

Národný projekt Otvorené údaje 2.0

# PoC Transformačné a Validačné API pre otvorené údaje

ŠPECIFIKÁCIA DIELA



**Miroslav Líška**  
**Dátová kancelária MIRRI**  
**v. 2023-02-19**

## Obsah

.....	1
Obsah.....	2
1 Úvod.....	3
1.1 Použité pojmy a skratky.....	3
2 Definícia projektu OU 2.0.....	4
2.1 Komponenty OD2.0.....	4
3 Ciele PoC.....	6
3.1 Definícia cieľov PoC.....	6
4 Požiadavky.....	7
4.1 Katalóg požiadaviek.....	7
4.2 Aktéri.....	7
4.3 Prípady použitia.....	8
4.3.1 Subsystem Validácie distribúcií datasetov.....	8
4.3.2 Subsystem Transformácie distribúcií datasetov.....	9
6 Návrh riešenia.....	10
6.2 Softvérová architektúra.....	10
7 Implementácia.....	11
8 Referencie.....	12

# 1 Úvod

Účelom tohto dokumentu je špecifikácia PoC komponentu OD2.0 -

## 1.1 Použité pojmy a skratky

V tomto dokumente sú použité nasledujúce skratky, pojmy a značky:

Skratka / Pojem	Vysvetlenie / Popis
5★ OpenData	Metodika kvality otvorených údajov založená stupňoch strojovej spracovateľnosti a interoperability- <a href="https://5stardata.info/en/">https://5stardata.info/en/</a>
CRUD	Skratka pre operácie s dátovými entitami alebo operácie služieb, CREATE, READ, UPDATE, DELETE
CKAN	Publikačná platforma pre otvorené údaje
DCAT	Data Catalog Vocabulary. Štandard pre reprezentáciu metadát o katalógoch, datasetoch i ich distribúciách
DCAT-AP-SK	Platná súčasná národná implementácia DCAT štandardu. Definuje štruktúru, povinné a nepovinné metadáta publikovaných otvorených údajov publikovaná na <a href="https://metais.vicemier.gov.sk/publicspace?pagelid=23990179">https://metais.vicemier.gov.sk/publicspace?pagelid=23990179</a>
EÚ	Európska únia
Formát s vysokým potenciálom na znovupoužitie	Formát s vysokým potenciálom na znovupoužitie je definovaný ako štruktúrovaný formát údajov v otvorenom štandarde, t.j. predstavuje úroveň 3★ kvality otvorených údajov, tj. CSV, XML, ODS
IAM	Identity and Access Management (Modul pre identifikáciu používateľa a riadenie prístupov)
IKT	Informačné a komunikačné technológie
ISVS, ITVS	Informačné systémy verejnej správy. Pre vývoj a prevádzku ISVS platí vyhláška o štandardoch pre informačné technológie verejnej správy (dostupné na <a href="https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2020/78/20200501">https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2020/78/20200501</a> ), prípadne verzia aktualizovaná počas plnenia PZ
KPI	Key performance indicator, Kľúčový ukazovateľ výkonnosti
MIRRI	Ministerstvo investícií, regionálneho rozvoja a informatizácie SR
NASES	Národná agentúra pre sieťové a elektronické služby
NKIVS	Národná koncepcia informatizácie verejnej správy
NKOD	Národný katalóg otvorených dát
OD	Open Data, otvorené údaje
OPZ	Opis predmetu zákazky
OVM	Orgány verejnej moci

## 2 Definícia projektu OU 2.0

Národný projekt **Otvorené údaje 2.0 - Rozvoj centrálnych komponentov pre kvalitné zabezpečenie otvorených údajov** (ďalej iba „OÚ 2.0“) nadväzuje a je pokračovaním národného projektu Elektronické služby Úradu vlády Slovenskej republiky - eDemokracia a otvorená vláda, ktorý bol realizovaný v rámci programového obdobia 2007 – 2013 z prostriedkov Operačného programu Informatizácia spoločnosti a bol realizovaný Národnou agentúrou pre sieťové a elektronické služby, prostredníctvom dohody o spolupráci.

