|  |  |
| --- | --- |
| **Detailný návrh riešenia (DNR)**   * Časť 1: Detailný návrh riešenia pre **Funkčné** a**Nefunkčné požiadavky**. * Časť 2: Detailný návrh riešenia pre **Technické požiadavky**. * Časť 3: Detailný návrh postupov a produktov pre fázy Testovanie a Nasadenie do produkcie.   **Národný katalóg otvorených dát NKOD**  **- Dátový model, databáza, dátový procesing, harvesting, meranie kvality, SPARQL Endpoint** | |
| Verzia dokumentu: | 13062022 0.3 |
| Projektový manažér - objednávateľa: | Mgr. Tibor Jedinák (MIRRI) |
| Projektový manažér - dodávateľa: | RNDr. Jakub Klimek, PhD. (UK-MATFYZ) |
| Predseda Riadiaceho výboru projektu: | Ing. Ján Hargaš |

Obsah

[1 Úvod 5](#_Toc106021603)

[1.1 Identifikácia a určenie 5](#_Toc106021604)

[1.1.1 Zmenový list 5](#_Toc106021605)

[1.1.2 Referencie na dokumenty 5](#_Toc106021606)

[1.1.3 Skratky 5](#_Toc106021607)

[1.1.4 Slovník pojmov 6](#_Toc106021608)

[1.1.5 Zoznam obrázkov 6](#_Toc106021609)

[1.1.6 Zoznam tabuliek 7](#_Toc106021610)

[2 Manažérske zhrnutie 7](#_Toc106021611)

[2.1 Co není součástí řešení 9](#_Toc106021612)

[2.2 Legislativní předpoklady 9](#_Toc106021613)

[Časť 1: Funkčná špecifikácia a detailný návrh riešenia 12](#_Toc106021614)

[3 Zámer riešenia 12](#_Toc106021615)

[3.1 Cieľ riešenia 13](#_Toc106021616)

[3.2 Popis navrhovaného riešenia 14](#_Toc106021617)

[3.3 Merateľné a výkonnostné ukazovatele (KPI) 14](#_Toc106021618)

[3.4 KATALÓG POŽIADAVIEK, Akceptačné kritériá 15](#_Toc106021619)

[4 Popis postupu analýzy 15](#_Toc106021620)

[5 Procesy podporované navhrovaným riešením 16](#_Toc106021621)

[5.1 Hledání informací o otevřených datech 16](#_Toc106021622)

[5.1.1 Hledání datasetů otevřených dat 16](#_Toc106021623)

[5.1.2 Vyhodnocení SPARQL dotazu 16](#_Toc106021624)

[5.2 Poskytování informací o otevřených datech 16](#_Toc106021625)

[5.2.1 Publikace otevřených dat 17](#_Toc106021626)

[5.2.2 Tvorba katalogizačního záznamu dle DCAT-AP-SK 17](#_Toc106021627)

[5.2.3 Vložení záznamu do NKOD 17](#_Toc106021628)

[5.2.4 Uložení registračního záznamu datasetu 17](#_Toc106021629)

[5.2.5 Vložení záznamu do LKOD 17](#_Toc106021630)

[5.2.6 Registrace LKOD do NKOD 17](#_Toc106021631)

[5.2.7 Uložení registračního záznamu LKODu 17](#_Toc106021632)

[5.3 Pravidelná aktualizace NKOD a měření kvality 18](#_Toc106021633)

[5.3.1 Načtení a kontrola registrací datových sad 18](#_Toc106021634)

[5.3.2 Načtení a kontrola registrací LKODů 18](#_Toc106021635)

[5.3.3 Harvestace záznamů z registrovaných LKODů 18](#_Toc106021636)

[5.3.4 Poskytnutí záznamů z LKOD dle DCAT-AP-SK 18](#_Toc106021637)

[5.3.5 Naplnění RDF databáze aktuálním stavem 19](#_Toc106021638)

[5.3.6 Zhodnocení souladu záznamů s DCAT-AP-SK 19](#_Toc106021639)

[5.3.7 Kontrola dostupnosti registrovaných zdrojů 19](#_Toc106021640)

[5.3.8 Poskytnutí odpovědi na kontrolní požadavky 19](#_Toc106021641)

[5.3.9 Generování reportů o datové kvalitě 19](#_Toc106021642)

