# 

**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](#__RefHeading___Toc1072_968028077)

[Obsah 2](#__RefHeading___Toc1074_968028077)

[1 Úvod 3](#__RefHeading___Toc1076_968028077)

[1.1 Použité pojmy a skratky 3](#__RefHeading___Toc1078_968028077)

[2 Definícia projektu OU 2.0 4](#__RefHeading___Toc1080_968028077)

[2.1 Komponenty OD2.0 4](#__RefHeading___Toc1082_968028077)

[3 Ciele PoC 6](#__RefHeading___Toc1084_968028077)

[3.1 Definícia cieľov PoC 6](#__RefHeading___Toc1086_968028077)

[4 Požiadavky 7](#__RefHeading___Toc1088_968028077)

[4.1 Katalóg požiadaviek 7](#__RefHeading___Toc1090_968028077)

[4.2 Aktéri 7](#__RefHeading___Toc1092_968028077)

[4.3 Prípady použitia 8](#__RefHeading___Toc1094_968028077)

[4.3.1 Subsystém Validácie distribúcií datasetov 8](#__RefHeading___Toc1096_968028077)

[4.3.2 Subsystém Transformácie distribúcií datasetov 9](#__RefHeading___Toc1098_968028077)

[6 Návrh riešenia 10](#__RefHeading___Toc1100_968028077)

[6.2 Softvérová architektúra 10](#__RefHeading___Toc1100_9680280771)

[7 Implementácia 11](#__RefHeading___Toc1111_968028077)

[8 Referencie 12](#__RefHeading___Toc1104_968028077)

# Úvod

Účelom tohto dokumentu je špecifikácia PoC kompomentu OD2.0 -

## 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- https://5stardata.info/en/ |
| 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 https://metais.vicepremier.gov.sk/publicspace?pageId=23990179 |
| 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 <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2020/78/20200501>), 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 |

# 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> ).

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

|  |  |  |
| --- | --- | --- |
| **PRIORITNÁ OS Operačného programu integrovaná infraštruktúra** | PO 7 | Informačná spoločnosť |
| **Tematický cieľ** | TC2 | Zlepšenie prístupu k informačným a komunikačným technológiám a zlepšenie ich využívania a kvality |
| **Investičná priorita** | 2c) | Posilnenie aplikácií IKT v rámci elektronickej štátnej správy, elektronického vzdelávania, elektronickej inklúzie, elektronickej kultúry a elektronického zdravotníctva |
| **Špecifické ciele** | 7.5 | Zlepšenie celkovej dostupnosti dát verejnej správy vo forme otvorených dát |
| **Kód intervencie** | 079 | Prístup k informáciám verejného sektora (vrátane elektronickej kultúry s otvoreným prístupom k údajom, digitálnych knižníc, digitálneho obsahu a elektronického cestovného ruchu) |
| **Merateľné ukazovatele** | 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 |

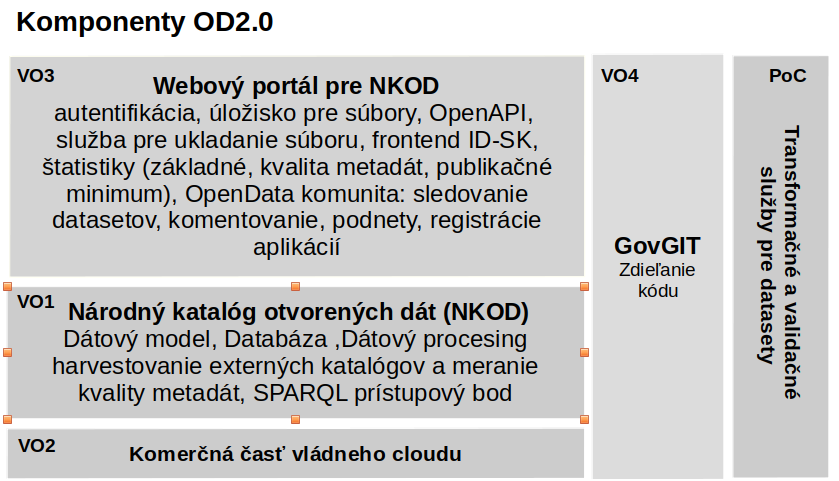
## Komponenty OD2.0

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 znovapoužitia existujúcich riešení. V tomto smere bolo odstúpené od zámeru vyvinutia úplne nového riešenia portálu s otvorenými údajmi a ako riešenie bolo vybraté znovapoužitie 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 znovapoužiitím 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.



# Ciele PoC

## 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 distríbúcií datasetov sa neimplementujú priamo v kóde, ale prostredníctvom použitia vhodného ETL systému

# Požiadavky

V tejto časti sú uvedené všetky požiadavky na predmet zákazky.

## Katalóg požiadaviek

Katalóg požiadaviek je nedeliteľnou súčasťou tohto Opisu predmetu dokumentu.

## Aktéri

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

## Prípady použitia

## Subsystém Validácie distribúcií datasetov

|  |  |  |
| --- | --- | --- |
| Pírpad použitia |  | **Subsystém** |
| 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 |

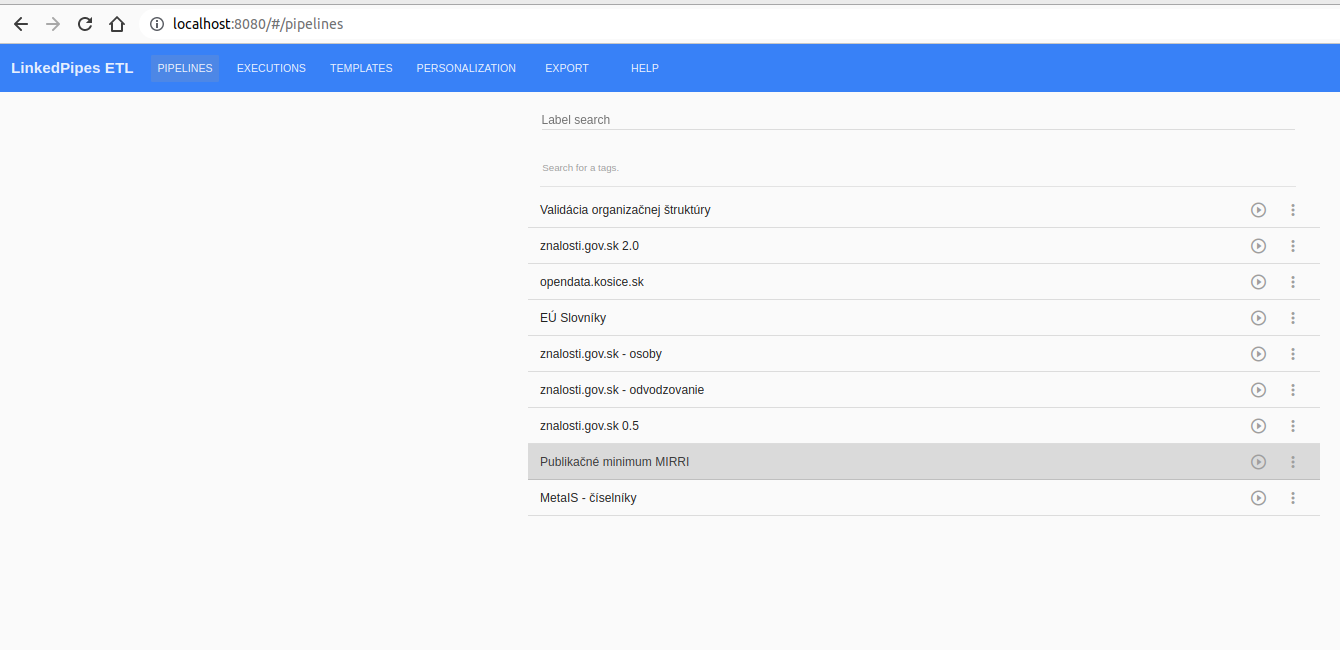
## Subsystém Transformácie distribúcií datasetov

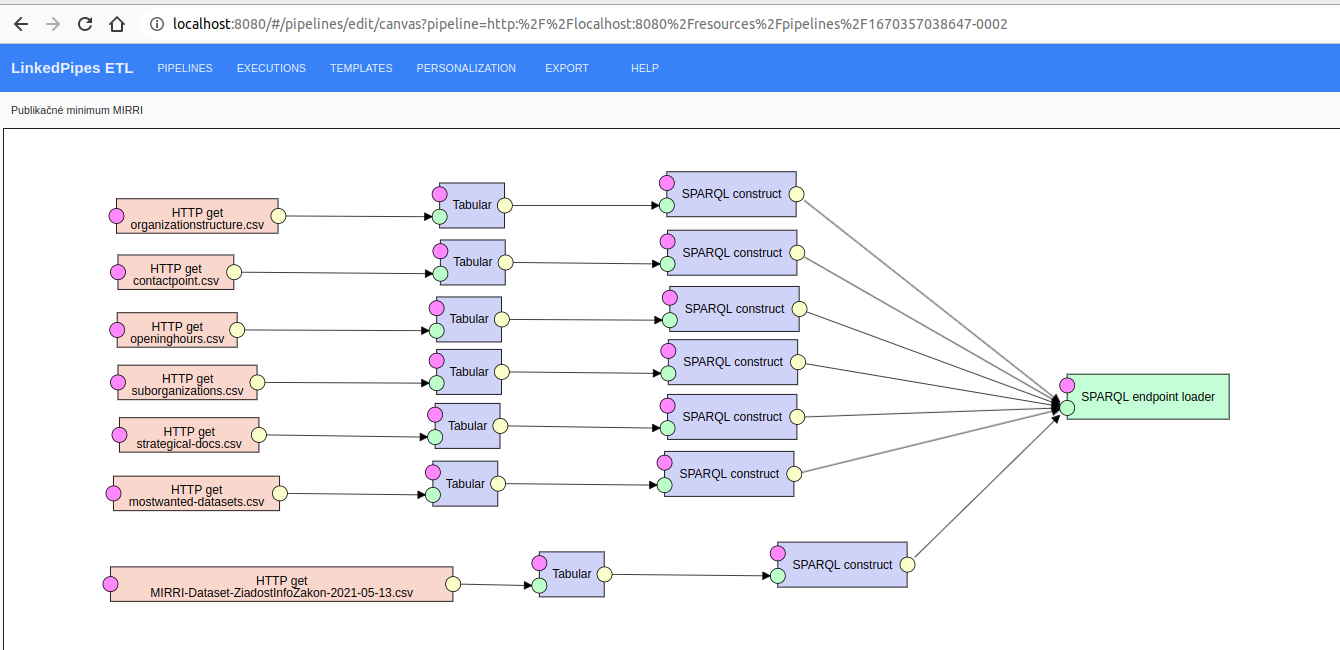
|  |  |  |
| --- | --- | --- |
| Pírpad použitia |  | **Subsystém** |
| UC Transformácia distribúcií datasetu Organizačná štruktúra | | Publikačné minimum |
| UC Transformácia distribúcií datasetu Kontaktné miesta | | Publikačné minimum |
| UC Transformácia distribúcií datasetu Úradné hodiny | | Publikačné 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)

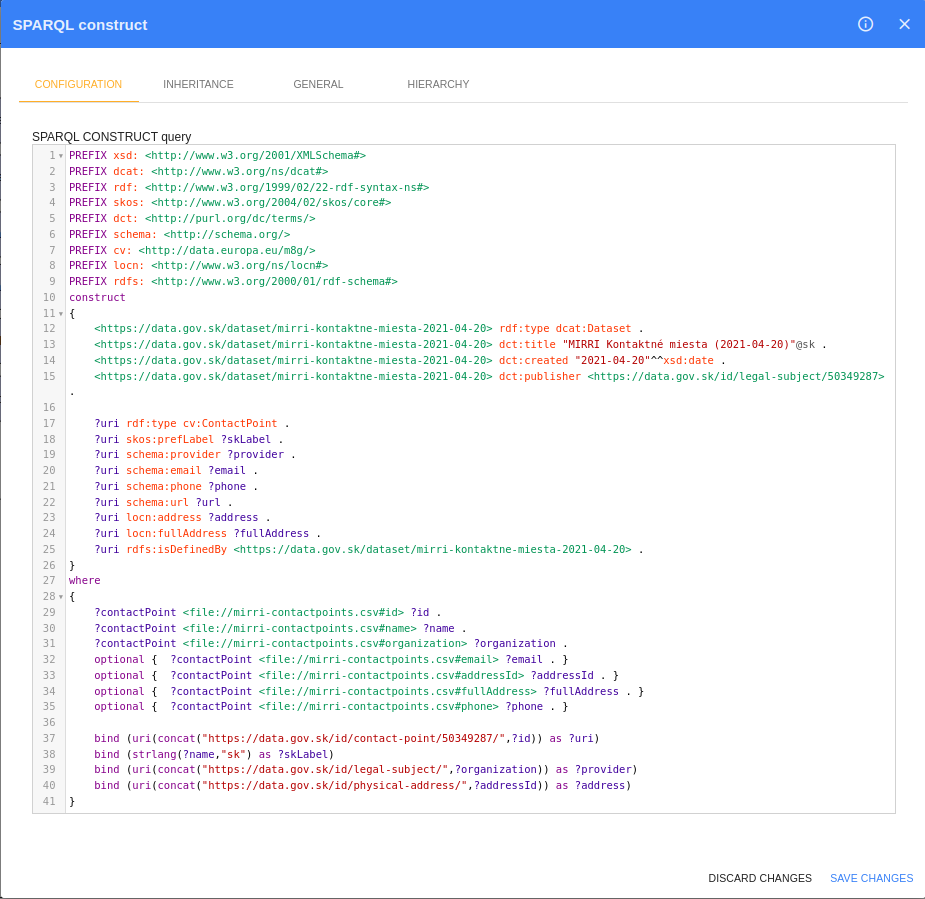




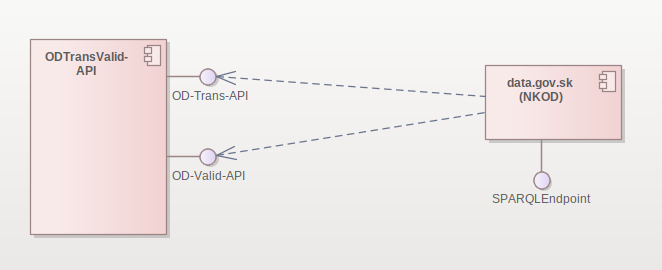
Príklad:

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

**Výstup**: RDF dataset kontaktných miest



## 6.2 Softvérová architektúra



# 7 Referencie

https://linkedpipes.com/