Projekt OÚ 2.0 realizuje Ministerstvo investícií, regionálneho rozvoja a informatizácie SR a jeho zámerom je rozvoj a dobudovanie centrálnych technických komponentov a prostriedkov na trvalé sprístupnenie otvorených údajov verejnej správy SR. MIRRI bude zároveň prijímateľom nenávratného finančného príspevku z prostriedkov Prioritnej osi 7 Operačného programu Integrovaná infraštruktúra (ďalej len „PO7 OP II“). Nižšie opísané parametre sú uvedené v Zmluve o poskytnutí nenávratného finančného príspevku č. 311071Y652 uzatvorenej dňa 13.12.2019 (dostupná: [https://www.crz.gov.sk/data/att/4376772\\_dokument1.pdf](https://www.crz.gov.sk/data/att/4376772_dokument1.pdf)).

Cieľom projektu je naplnenie nižšie uvedených cieľov schválených riadiacim orgánom MIRRI:

<b>PRIORITNÁ OS Operačného programu integrovaná infraštruktúra</b>	PO 7	Informačná spoločnosť
<b>Tematický cieľ</b>	TC2	Zlepšenie prístupu k informačným a komunikačným technológiám a zlepšenie ich využívania a kvality
<b>Investičná priorita</b>	2c)	Posilnenie aplikácií IKT v rámci elektronickej štátnej správy, elektronického vzdelávania, elektronickej inklúzie, elektronickej kultúry a elektronického zdravotníctva
<b>Špecifické ciele</b>	7.5	Zlepšenie celkovej dostupnosti dát verejnej správy vo forme otvorených dát
<b>Kód intervencie</b>	079	Prístup k informáciám verejného sektora (vrátane elektronickej kultúry s otvoreným prístupom k údajom, digitálnych knižníc, digitálneho obsahu a elektronického cestovného ruchu)
<b>Merateľné ukazovatele</b>	P0217	Počet nových datasetov publikovaných vo formáte s vysokým potenciálom na znovupoužitie
	P0050	Dodatočný počet inštitúcií verejnej správy prepojených s centrálnou platformou pre otvorené dáta

### 2.1 Komponenty OD2.0

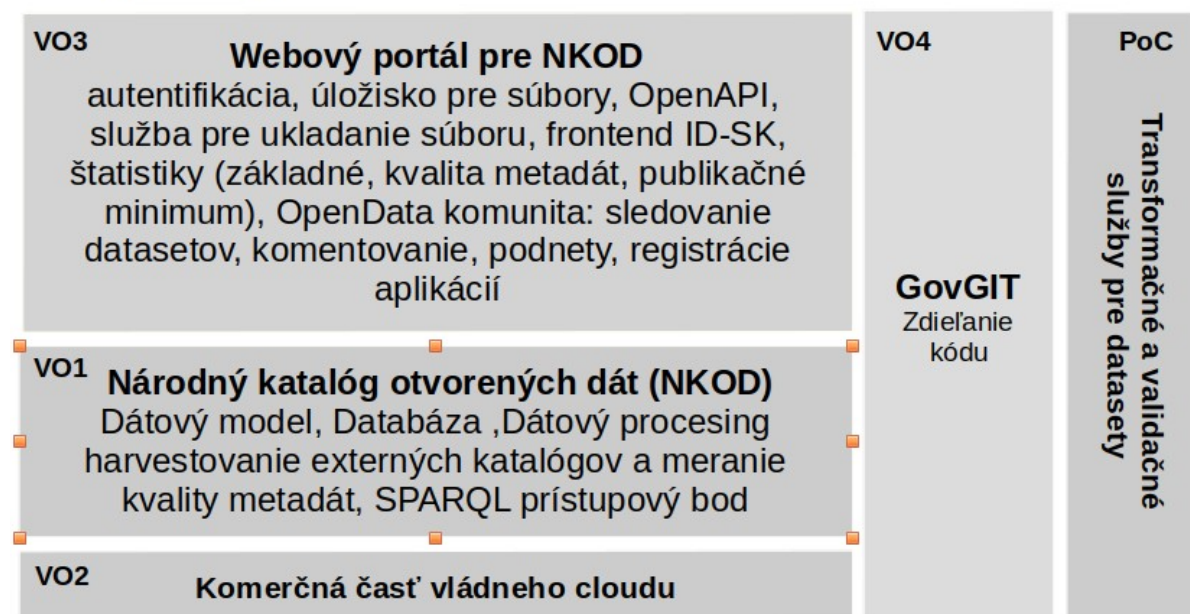
Od času schválenia štúdie uskutočniteľnosti v roku 2019 prešiel projekt, ako aj samotná oblasť otvorených dát mnohými spresneniami, pričom projekt sa logicky rozdelil na samostatné komponenty, ktoré sa vzájomne dopĺňajú. Z dôvodu že je oblasť portálov s otvorenými údajmi výrazne rozvinutá a existuje množstvo štandardov a odporúčaní v rámci EÚ, sú jednotlivé komponenty OD2.0 realizované so zámerom znovupoužitia existujúcich riešení. V tomto smere bolo odstúpené od zámeru vyvinutia úplne nového riešenia portálu s otvorenými údajmi a ako riešenie bolo vybrané znovupoužitie jadra portálu data.gov.cz, ktoré vytvára tzv. **Národný katalóg otvorených údajov**.

Medzi základné komponenty projektu OD2.0 patria:

- **Národný katalóg otvorených dát (NKOD)** – dátové jadro vytvorené znovapoužitím data.gov.cz – tj. RDF databáza, Dátový procesing: harvestovanie lokálnych katalógov, meranie kvality metadát otvorených údajov, SPARQL prístupový bod
- **Komerčná časť vládneho cloudu** – projekt bude bežať v komerčnej časti vládneho cloudu
- **Webový portál pre NKOD** – Webový portál využívajúci NKOD ako zdroj údajov vytvorení v súlade s jednotným design manuálnom IDSK.
- **GovGit** – Platforma pre zdieľanie otvorených zdrojových kódov informačných systémov verejnej správy.
- **PoC Transformačné a validačné služby pre otvorené údaje** – samostatný modul, ktorý pre vybranú množinu datasetov (publikačné minimum orgánu verejnej správy) umožňuje validovať ich správnosť a transformovať datasety do distribúcií s iným ako dostupným formátom.

Predmetom špecifikácie sú Transformačné a validačné služby pre datasety otvorených údajov.

## Komponenty OD2.0



## 3 Ciele PoC

### 3.1 Definícia cieľov PoC

Cieľom tohto PoC vytvorenie elektronických služieb v podobe samostatného modulu (samostatného servera) poskytujúcu množinu aplikačných programovaných rozhraní API, ktoré pre vybrané datasety:

- poskytnú validačnú službu distribúcie daného datasetu, ktorá vyhodnotí jednak syntaktickú správnosť datasetu (formát distribúcie a formáty údajov) ako aj obsahovú časť (dátovú štruktúru a údaje daného datasetu).
- poskytnú transformačnú službu danej distribúcie datasetu do iného formátu (napr. CSV do RDF alebo opačne)
- rozsah podporovaných formátov ako aj použitých štandardov, zoznam validácií a transformácií je definovaný v Katalógu požiadaviek
- Systém poskytujúci API určené pre validáciu a transformáciu distribúcií datasetov otvorených údajov musí byť z pohľadu projektu NKOD (Národný katalóg otvorených dát) realizovaný ako externý komponent. NKOD je na tomto systéme nezávislý.
- Jednotlivé validácie a transformácie distribúcií datasetov sa neimplementujú priamo v kóde, ale prostredníctvom použitia vhodného ETL systému

## 4 Požiadavky

V tejto časti sú uvedené všetky požiadavky na predmet zákazky.

### 4.1 Katalóg požiadaviek

Katalóg požiadaviek je nedeliteľnou súčasťou tohto Opisu predmetu dokumentu.

### 4.2 Aktéri

Názov aktéra	Typ	Popis
Používateľ API	Človek, Stroj	Klient volajúci API
Správca transformácie	Človek	Tvorba validáčnej, resp. transformačnej pipeline.
Externý ISVS	Stroj	Externý ISVS používajúci validačné a transformačné API

## 4.3 Prípady použitia

### 4.3.1 Subsystem Validácie distribúcií datasetov

Případ použitia	Subsystem
UC Validácia distribúcií datasetu Organizačná štruktúra	Publikačné minimum
UC Validácia distribúcií datasetu Kontaktné miesta	Publikačné minimum
UC Validácia distribúcií datasetu Úradné hodiny	Publikačné minimum
UC Validácia distribúcií datasetu Podriadené organizácie	Publikačné minimum
UC Validácia distribúcií datasetu Pracovné skupiny	Publikačné minimum
UC Validácia distribúcií datasetu Prijaté žiadosti podľa infozákona	Publikačné minimum
UC Validácia distribúcií datasetu Strategické dokumenty	Publikačné minimum
UC Validácia distribúcií datasetu Najžiadanejšie otvorené údaje OVM	Publikačné minimum
UC Validácia distribúcií datasetu Katalóg otvorených údajov orgánu verejnej správy	Publikačné minimum
UC Validácia distribúcií datasetu Časový harmonogram sprístupňovania datasetov	Publikačné minimum
UC Validácia distribúcií datasetu Objednávky	Publikačné minimum
UC Validácia distribúcií datasetu Faktúry	Publikačné minimum
UC Validácia distribúcií datasetu Dotácie (schémy, obmedzenia, výzvy, žiadosti)	Publikačné minimum
UC Validácia distribúcií datasetu Zahraničné služobné cesty	Publikačné minimum



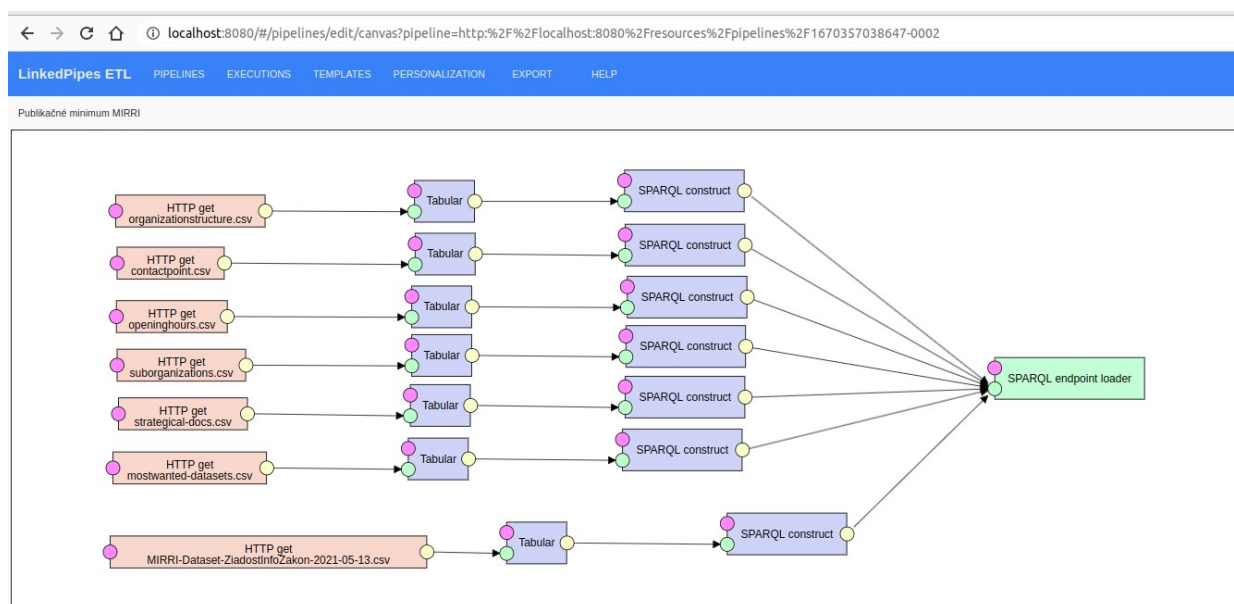
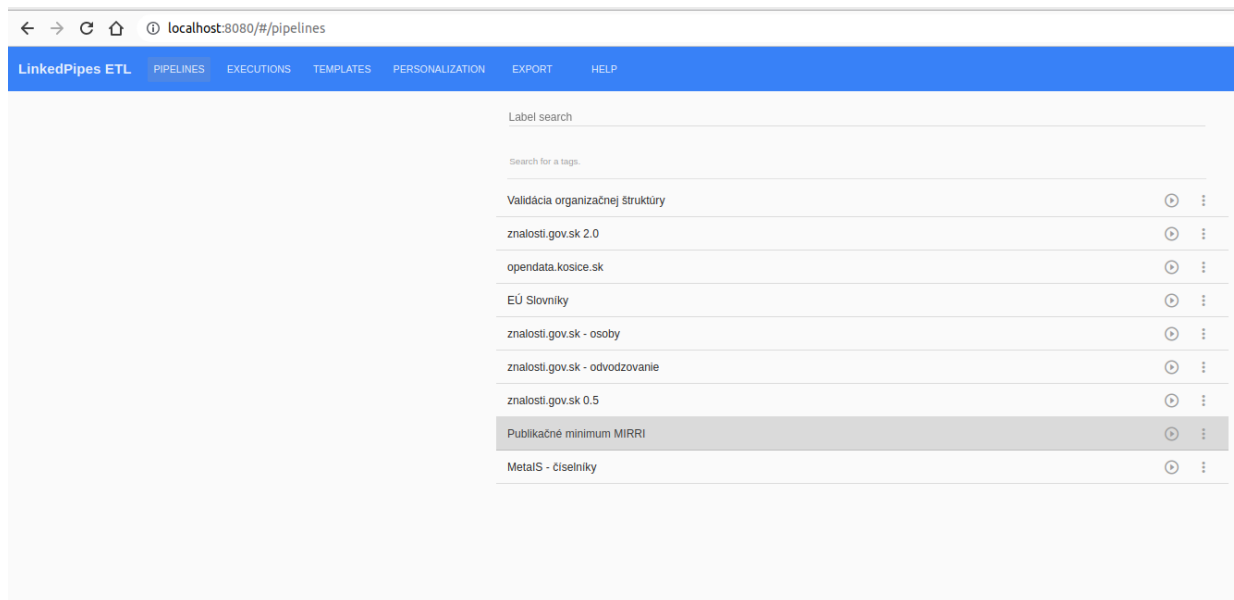
#### 4.3.2 Subsystem Transformácie distribúcií datasetov

Pírpád použítia	Subsystem
UC Transformácia distribúcií datasetu Organizačná štruktúra	Publikačné minimum
UC Transformácia distribúcií datasetu Kontaktné miesta	Publikačné minimum
UC Transformácia distribúcií datasetu Úradné hodiny	Publikačné minimum
UC Transformácia distribúcií datasetu Podriadené organizácie	Publikačné minimum
UC Transformácia distribúcií datasetu Pracovné skupiny	Publikačné minimum
UC Transformácia distribúcií datasetu Prijaté žiadosti podľa infozákona	Publikačné minimum
UC Transformácia distribúcií datasetu Strategické dokumenty	Publikačné minimum
UC Transformácia distribúcií datasetu Najžiadanejšie otvorené údaje OVM	Publikačné minimum
UC Transformácia distribúcií datasetu Katalóg otvorených údajov orgánu verejnej správy	Publikačné minimum
UC Transformácia distribúcií datasetu Časový harmonogram sprístupňovania datasetov	Publikačné minimum
UC Transformácia distribúcií datasetu Objednávky	Publikačné minimum
UC Transformácia distribúcií datasetu Faktúry	Publikačné minimum
UC Transformácia distribúcií datasetu Dotácie (schémy, obmedzenia, výzvy, žiadosti)	Publikačné minimum
UC Transformácia distribúcií datasetu Zahraničné služobné cesty	Publikačné minimum

## 6 Návrh riešenia

### 6.1 Použitie LinkedPipes

#### 6.1.1 Publikačné minimum MIRRI - Pipeline (CSV->RDF)



Príklad:

**Vstup:** tabularizovaný CSV dataset kontaktných miest MIRRI

**Výstup:** RDF dataset kontaktných miest

SPARQL construct ⓘ ✕

CONFIGURATION INHERITANCE GENERAL HIERARCHY

SPARQL CONSTRUCT query

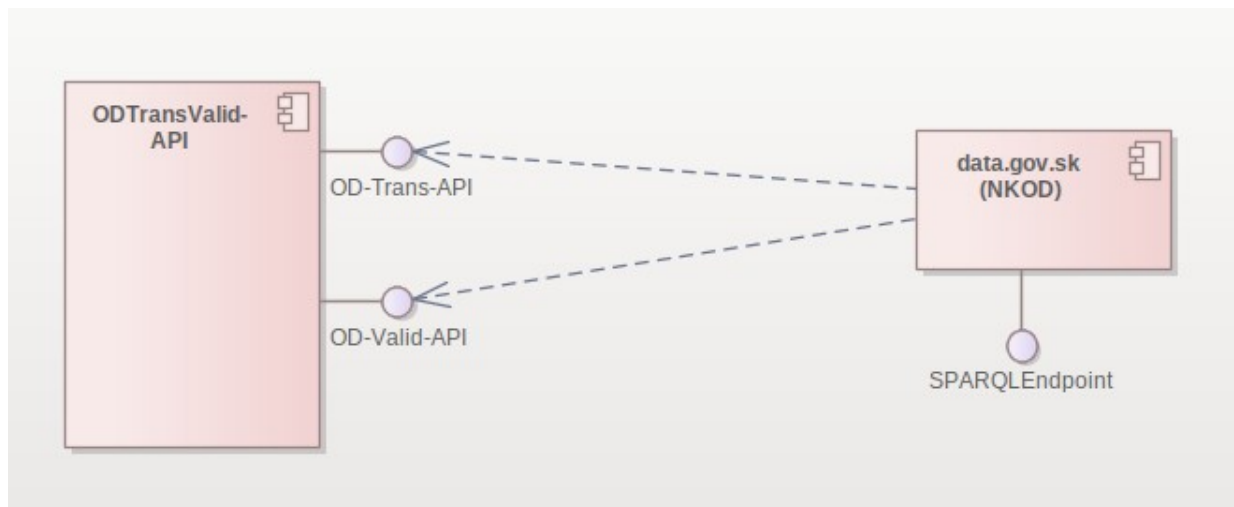
```

1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX dcat: <http://www.w3.org/ns/dcat#>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5 PREFIX dct: <http://purl.org/dc/terms/>
6 PREFIX schema: <http://schema.org/>
7 PREFIX cv: <http://data.europa.eu/m8g/>
8 PREFIX locn: <http://www.w3.org/ns/locn#>
9 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
10 construct
11 {
12   <https://data.gov.sk/dataset/mirri-kontaktne-miesta-2021-04-20> rdf:type dcat:Dataset .
13   <https://data.gov.sk/dataset/mirri-kontaktne-miesta-2021-04-20> dct:title "MIRRI Kontaktné miesta (2021-04-20)"@sk .
14   <https://data.gov.sk/dataset/mirri-kontaktne-miesta-2021-04-20> dct:created "2021-04-20"^^xsd:date .
15   <https://data.gov.sk/dataset/mirri-kontaktne-miesta-2021-04-20> dct:publisher <https://data.gov.sk/id/legal-subject/50349287> .
16
17   ?uri rdf:type cv:ContactPoint .
18   ?uri skos:prefLabel ?skLabel .
19   ?uri schema:provider ?provider .
20   ?uri schema:email ?email .
21   ?uri schema:phone ?phone .
22   ?uri schema:url ?url .
23   ?uri locn:address ?address .
24   ?uri locn:fullAddress ?fullAddress .
25   ?uri rdfs:isDefinedBy <https://data.gov.sk/dataset/mirri-kontaktne-miesta-2021-04-20> .
26 }
27 where
28 {
29   ?contactPoint <file://mirri-contactpoints.csv#id> ?id .
30   ?contactPoint <file://mirri-contactpoints.csv#name> ?name .
31   ?contactPoint <file://mirri-contactpoints.csv#organization> ?organization .
32   optional { ?contactPoint <file://mirri-contactpoints.csv#email> ?email . }
33   optional { ?contactPoint <file://mirri-contactpoints.csv#addressId> ?addressId . }
34   optional { ?contactPoint <file://mirri-contactpoints.csv#fullAddress> ?fullAddress . }
35   optional { ?contactPoint <file://mirri-contactpoints.csv#phone> ?phone . }
36
37   bind (uri(concat("https://data.gov.sk/id/contact-point/50349287/",?id)) as ?uri)
38   bind (strlang(?name,"sk") as ?skLabel)
39   bind (uri(concat("https://data.gov.sk/id/legal-subject/",?organization)) as ?provider)
40   bind (uri(concat("https://data.gov.sk/id/physical-address/",?addressId)) as ?address)
41 }

```

[DISCARD CHANGES](#) [SAVE CHANGES](#)

## 6.2 Softvérová architektúra



## 7 Referencie

<https://linkedpipes.com/>