[5.4 Monitorování kvality poskytování informací o otevřených datch 19](#_Toc106021643)

[5.4.1 Dotaz na datovou kvalitu 20](#_Toc106021644)

[5.4.2 Vyhodnocení SPARQL dotazu 20](#_Toc106021645)

[5.5 Zlepšování kvality poskytování otevřených dat 20](#_Toc106021646)

[5.5.1 Dotaz na datovou kvalitu 21](#_Toc106021647)

[5.5.2 Vyhodnocení SPARQL dotazu 21](#_Toc106021648)

[5.5.3 Vyřešení problémů s datovou kvalitou 21](#_Toc106021649)

[6 Biznis architektúra riešenia 21](#_Toc106021650)

[6.1 Aktéři 23](#_Toc106021651)

[6.1.1 Poskytovatel dat 23](#_Toc106021652)

[6.1.2 Návštevník 23](#_Toc106021653)

[6.1.3 Externí IS 23](#_Toc106021654)

[6.1.4 Administrátor NKOD (Správca NKOD) 23](#_Toc106021655)

[6.2 MapovAnIE Funkčných požIadaviek 23](#_Toc106021656)

[7 Popis dátových entít 24](#_Toc106021657)

[7.1 Diagram entít – Záznam dÁtové sady 24](#_Toc106021658)

[7.2 Diagram entít – MERANIA KVALITY 24](#_Toc106021659)

[7.2.1 Popis datových entit 25](#_Toc106021660)

[7.2.2 Měřené metriky kvality 26](#_Toc106021661)

[7.2.2.1 Dostupnost a dostupnost CORS 26](#_Toc106021662)

[7.2.2.2 Kvalita metadat 26](#_Toc106021663)

[8 Návrh riešenia dizajnu koncových služieb 28](#_Toc106021664)

[8.1 Identifikácia vlastníka dizajnu koncových služieb (UX) 29](#_Toc106021665)

[8.1.1 Pomenovanie vlastníka 29](#_Toc106021666)

[8.1.2 Kontakt na vlastníka 29](#_Toc106021667)

[8.2 Špecifikácia potrieb koncového používateľa 29](#_Toc106021668)

[8.3 Návrh mapy stránky (príp. toky používateľov) 29](#_Toc106021669)

[8.3.1 Návrh prototypov 29](#_Toc106021670)

[9 Prípady použitia (Use case model) 30](#_Toc106021671)

[9.1 Návrh vizuálov obrazoviek, podporených E2E procesom 31](#_Toc106021672)

[9.2 Používateľské roly – aktéri 31](#_Toc106021673)

[9.2.1 Poskytovatel dat 31](#_Toc106021674)

[9.2.2 Návštevník 32](#_Toc106021675)

[9.2.3 Externí IS 32](#_Toc106021676)

[9.2.4 Administrátor NKOD (Správca NKOD) 32](#_Toc106021677)

[9.2.5 Agent NKOD 32](#_Toc106021678)

[9.3 Všebecné prípady použitia 33](#_Toc106021679)

[9.3.1 UC Vykonanie dotazu cez prístupový bod 33](#_Toc106021680)

[9.3.2 UC Konfigurácia preddefinovaných dotazov 34](#_Toc106021681)

[9.3.3 UC Registrácia LKOD 34](#_Toc106021682)

[9.3.4 UC Harvestovanie metadát z lokálnych katalógov 34](#_Toc106021683)

[9.3.5 UC Meranie dátovej kvality 34](#_Toc106021684)

[9.3.6 UC Tvorba štatistických zostáv 34](#_Toc106021685)

[9.4 Sekvenčné diagramy (model správania systému) 35](#_Toc106021686)

[9.4.1 Harvestace lokálních katalogů 35](#_Toc106021687)

[9.4.2 Měření dostupnosti registrovaných zdrojů 35](#_Toc106021688)

[Časť 2: Technická špecifikácia a detailný návrh riešenia 36](#_Toc106021689)

[10. APLIKAČNÁ ARCHITEKTÚRA 36](#_Toc106021690)

[10.1 prehľad 36](#_Toc106021691)

[10.1.1 SPARQL Endpoint 36](#_Toc106021692)

[10.1.2 DCAT-AP-SK 36](#_Toc106021693)

[10.1.3 RDF úložiště 37](#_Toc106021694)

[10.1.4 Národní katalog otevřených dat (NKOD) 37](#_Toc106021695)

[10.1.5 LKOD (1,2,3) 37](#_Toc106021696)

[10.1.6 Metadatový procesor (LinkedPipes ETL) 37](#_Toc106021697)

[10.1.7 Původní data.gov.sk 37](#_Toc106021698)

[10.1.8 Filesystem 37](#_Toc106021699)

[10.1.9 Webová stránka 37](#_Toc106021700)

[10.1.10 Updater registrací 37](#_Toc106021701)

[10.2 KOMUNIKÁCIA MEDZI KOMPONENTAMI 38](#_Toc106021702)

[10.2.1 Komunikácia medzi UI a servisnou vrstvou 38](#_Toc106021703)

[10.2.2 Komunikácia medzi servisnou vrstvou a dátovou vrstvou 38](#_Toc106021704)

[10.2.3 Iné formu komunikácie v ráci riešenia 38](#_Toc106021705)

[10.3 KOMUNIKÁCIA MEDZI INTERNÝMI A EXTERNÝMI SYSTÉMAMI 39](#_Toc106021706)

[10.3.1 Zoznam a popis existujúcich integrácií 39](#_Toc106021707)

[10.3.2 Zoznam a popis navrhovaných integrácií 39](#_Toc106021708)

[11. FYZICKÁ ARCHITEKTÚRA 40](#_Toc106021709)

[11.1 prehľad 40](#_Toc106021710)

[11.1.1 Požadavky na běhové prostředí 40](#_Toc106021711)

[11.1.2 Konfigurace zdrojů (resources) 41](#_Toc106021712)

[11.1.3 Definice nasazení (Deployments) 41](#_Toc106021713)

[11.1.4 Definice sekundárního úložiště (PersistentVolumeClaim) 41](#_Toc106021714)

[11.1.5 Definice služeb (Service) 42](#_Toc106021715)

[11.1.6 Docker (Docker Image, Docker Container) 42](#_Toc106021716)

[11.1.7 Příchozí komunikace 43](#_Toc106021717)

[11.2 KOMUNIKÁCIA 43](#_Toc106021718)

[11.2.1 Komunikace pomocí sekundární paměti 43](#_Toc106021719)

[11.2.2 Komunikace pomocí síťového rozhraní 44](#_Toc106021720)

[11.3 CI/CD PROCESS 44](#_Toc106021721)

[11.4 KONTAJNERIZÁCIA A ORCHESTRÁCIA 44](#_Toc106021722)

[11.5 VYUŽITIE VLÁDNEHO CLOUDU (EZ) 45](#_Toc106021723)

[11.6 POPIS TESTOVACIEHO PROSTREDIA 45](#_Toc106021724)

[11.7 POPIS PRODUKČNÉHO PROSTREDIA 45](#_Toc106021725)

[12. SOFTVÉROVÉ LICENCIE 45](#_Toc106021726)

[13. ZÁLHOVANIE A OBNOVA RIEŠENIA 45](#_Toc106021727)

[13.1 POPIS RIEŠENIA PRE ZÁLOHOVANIE a OBNOVu 45](#_Toc106021728)

[14. ZABEZPEČENIE DOSTUPNOSTI 45](#_Toc106021729)

[14.1 POPIS RIEŠENIA NA DOSTUPNOSŤ SYSTÉMU IS 45](#_Toc106021730)

[15. BEZPEČNOSŤ, AUTORITY MANAŽMENT 46](#_Toc106021731)

[15.1 ZABEZPEČENIE KOMUNIKÁCIE A INTEGRITY 46](#_Toc106021732)

[15.2 RIADENIE PRÍSTUPU (AUDIT TRAILS – SLEDOVANIE LOGOVANIA DO SYSTÉMU) 46](#_Toc106021733)

[15.3 DEFINOVANIE POUŽÍVATEĽSKÝCH ROLÍ 46](#_Toc106021734)

[15.4 METODIKA MANAŽMENTU POUŽÍVATEĽOV 46](#_Toc106021735)

[16. KONVERZIE DÁT, MIGRÁCIA DÁT, DÁTOVÝ MODEL A NÁVRH RIEŠENIA 47](#_Toc106021736)

[16.1 POSTUP VYKONANIA MIGRÁCIE DÁT 47](#_Toc106021737)

[16.2 OVERENIE MIGRÁCIE DÁT 47](#_Toc106021738)

[Časť 3: Detailný návrh postupov a produktov pre fázy Testovanie a Nasadenie do produkcie. 48](#_Toc106021739)

[16.3 FÁZY PROJEKTU A ČASOVÝ HARMONOGRAM – KONSOLIDOVANÝ PREHĽAD 48](#_Toc106021740)

[16.3.1 Fáza - Analýzy a Dizajnu 48](#_Toc106021741)

[16.3.2 Fáza - Implementácie a Testovania 48](#_Toc106021742)

[16.3.3 Fáza - Nasadenia a Post-Implementačnej podpory – príprava a nasadenie do prevádzky 48](#_Toc106021743)

[16.3.4 Ukončenie projektu 48](#_Toc106021744)

[17. ZÁVISLOSTI A NÁVRH RIEŠENIA 48](#_Toc106021745)

[17.1 ZÁVISLOSTI (PROJEKTOVÉ/MEDZIPROJEKTOVÉ) 49](#_Toc106021746)

[18. DOKUMENTÁCIA 49](#_Toc106021747)

[19. OTVORENÉ OTÁZKY 50](#_Toc106021748)

[20. PRÍLOHY 50](#_Toc106021749)

# Úvod

## Identifikácia a určenie

### Zmenový list

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Verzia | Dátum | Autor | Dôvod zmeny | Kapitola, Bod |
| 0.3 | 2022-06-13 | Jakub Klímek | Iniciální verze |  |
|  |  |  |  |  |
|  |  |  |  |  |

### Referencie na dokumenty

|  |  |
| --- | --- |
| Identikácia | Názov dokumentu |
|  |  |
|  |  |
|  |  |

### Skratky

|  |  |
| --- | --- |
| Skratka | Vysvetlenie |
| NKOD | Národní katalog otevřených dat |
| LKOD | Lokální katalog otevřených dat |
| DCAT-AP-SK | Data Catalog Vocabulary Application Profile Slovakia |
| EDP | Oficiálny portál pre údaje o Európe (<https://data.europa.eu>) |
| LP-ETL | LinkedPipes ETL |
| RDF | Resource Description Framework |
| CORS | Cross-Origin Resource Sharing |
| HTTP | Hyper-Text Transfer Protocol |
| VPN | Virtual Private Networking |
| SSH | Secure Shell |
| HTTP | Hyper-text Transfer Protocol |

### Slovník pojmov

|  |  |
| --- | --- |
| Pojem | Vysvetlenie |
| Národní katalog otevřených dat (NKOD) | Hlavní komponenta tohoto projektu. Jedná se o metadata datasetů otevřených dat zahrnující jejich popis a způsob přístupu k otevřeným datům. Metadata jsou přístupná přes SPARQL endpoint, nad kterým lze do budoucna stavět uživatelské rozhraní. |
| Lokální katalog otevřených dat (LKOD) | Pokud některý poskytovatel otevřených dat nechce registrační záznamy spravovat přímo v NKOD, může provozovat vlastní Lokální katalog otevřených dat, ve kterém si spravuje metadata dle svých potřeb. Aby se metadata z něj dostala do Národního katalogu otevřených dat, musí mít každý LKOD rozhraní dle specifikace DCAT-AP-SK, které je pak registrováno v NKOD. |
| Harvestace | Proces kopírování metadat datasetů z LKOD do NKOD (nebo také z NKOD do EDP). V NKOD běží pravidelné procesy, které přes rozhraní jednotlivých LKOD zrcadlí jejich metadatové záznamy do své databáze, aby byly přístupné na jednom místě. |
| DCAT-AP-SK | Specifikace položek metadatového záznamu datasetu a rozhraní LKOD |
| Pipeline | Definovaný datově transformační proces v nástroji LinkedPipes ETL |
| LinkedPipes ETL | Nástroj, ve kterém jsou definovány datově transformační procesy spravující NKOD a počítající ukazatele kvality metadatových záznamů. |
| SPARQL | Dotazovací jazyk nad daty v RDF |
| SPARQL Endpoint | Webová služba běžící nad RDF databází, přijímající dotazy v jazyce SPARQL, a vracející výsledky vyhodnocení dotazu nad daty v databázi. |

### Zoznam obrázkov

[Obrázok 1 Komponenty projektu Otvorené údaje 2.0 7](#_Toc106018772)

[Obrázok 2 Hledání informací o otevřených datech 15](#_Toc106018773)

[Obrázok 3 Poskytování informací o otevřených datech 16](#_Toc106018774)

[Obrázok 4 Pravidelná aktualizace NKOD a měření kvality 17](#_Toc106018775)

[Obrázok 5 Monitorování kvality poskytování informací o otevřených datech 19](#_Toc106018776)

[Obrázok 6 Zlepšování kvality poskytování otevřených dat 20](#_Toc106018777)

[Obrázok 7 Byznys architektura NKOD 21](#_Toc106018778)

[Obrázok 8 Datový model pro záznam datasetu 23](#_Toc106018779)

[Obrázok 9 Datový model pro měření datové kvality 24](#_Toc106018780)

[Obrázok 10 Obrazovka pro zadání SPARQL dotazu 29](#_Toc106018781)

[Obrázok 11 Obrazovka pro zadání SPARQL dotazu 30](#_Toc106018782)

[Obrázok 12 Aplikační architektura 35](#_Toc106018783)

[Obrázok 13 Technologická architektura 39](#_Toc106018784)

[Obrázok 14 Úložiště na technologické vrstvě 43](#_Toc106018785)

[Obrázok 15 Komunikace na technologické vrstvě 44](#_Toc106018786)

### Zoznam tabuliek

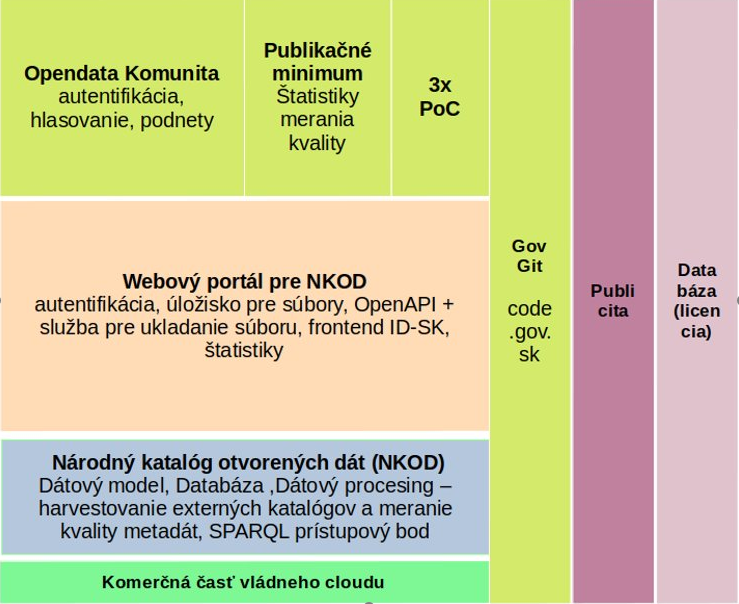
[Tabuľka 1 Analytické produkty 14](#_Toc106018787)

[Tabuľka 2 Mapování funkčních požadavků na byznys procesy 22](#_Toc106018788)

[Tabuľka 3 Přehled metrik kvality metadat 25](#_Toc106018789)

# Manažérske zhrnutie

Projekt „Národný katalóg otvorených dát (NKOD)“ tvoří jednu z komponent projektu „Otvorené údaje 2.0“.



Obrázok 1 Komponenty projektu Otvorené údaje 2.0

V projektu NKOD bude vybudována databáze pro metadatové záznamy datasetů otevřených dat poskytovaných různými institucemi veřejné správy, která bude poskytovat SPARQL endpoint pro dotazování. V databázi se budou zrcadlit metadataové záznamy datasetů registrovaných jednotlivě přímo v NKOD a záznamy pocházející z Lokálních katalogů otevřených dat provozovaných přímo poskytovateli dat. Metadatové záznamy budou odpovídat specifikaci DCAT-AP-SK, která bude také vytvořena v rámci projektu. Databáze NKOD bude tvořena pravidelně, např. denně. Po každém vytvoření databáze NKOD bude i zhodnocena kvalita metadatových záznamů vzhledem k DCAT-AP-SK a dostupnost registrovaných zdrojů. Na základě naměřených hodnot budou vygenerovány reporty obsahující zjištěné skutečnosti.

V rámci projektu proběhne i migrace a čištění dat ze stávajícího portálu data.gov.sk, který bude v rámci projektu Otvorené údaje 2.0 nahrazen.

Cíle projektu NKOD jsou:

1. Vytvořit specifikaci DCAT-AP-SK pro registrační záznamy datasetů a rozhraní LKOD,
2. vybudovat backend Národního katalogu otevřených dat, který bude harvestovat LKODy a přijaté registrační záznam datových sad a ukládat je do RDF databáze,
3. vystavit rozhraní SPARQL endpoint pro dotazování nad metadaty v katalogu obsaženými včetně webové stránky umožňující zaslání dotazu a zobrazení výsledků dotazu,
4. nad metadaty spočítat statistiky kvality registrovaných záznamů a indikátory dostupnosti registrovaných zdrojů,
5. zmigrovat stávající obsah katalogu data.gov.sk dle vytvořené specifikace.

Projekt bude vybudován na základě řešení, které je již v provozu v České republice jako český Národní katalog otevřených dat.

## Co není součástí řešení

V projektu je tvořena databáze NKOD a datové procesy databázi tvořící a zpracovávající. Součástí řešení projektu není uživatelské rozhraní mimo stránku zpřístupňující SPARQL endpoint. Zejména tedy součástí není

* Uživatelské rozhraní pro zobrazování metadatových záznamů mimo SPARQL endpoint
* Uživatelské rozhraní pro pořizování a správu metadatových záznamů a registrací lokálních katalogů
* Uživatelské rozhraní pro zobrazování naměřených indikátorů datové kvality a dostupnosti zdrojů mimo SPARQL endpoint

## Legislativní předpoklady

|  |
| --- |
| Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 78/2020 Z. z. o štandardoch pre informačné technológie verejnej správy |
| Zákon č. č. 305/2013 Z. z. o elektronickej podobe výkonu pôsobnosti orgánov verejnej moci a o zmene a doplnení niektorých zákonov (zákon o e-Governmente) v znení neskorších predpisov |
| Metodika Jednotný dizajn manuál elektronických služieb verejnej správy (dostupný na <https://idsk.gov.sk/>) |
| Metodika Tvorba používateľsky kvalitných digitálnych služieb verejnej správy (dostupný na <https://www.vicepremier.gov.sk/sekcie/oddelenie-behavioralnych-inovacii/index.html>) |
| Nariadenie Európskeho parlamentu a Rady (EÚ) č. 910/2014 z 23. júla 2014 o elektronickej identifikácii a dôveryhodných službách pre elektronické transakcie na vnútornom trhu a o zrušení smernice 1999/93/ES |
| Nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov) |
| Zákon č. 18/2018 Z. z. o ochrane osobných údajov a o zmene a doplnení niektorých zákonov v znení neskorších predpisov |
| Vyhláška Úradu na ochranu osobných údajov č. 158/2018 Z. z. o postupe pri posudzovaní vplyvu na ochranu osobných údajov |
| Zákon č. 69/2018 Z. z. o kybernetickej bezpečnosti a o zmene a doplnení niektorých zákonov v znení neskorších predpisov |
| Zákon č. 272/2016 Z. z. o dôveryhodných službách pre elektronické transakcie na vnútornom trhu a o zmene a doplnení niektorých zákonov  (zákon o dôveryhodných službách) v znení neskorších predpisov. |
| Zákon č. 95/2019 Z. z. o informačných technológiách vo verejnej správe a o zmene a doplnení niektorých zákonov. Je požadované, aby jednotlivé IS boli vyvinuté formou open-source (EUPL licenčný model). |
| Metodika merania dátovej kvality vo verejnej správe MIRRI (ÚPVII) (dostupná na <https://datalab.digital/wp-content/uploads/Metodika-merania-dátovej-kvality-vo-verejnej-správe.pdf> ). |
| Metodické usmernenie Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 3639/2019/oDK-1 o postupe zaraďovania referenčných údajov do zoznamu referenčných údajov vo väzbe na referenčné registre a vykonávania postupov pri referencovaní (dostupný na <https://datalab.digital/wp-content/uploads/Metodické-usmernenie-ÚPVII-č.-3639-2019-oDK-1-FINAL-1.pdf>). |
| Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 85/2020 Z. z. o riadení projektov |
| