprístupnených správcom transformácií.
 - Nakonfigurovanie načítavania (harvestovanie) vstupných dát v zmysle podporovaných foriem extrakcie:
 - SQL dotaz cez JDBC/ODBC priamo do DB.
 - Načítanie dátového súboru z filesystemu - podporované formáty: CSV, XML, JSON, TXT, HTML, TIFF, JPG, PDF.
 - Načítanie dát cez SCP/SFTP protokol.
 - Nakonfigurovanie transformácie (samotnej transformačnej logiky) v obmedzenej miere (v takej, akú povoľuje daný transformačný predpis) - napríklad voľba požadovaných parametrov pre transformáciu, či sa majú publikovať dáta a/alebo metadáta.
 - Nakonfigurovanie umiestnenia výsledkov transformácie - do ktorého datasetu má byť výsledný dátový zdroj zaradený.
- Spustenie transformačného predpisu – manuálne alebo časovačom:
 - Výber transformačného predpisu na spustenie manuálne alebo časovačom.
 - Pred manuálnym spustením predpisu nahratie vstupných dát (iba ak nakonfigurovaný transformačný predpis vyžaduje na vstupe manuálne nahratie dát) alebo nakonfigurovanie časovača - v akých intervaloch/termínoch sa má predpis automaticky spúšťať a poskytovateľ údajov odošle nastavenie.
 - Manuálne spustenie transformačného predpisu alebo automatické spustenie časovačom podľa konfigurácie.
 - Po dobehnutí transformácie je v GUI zobrazený poskytovateľovi údajov výsledok vykonania predpisu so stavom:
 - Úspech.
 - Neúspech.
 - Prípadné chybové hlásenia.

7.3 Popis datasetov (pridanie metadát)

Po vykonaní transformačných procedúr je nutné pridať k datasetom ich popis vo forme metadát. Metadáta sú štruktúrované údaje nesúce informácie o primárnych údajoch. V kontexte OpenData ide o údaje popisujúce jednotlivé dátové zdroje prístupné cez portál otvorených dát a nie všetky informácie v nich obsiahnuté je možné vygenerovať automaticky nástrojom na vykonanie transformácií. Metadáta spravidla zahŕňajú popisné údaje, napr. meno dátového zdroja, popis, jeho účel, typy údajov v dátovom zdroji, identifikáciu správcu, miesto sprístupnenia údajov z dátového zdroja, ale aj hodnotiace údaje, napr. hodnotenie kvality dátového zdroja, jeho aktuálnosti a pod.

Obrázok 19: Popis datasetov (pridanie metadát)



Metaúdaje datasetu s otvorenými údajmi presne definuje príloha č. 9 Výnosu Ministerstva financií Slovenskej republiky o štandardoch pre informačné systémy verejnej správy.¹⁸ Ak želaný dataset ešte neexistuje, vytvorí povinná osoba nový dataset, pričom vyplní html formulár, do ktorého zadá metaúdaje datasetu:

- Názov datasetu.
- Jednoznačný identifikátor poskytovateľa datasetu (generovaný automaticky v predchádzajúcom kroku).
- Popis datasetu prípadne aj s URL na jeho technickú dokumentáciu a s prepojením na dátovú schému (popísanú nižšie), aby nedochádzalo k dezinterpretácii datasetu.
- Dátum prvého zverejnenia datasetu (generovaný automaticky v predchádzajúcom kroku).
- Semantické značky (tagy) súvisiace s datasetom - kľúčové slová, popis datasetu.
- Prístupnosť datasetu (zverejnený/nezverejnený).

Ďalšie odporúčané metadáta datasetu sú nasledujúce:

- Identifikácia datasetu (generovaná automaticky).
- Priestorová informácia vo forme geodimenzie, prípadne zadanie priestorovej jednotky z číselníka.
- Jazyk datasetu.

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 dopĺňovaný pomocou autocomplete funkcie. V prípade, že sa formát nenachádza v databáze, bude nový formát do nej uložený.
- Licencia datasetu (pridelená podľa kapitoly 7.4 Definovanie licencií).
- Referencia na miesto sprístupnenia dátového zdroja a jeho schémy údajov.
- Informácia o aktuálnosti datasetu podľa § 53 písm. g):
 - Uvádza sa okamih alebo obdobie, v ktorom bol platný.
 - Po aktualizácii údajov sa vyznačí, ktoré údaje boli zmenené.

¹⁸ Výnos č. 55/2014 Z. z. Ministerstva financií Slovenskej republiky o štandardoch pre informačné systémy verejnej správy.

- Uvedie sa, či je dataset aktualizovaný v reálnom čase, alebo v určitej periodicite a v akej.
- Odlíšia sa chybné alebo nepresné údaje od správnych alebo presných údajov; ak to nie je možné, odlíši sa celý dataset ako dataset obsahujúci chybné alebo nepresné údaje.
- Informácia o začiatku a konci účinnosti datasetu alebo jeho údajov, najmä ak je účinnosť odlišná od platnosti.
- Prístupnosť dátového zdroja (zverejnený/nezverejnený).

Definované metaúdaje musia byť v JSON formáte, aby ich vedel systém spracovať. Ak sa metadáta vytvárajú cez grafické rozhranie MOD, systém sa automaticky postará o príslušný formát metadát.

Systém správy katalógov zvaliduje metaúdaje, vygeneruje ďalšie metaúdaje a uloží všetky metaúdaje dátového zdroja.

- Identifikácia dátového zdroja (generovaná automaticky).
- Dátum vytvorenia dátového zdroja.
- Dátum poslednej úpravy (pri vytváraní zhodný s dátumom vytvorenia).
- Dostupnosť dátového zdroja na základe vyhodnotenia referencie na miesto sprístupnenia.

Ku každému datasetu je dostupná spomínaná dátová schéma, 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 údajov. Formát dátovej schémy závisí od formátu datasetu:

- Pre CSV formát sa používa Metadata Vocabulary for Tabular Data.¹⁹
- Pre XML formát sa vyberie XML schéma.²⁰
- Pre JSON formát sa vyberie JSON schéma.²¹

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

7.4 Definovanie licencií

Po popísaní datasetu povinná osoba určí pre daný dataset konkrétnu licenčnú zmluvu (ďalej len licenciou) na základe nasledujúcej metodiky. Licenciou v autorskom práve poskytovateľ (autor diela) poskytuje oprávnenie užívať dielo nadobúdateľovi – poskytovateľovi otvorených údajov, ale aj používateľom otvorených údajov. Povinnosť prideliť datasetu licenciou vyplýva zo:

- Zákona č. 618/2003 o autorskom práve a právach súvisiacich s autorským právom (autorský zákon), na základe ktorého je nevyhnutné chrániť práva a oprávnené záujmy autora prostredníctvom licenčnej zmluvy.
- Výnosu 55/2014 Z. z., podľa ktorého musia byť údaje poskytované otvoreným spôsobom umožňujúcim voľné opakované použitie informácií s právnou oporou.
- Zákona č. 211/2000 o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov (zákon o slobode informácií) § 5 a § 21, ktoré definujú povinné zverejňovanie informácií, ktoré nespádajú pod autorský zákon, ako aj ďalších informácií spadajúcich pod opakované použitie informácií. Aj pri informáciách podľa § 5, na ktoré sa neviaže autorský zákon a sú automaticky v takzvanej verejnej doméne („public domain“), musí byť v rámci

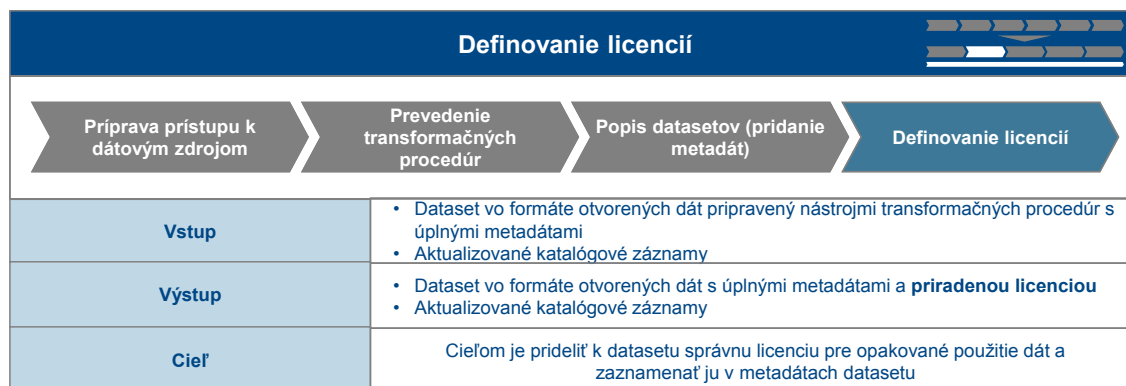
¹⁹ <http://w3c.github.io/csvw/metadata/>

²⁰ <http://www.w3.org/XML/Schema>

²¹ <http://tools.ietf.org/html/draft-zyp-json-schema-03>

datasetu tiež pridelená licencia, aby sa zabezpečilo poskytovanie týchto údajov otvoreným spôsobom (v tomto prípade sa udáva licencia CC0, ktorá bude popísaná nižšie).

Obrázok 20: Definovanie licencií



Dielo zverejnené v datasete otvorených údajov musí spadať pod autorský zákon, aby bolo možné autorské právo poskytovateľa k dielu licenciou zachovať. Kľúčovými typmi diela relevantnými pre oblasť otvorených údajov sú:

- **Audiovizuálne dielo** je dielo, ktoré je vnímateľné prostredníctvom technického zariadenia ako rad súvisiacich obrazov, či už sprevádzaných zvukom, alebo nie, ak je určené na uvádzanie na verejnosti; originálom audiovizuálneho diela je prvý zvukovo-obrazový záznam tohto diela určený na uvádzanie na verejnosti. Za spoluautorov tohto diela sa považujú najmä hlavný režisér, autor scenára, autor dialógov a autor hudby, ktorá bola vytvorená osobitne pre toto dielo.
- **Zvukovo-obrazový záznam** je záznam zvukov aj obrazov, ktoré sú vnímateľné sluchom aj zrakom bez ohľadu na to, akým spôsobom a na akom nosiči sa tieto zvuky a obrazy zaznamenávajú.
- **Zvukový záznam** je len záznam zvukov vnímateľný sluchom bez ohľadu na to, akým spôsobom a na akom nosiči sa tieto zvuky zaznamenávajú; záznam zvuku, ktorý je obsiahnutý v zázname audiovizuálneho diela, sa za zvukový záznam nepovažuje.
- **Kartografické dielo** v analógovej alebo v inej forme.
- **Súborné dielo** vyjadrené v akejkoľvek forme vrátane elektronickej formy zahŕňajúcej analógové aj digitálne vyjadrenie, najmä zborník, noviny, časopis, encyklopédia, antológia, pásmo, výstava alebo iná databáza, ak je súborom nezávislých diel alebo iných prvkov, ktorý je spôsobom výberu alebo usporiadaním obsahu výsledkom vlastnej tvorivej duševnej činnosti autora.
- **Databáza** ako súbor nezávislých diel, údajov alebo iných materiálov systematicky, alebo metodicky usporiadaných a jednotlivo prístupných elektronickými alebo inými prostriedkami. Za databázu sa nepovažuje počítačový program použitý pri zhotovení alebo prevádzke databázy prístupnej elektronickými prostriedkami.

Najčastejšie používaným dielom v otvorených údajoch je databáza, ktorú môže vytvoriť zamestnanec verejnej správy manuálne kopírovaním dostupných informácií o istej téme zachytenej v datasete, kombinovaním dostupných datasetov, obohacovaním datasetov o ďalšie informácie alebo automaticky exportom časti informácií z databáz informačných systémov. Hoci autorský zákon sa nevzťahuje na:

- text právneho predpisu,
- úradné rozhodnutie,
- verejnú listinu - verejná listina je písomný dokument, na ktorom je odtlačok pečiatky slovenského orgánu/úradu alebo úradnej osoby a podpis úradnej osoby. Verejnou listinou je nielen listina, ktorú vyhotovil slovenský orgán, ale aj listina, ktorú takýto orgán len potvrdil,

napríklad splnomocnenie, na ktorom notár osvedčil správnosť podpisu osoby. Rozhodujúce je, či na písomnosti je odtlačok úradnej pečiatky, alebo podpis úradnej osoby,

- verejne prístupný register,
- úradný spis,
- slovenskú technickú normu vrátane ich prípravnej dokumentácie a prekladu,
- prejavy prednesené pri prerokúvaní vecí verejných, avšak na súborné vydanie týchto prejavov a na ich zaradenie do zborníka je potrebný súhlas toho, kto ich predniesol,
- na dennú správu
- a ďalšie informácie, ktoré je povinná osoba povinná zverejniť na základe zákona č. 211/2000.

Pri vytvorení datasetu otvorených údajov sa udáva licencia CC0, ktorou zamestnanec verejnej správy dáva dataset na voľné opakované použitie a ktorá je v súlade s Výnosom 55/2014.

Licencia nemusí byť v písomnej forme a musí obsahovať najmä spôsob použitia diela, rozsah licencie, čas, na ktorý autor licenciu udeľuje, alebo spôsob jeho určenia a odmenu, pričom v prípade poskytovania údajov otvoreným spôsobom sa autor s nadobúdateľom (používateľom údajov) musí dohodnúť na bezodplatnom poskytnutí licencie. Keďže konkrétny používateľ údajov nie je vopred známy, využíva sa osobitný koncept ponuky licencie smerujúci voči neurčitým osobám, a prijatia tejto ponuky konaním, z ktorého možno vyvodiť súhlas s podmienkami licencie aj bez vyznenia autora. Licencia musí teda spĺňať aspoň nasledujúce požiadavky, definované vo Výnose 55/2014:

- právne aspekty prístupu k údajom a jeho používaniu sú explicitne vyrovnané – to znamená, že aplikovaním licencie je autor uzrozumený o všetkých pravidlách spojených s opakovaným použitím údajov,
- je umožnené vytvorenie právnych vzťahov na používanie údajov aj prostredníctvom anonymného vzdialeného automatizovaného prístupu, keďže datasety budú používané vopred neidentifikovateľnými používateľmi údajov na ich účely,
- prístup k údajom je umožnený všetkým osobám za rovnakých podmienok, pričom tieto podmienky sú explicitne uvedené – v prípade zverejnenia údajov na portáli data.gov.sk platia podmienky prístupu popísané a zrealizované na portáli,
- údaj je možné použiť na nekomerčný aj komerčný účel a je možné ho kombinovať s inými údajmi, dopĺňať, opravovať, modifikovať alebo použiť z datasetu bez povinnosti použitia ostatných údajov datasetu – táto požiadavka je kľúčová pre rozvoj podnikania a kreatívnych komunít nad otvorenými,
- činnosti podľa predchádzajúceho bodu sú bezodplatné.

Posledná časť prípravy na publikovanie údajov je výber licencie zo zoznamu na základe definovaných kritérií. V prípade existujúcej licencie na obsah datasetu, ktorú nemožno zmeniť, je možné ju pridať do poľa licencie pri zverejnení. Ak však neexistujú žiadne obmedzenia zo strany autora, je želaným cieľovým stavom mať jedinou dostupnú licenciu pre otvorené dáta **Creative Commons 0 licencie 4.0** (ďalej len CC0). CC0 umožňuje vedcom, učiteľom, zamestnancom verejnej správy a ďalším tvorcom a vlastníkom autorských práv alebo chráneného databázového obsahu vzdať sa svojich výhradných práv na používanie ich práce, a tým ich verejne zverejniť. Týmto verejným zverejnením poskytnú možnosť, aby ktokoľvek mohol voľne opakovane použiť, zlepšiť a budovať na ich diele na akékoľvek účely – komerčné aj neziskové – bez obmedzení daných autorským právom. V porovnaní s ostatnými Creative Commons licenciami, ktoré umožňujú držiteľom vybrať si z rozsahu oprávnení a zároveň si zachovať autorské právo, CC0 predstavuje možnosť úplne sa vzdať autorských práv alebo práv vzťahujúcich sa na ochranu databázy do najvyššej možnej miery, ktorú povoľuje legislatíva, a to s celosvetovou platnosťou. CC0 je univerzálny nástroj, ktorý nie je zakotvený do zákona žiadnej konkrétnej jurisdikcie – podobne ako je to v prípade open source softvérových licencií. CC0 je

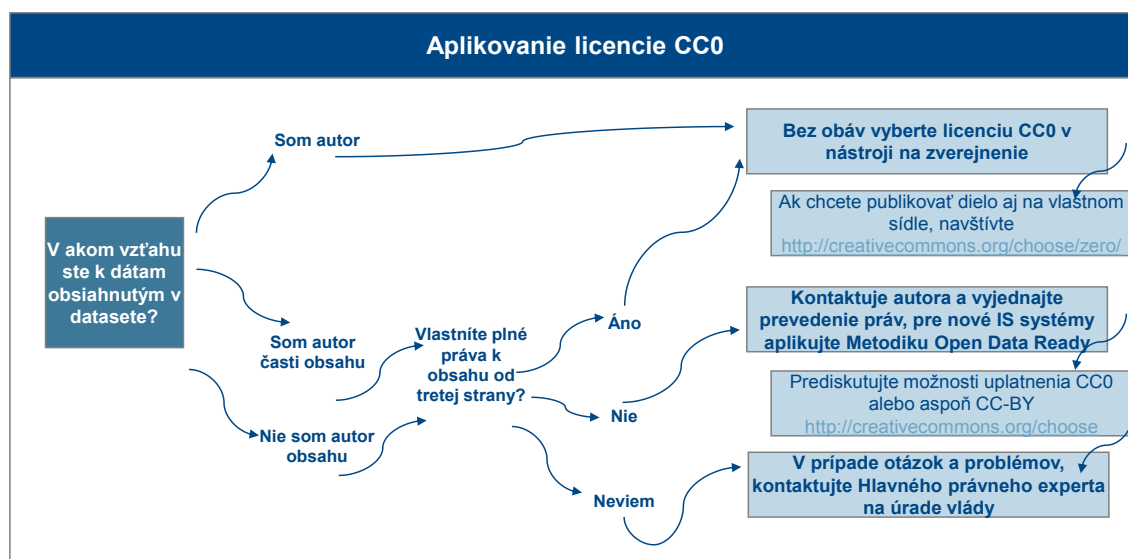
v súlade so zámermi iniciatívy otvorených dát na jednoduché opakované použitie, ktoré podporuje vznik nových biznis modelov a kreatívnych komunít.

Pred aplikovaním licencie na dataset je nutné splniť nasledujúci postup:

1. CC0 možno použiť, ak existujú autorské práva alebo obmedzenia aspoň v rámci jednej jurisdikcie. Aplikovaním CC0 sa tak umožní slobodné opakované používanie na celom svete bez pochybností, čím sa podporí ochota verejnosti využívať datasety.
2. CC0 môže použiť akákoľvek fyzická alebo právnická osoba len na vlastné dielo alebo, pokiaľ fyzická alebo právnická osoba disponuje nevyhnutnými právami na aplikovanie CC0 licencie, aj na cudzie dielo alebo na obsah databázy (pri splnení Metodiky Open Data Ready pre všetky nové informačné systémy bude táto podmienka splnená (to znamená, že povinná osoba bude disponovať patričnými právami) vďaka vzájomnej dohode s dodávateľom informačného systému).
3. Aj keď sa v rámci CC0 **neposkytuje záruka vlastníctva autorských práv**, povinná osoba stále zodpovedá za tretie strany, ktoré by mohli mať existujúce práva na aktuálne zverejňované dielo alebo na jeho časť. Vďaka ostatným pravidlám zverejňovania datasetov nikdy nejde o osobné údaje alebo utajované skutočnosti, na ktoré by sa mala aplikovať licencia CC0. Ak je napríklad časťou datasetu dielo inej osoby, ktorá ho zverejnila napríklad pod CC BY licenciou, je nevyhnutné publikovať túto časť datasetu samostatne aj s odkazom na autora a licenciu. Takýto postup však veľmi komplikuje zverejňovanie a opakované používanie, preto je veľmi dôležité vyhľadať autora a požiadať ho o zvolenie aplikovať CC0 na celý dataset.
4. **Dôležitá poznámka:** Ak sa raz aplikuje CC0, už nie je možné si to rozmyslieť neskôr a znovu aplikovať autorské práva na dielo alebo databázu. V niektorých prípadoch nie je jednoduché rozhodnúť, či sa niektorá časť diela kvalifikuje pre nevyhnutnosť ponechania autorských práv. V tomto prípade je potrebné obrátiť sa na odborníkov na Úrade vlády SR. Zámerom je však do najvyššej možnej miery pozbaviť všetky datasety akýchkoľvek autorských a licenčných práv, ktoré by bránili voľnému opakovanému použitiu.
5. Samotný proces pridelenia licencie prebieha na portáli <http://creativecommons.org/choose/zero/>, kde je poskytovateľ údajov navádzaný krok za krokom. Po ukončení sa vytvorí HTML kód, ktorý možno nakopírovať na vlastné webové sídlo v prípade potreby. Na portáli data.gov.sk sa momentálne tento kód nevyužíva.
6. Napriek aplikovaniu CC0 sa však všade tam, kde je to vhodné, odporúča uviesť poďakovanie a autora.
7. V prípade, že je nevyhnutné uvádzať autora – teda mu uznať dostatočný kredit za jeho prácu, je nevyhnutné aplikovať licenciu **Creative Attribution 4.0** – (ďalej len CC BY). Pre tento typ licencie platia body 1 až 3, avšak je potrebné zistiť u tvorca jeho požiadavky na uznanie kreditu za prácu. V tomto prípade sa sčasti komplikuje opakované používanie a kombinovanie datasetov, pretože je vždy nutné uviesť:
 - a. Meno autora a označenie zúčastnených strán ako napríklad spoluautora, sponzora alebo vydavateľa.
 - b. Oznámenie o autorských právach.
 - c. Oznámenie o licencií buď v plnom znení, napríklad Creative Commons Attribution-Noncommercial Australia 2.5, v skrátenom znení, napríklad CC BY-NC Aus 2.5, alebo cez tlačidlá licencie, ktoré možno získať postupom uvedeným na linke <http://creativecommons.org/choose>.
 - d. Linku na licenciu – napríklad <https://creativecommons.org/licenses/by/2.5/au/>.
 - e. Vyhlásenie o vylúčení zodpovednosti.
 - f. Odkaz na pôvodný materiál – jeho URL, ak je to relevantné.

- g. Vyhlásenie, či došlo k zmenám pôvodného diela a identifikovanie zmien, ktoré boli vykonané.

Obrázok 21: Postup aplikovania licencie CC0



Pridelenie CC licencie je v súlade s autorským zákonom, ktorý bol novelizovaný na umožnenie používania takejto licencie, ktoré však nie je povinné. Neodporúča sa zverejňovať dataset s duálnou licenciou jeho obsahu – čo znamená, že na rôzne časti datasetu sa vzťahuje rôzna licencia. Takýto postup veľmi komplikuje opakované použitie. V tomto prípade sa odporúča rozdeliť dataset na viacero datasetov, aby pre každý obsah takto vzniknutého datasetu platila len jedna licencia.

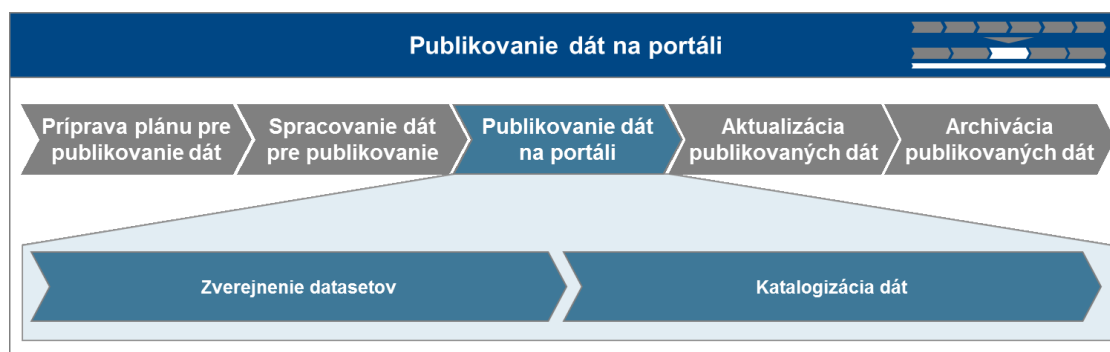
Ďalším zdrojom informácií o ochrane autorských práv je:

- Smernica Rady a Európskeho parlamentu 96/9/ES z 11. marca 1996 o právnej ochrane databáz,
- Smernica Rady a Európskeho parlamentu 2001/29/ES z 22. mája 2001 o harmonizácii niektorých aspektov autorských práv a s nimi súvisiacich práv v informačnej spoločnosti.

8 Publikovanie dát na portáli

V tejto časti metodiky sa definuje prístup k publikovaniu dát na portáli. Dodefinujú sa posledné manuálne kroky prípravy datasetov na publikovanie, ktoré nebolo možné vykonať automaticky podľa kapitoly 7.2. Publikovanie pozostáva zo samotného zverejnenia datasetov a z následnej katalogizácie v centrálnom katalógu otvorených dát.

Obrázok 22: Publikovanie dát na portáli



8.1 Zverejnenie datasetov

Po prevedení transformačných procedúr a pridaní metadát z predchádzajúcej kapitoly nasleduje zverejnenie prvej verzie datasetov alebo zverejnenie pripravenej aktualizácie.

Obrázok 23: Zverejnenie datasetov



Pred zverejnením samotného datasetu je potrebné overiť, či nie je nutné aplikovať zložitejšie alebo manuálne postupy anonymizácie ako:

- Projekcia – odobratie atribútov z datasetu, ktoré neboli automatickým algoritmom vyhodnotené ako osobné údaje alebo utajované skutočnosti, či obchodné tajomstvo.
- Agregácia – zoskupenie viacerých prvkov údajov do jedného štatistického prvku s cieľom zverejniť dataset vo vyššej kvalite požadovanej používateľmi údajov (napríklad zoskupenie osôb a ich veku v danom regióne a publikovanie len priemerného veku osôb v danom regióne).
- Odstránenie liniek – v prípade prelinkovania údajov v datasetoch zverejňovaných na úrovni 5* je nevyhnutné skontrolovať, či nebudú odhalené niektoré utajované skutočnosti. Ak áno, je nutné pred publikovaním linky odobrať.

Nevyhnutné je tiež overiť, či existuje spôsob, ktorým je možné údaje po anonymizácii spätne deanonymizovať.

Ďalšími parametrami, ktoré ovplyvňujú náročnosť zverejnenia, sú:

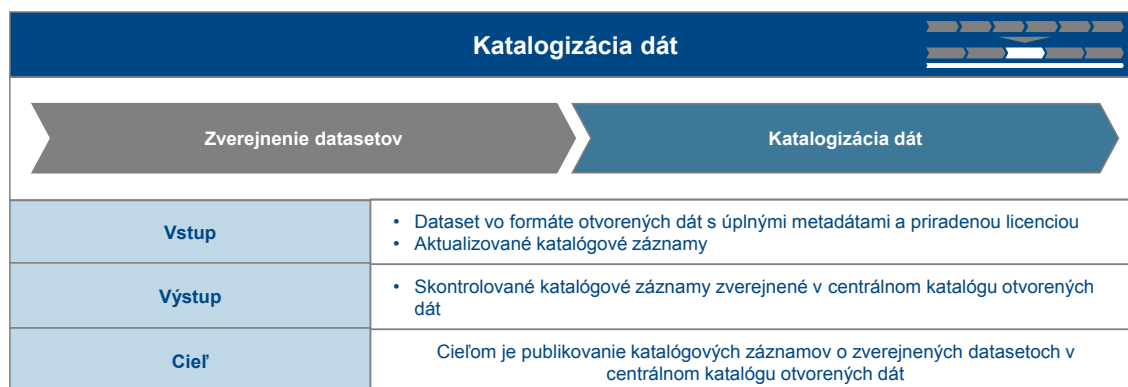
- Cieľová periodicita aktualizácie datasetu – jej stanoveniu sa venuje kapitola 9.
- Dostupnosť historických verzií datasetu.
- Zverejnenie zoznamu zmien oproti predchádzajúcej verzii datasetu (ak je takýto zoznam možné technicky vytvoriť).
- Ak je to technicky a finančne možné, odporúča sa publikovať aj API, ktoré umožní priamy prístup ku konkrétnym entitám datasetu:
 - Pre úroveň 3* a 4* musí byť API webová služba typu RESTful.
 - Pre úroveň 5* ide o koncový bod SPARQL.

Proces zverejnenia v centrálnom katalógu možno uskutočniť nasledujúcimi spôsobmi:

- Manuálnou zmenou príznaku private na public, ak neboli potrebné transformačné procedúry alebo ak tieto boli vykonané manuálne a dátový zdroj už bol nahraný vo formáte otvorených dát, prípadne ak je dátový zdroj len prelinkovaný s úložiskom povinnej osoby.
- Po prevedení transformačných procedúr automatickou zmenou príznaku datasetov na public, čím sa ukončí proces transformácie a následného zverejnenia (automatické nastavenie príznaku na public možno aj deaktivovať, potom musí povinná osoba zmeniť príznak z neverejných dát na verejné manuálne).
- Synchronizovaním centrálneho úložiska s podpornou aplikáciou.
- Použitím vlastného ETL nástroja nad databázou, ktorý po transformácií zosynchronizuje datasety cez API kľúč na centrálné úložisko.

8.2 Katalogizácia dát

Obrázok 24: Katalogizácia dát



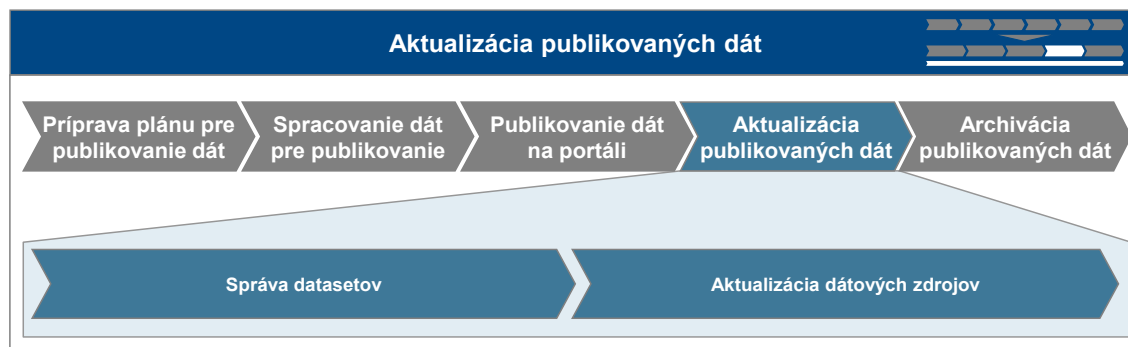
Cieľom tohto kroku je overenie správneho publikovania katalógových záznamov v katalógu dát. Ide napríklad o hodnoty, ktoré sú dostupné až po publikovaní dát - napríklad dátum publikovania, dátum poslednej zmeny datasetu, účinnosť datasetu alebo prepojenie datasetu s dátami na stiahnutie. Tieto hodnoty môžu byť zmenené manuálne.

Bez ohľadu na miesto zverejnenia samotného datasetu sa katalógové záznamy presunú do centrálneho katalógu otvorených dát.

9 Aktualizácia publikovaných dát

Cieľom tohto kroku je udržiavať datasety a katalógové záznamy aktuálne, zlepšovať kvalitu zverejňovaných dát a reagovať na podnety verejnosti a používateľov údajov.

Obrázok 25: Aktualizácia publikovaných dát



9.1 Správa datasetov

V rámci správy datasetov možno identifikovať nasledujúce typy úkonov:

- Aktualizácia katalógových záznamov.
- Oprava a doplnenie katalógových záznamov pri zistených nezrovnalostiach.
- Aktualizácia datasetov, ktorej sa venuje ďalšia kapitola.
- Oprava identifikovaných nedostatkov alebo chýb v samotnom datasete.
- Testovanie transformačných procedúr s ohľadom na zmeny vykonávané v dátových zdrojoch.

Uvedené aktivity sa plánujú na základe:

- Určenej periodicity aktualizácie datasetu.
- Pravidelného intervalu pre aktivity spravovania.
- Spätnej väzby od používateľov údajov, ktorá môže poukazovať na potrebu zvýšenia kvality datasetov, opravu chýb v datasetoch alebo katalógových záznamoch.

Obrázok 26: Správa datasetov

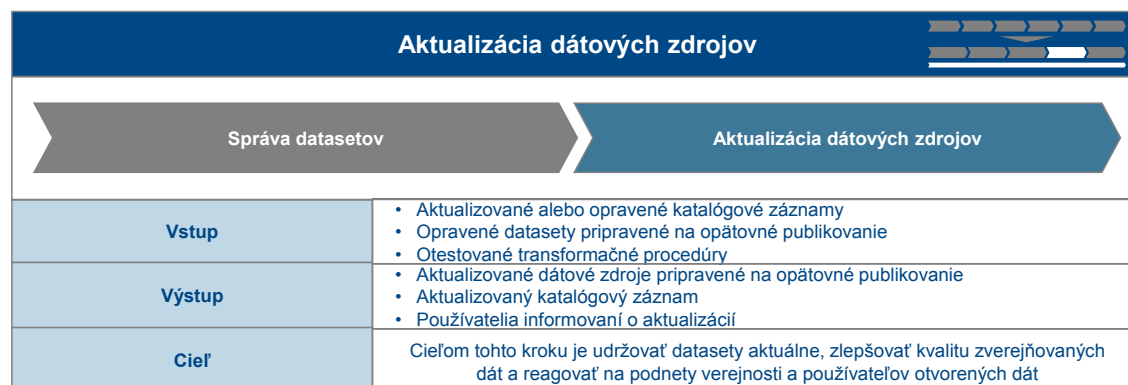


V prípade zmeny štruktúry datasetu je potrebné zanalyzovať, či dôjde aj k zmene dátovej schémy. Ak áno, odporúča sa dataset archivovať podľa nasledujúcej kapitoly a vytvoriť nový dataset s aktualizovanou dátovou schémou.

Spravovať a editovať metadáta a dátové zdroje je možné buď cez podpornú aplikáciu s následnou synchronizáciou, cez API kľúč, alebo priamo v nástroji centrálného katalógu, ktorý obsahuje osobitné tlačidlo Spravuj.

9.2 Aktualizácia dátových zdrojov

Obrázok 27: Aktualizácia dátových zdrojov



V závislosti od typu dát sa určí, či sa aktualizované dáta budú publikovať ako separátne dátové zdroje, alebo ako jednotný dátový zdroj, ktorý obsahuje kompletne historické údaje. Napríklad štatistiky zahraničného trhu sú publikované ako oddelené súbory na rok 2013 a 2014. V druhom prípade sa dataset prepíše novými údajmi, len čo sú tieto údaje aktualizované. Príkladom sú informácie o inšpekciách, pokutách a zákazoch Českej obchodnej inšpekcie, ktoré sa generujú štvrťročne a dataset s údajmi o poslednom štvrťroku prepíše ten predchádzajúci. Pri výbere typu aktualizácie je nutné brať do úvahy aj spätnú väzbu používateľov údajov.

Informovanie o aktualizácii prebehne cez notifikačné API a novým záznamom v sekcii noviniek na portáli data.gov.sk.

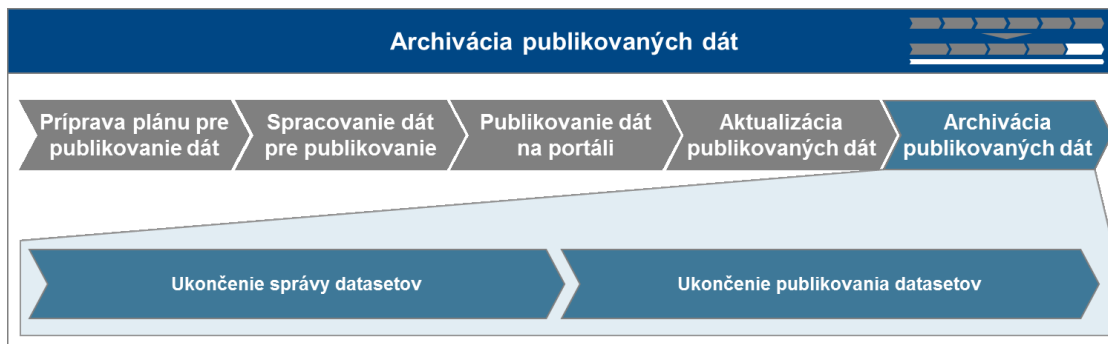
Periodicita aktualizácie zverejnených datasetov sa určuje na základe nasledujúceho:

- Aktualizácie nemusia byť publikované tak často, ako nastane zmena v dátovom zdroji. Napríklad ak sa dátový zdroj mení každú hodinu, aktualizácia datasetu môže byť zverejnená raz za deň – to znamená, že cieľová periodicita aktualizácie je jeden týždeň. Čím bližšie je však cieľová periodicita aktualizácie k frekvencii zmien dátového zdroja, tým kvalitnejší je poskytovaný dataset, avšak za cenu možných vyšších nákladov zverejnenia.
- Pri cieľovej periodicite aktualizácie je dôležité tiež zohľadniť spätnú väzbu používateľov údajov alebo overiť, že cieľová periodicita aktualizácie je pre nich postačujúca.

Poznámka: Dátové zdroje sa len pridávajú, nezmazávajú sa. Ak je to však nutné, môže administrátor dátový zdroj aj vymazať. Ak sú dátové zdroje dostupné cez linku a zdieľané z úložiska povinnej osoby, je potrebné ponechať aj staré verzie a verzionovať súbory – táto aktivita je ponechaná na povinnú osobu. Cez transformačný nástroj je možné urobiť spomínanú rozdielovú analýzu a vytvoriť nový dataset a rozdielový súbor.

10 Archivácia publikovaných dát

Obrázok 28: Archivácia publikovaných dát



10.1 Ukončenie správy datasetov

V prípade, ak primárne dáta pre datasety nie sú ďalej zbierané alebo ak dátová schéma, alebo význam dát prešiel významnou zmenou, z ktorej vyplýva nutnosť vytvoriť nový dataset, správa pôvodného datasetu musí byť ukončená. Ak je to možné a uskutočniteľné, zverejnené dáta by mali byť naďalej dostupné na opakované použitie týchto datasetov a preto by malo dôjsť k archivácii datasetu. Zodpovedajúce katalógové záznamy by mali byť taktiež zachované, ale je nutné, aby bola k záznamu pridaná informácia, že dataset nie je naďalej spravovaný. Používatelia údajov by mali byť informovaní o zmene statusu datasetu.

Obrázok 29: Ukončenie správy datasetov



Je veľmi dôležité informovať užívateľov o ukončení správy datasetu cez notifikačné API a novým záznamom v sekcii noviniek na data.gov.sk.

10.2 Ukončenie publikovania datasetov

Dôvody na ukončenie publikovania datasetov môžu byť napríklad spôsobené legislatívnymi zmenami, rozhodnutím súdu a podobne. Následkom týchto zmien môže publikovanie týchto datasetov byť porušením povinností alebo práv tretích strán. V takom prípade je nutné zablokovat verejný prístup k datasetu a k zdrojovým dátam (zmena príznaku na private). Je nutné kontaktovať známych používateľov údajov, ktorí si už dáta stiahli, a informovať ich o ukončení publikovania. Napriek ukončeniu publikovania datasetu by mal byť zachovaný katalógový záznam s aktualizovanou informáciou o tom, že dataset už nie je dostupný. Ak datasety boli publikované u povinnej osoby, odporúča sa ich nahrat' na portál a upraviť linky.

Následne môžu byť datasety lokálne zmazané. Zároveň je veľmi dôležité, aby používateľ údajov daného datasetu bol informovaný o ukončení publikovania datasetu.

Obrázok 30: Ukončenie publikovania datasetov



Dôležité je tiež poznamenať, že sa môže meniť platnosť a účinnosť údajov v datasete, čo je nutné v ňom nastaviť. Datasetu možno tiež nastaviť stav, že sa už neaktualizuje. Pri dátach je možné definovať presnosť a exaktnosť.

11 Prílohy

11.1 Checklist na splnenie aktivít

Povinná osoba si musí zvážiť plnenie nasledujúcich predpokladov, aby sa mohla úspešne zapojiť do iniciatívy otvorených dát.

Aktivita	Popis aktivity	Kapitola	Povinná	Vykonaná
Pridelenie oprávnení	Podanie žiadosti o pridelenie oprávnení na ÚPVS pre zastupovanie povinnej osoby v oblasti publikovania otvorených dát (napríklad štatutár podá prvú žiadosť a následne možno žiadať o pridelenie oprávnení viacerých používateľov za povinnú osobu)	Používateľská príručka	Áno	
Prevzatie API kľúča	Povinná osoba si po prihlásení zo svojho profilu prevezme svoj API kľúč (je možné ho v prípade potreby pregenerovať) – v prípade, že chce používať podpornú aplikáciu alebo publikovanie dát na centrálnom portáli cez API napríklad pomocou vlastného nástroja napojeného na databázu	Používateľská príručka	Nie	
Organizačná štruktúra	Zabezpečenie organizačnej štruktúry podľa Tabuľka 1: Odporúčaná definícia rolí a ich zodpovedností u povinnej osoby	5	Nie	
Identifikovanie datasetov	Povinná osoba musí mať jasnú predstavu, ktorými datasetmi disponuje a ktoré z nich možno zverejniť	6.1	Áno (Uznesenie vlády SR č. 59/2015 z 11. 2. 2015)	
	Analýza dostupných dátových zdrojov	6.1.1	Áno (Uznesenie vlády SR č. 59/2015 z 11. 2. 2015)	
	Identifikovanie datasetov na zverejnenie	6.1.2	Áno (Uznesenie vlády SR č. 59/2015 z 11. 2. 2015)	
	Identifikovanie údajov vylúčených z povinnosti zverejnenia	6.1.2.1	Áno (§ 8 - § 11 zákona č. 211/2000)	
	Stanovenie náročnosti zverejnenia jednotlivých datasetov	6.1.3	Nie	

Aktivita	Popis aktivity	Kapitola	Povinná	Vykonaná
	Stanovenie priorít na zverejnenie datasetov	6.1.4	Áno (Uznesenie vlády SR č. 59/2015 z 11. 2. 2015)	
Register rizík	Na riadenie rizík v oblasti zverejňovania údajov a na dohľad nad ochranou osobných údajov a utajovaných skutočností sa odporúča viesť Register rizík	6.1.2.2	Nie	
Plán zverejňovania	Odporúča sa udržiavať plán zverejňovania so zadanou náročnosťou a s prioritou zverejňovania jednotlivých datasetov.	6.2	Nie	
Publikovanie	Povinná osoba si nastaví alebo implementuje nástroje na publikovanie otvorených dát (podpornú aplikáciu, vlastné transformačné nástroje napojené na databázy informačných systémov, metódy transformácie v transformačnom module)	7 8	Áno (§ 5 - § 7 zákona č. 211/2000)	
	Príprava prístupu k dátovým zdrojom	7.1	Áno (§ 53 Výnos č. 55/2014)	
	Prevedenie transformačných procedúr	7.2	Áno (§ 53 Výnos č. 55/2014)	
	Pridanie metadát	7.3	Áno (Príloha č. 9 k Výnosu č. 55/2014 Z. z.)	
	Vytvorenie referencovateľných identifikátorov	6.1.3.1	Áno (§ 46 Výnos č. 55/2014)	
	Definovanie licencií	7.4	Áno (§ 52 Výnos č. 55/2014)	
	Zverejnenie datasetov	8.1	Áno (§ 6 zákona č. 211/2000)	
	Katalogizácia dát	8.2	Áno (§ 53f Výnos č. 55/2014)	
Aktualizácia	Povinná osoba si nastaví procesy na aktualizáciu údajov vo zverejnených datasetoch	9	Áno (§ 53g Výnos č. 55/2014)	

Aktivita	Popis aktivity	Kapitola	Povinná	Vykonaná
Archivácia	Povinná osoba si nastaví procesy na archiváciu údajov vo zverejnených datasetoch	10	Nie	

11.2 Aktualizácia informačných systémov verejnej správy

Cieľom programového obdobia 2014 až 2020 je dosiahnuť, aby všetky novovzniknuté datasety v kompetencii orgánov verejnej správy boli v otvorenom formáte a zároveň pokračovať v konverzii existujúcich datasetov podľa aktualizovaného zoznamu datasetov verejnej správy.

Ďalej sa pre aktualizáciu alebo pre implementáciu nových informačných systémov plánuje zavedenie metodika Open Data Ready. Metodika Open Data Ready je 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 na plnenie týchto kritérií. Kľúčovým kritériom metodiky je vyrovnanie 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 na čerpanie financií zo štrukturálnych fondov EÚ.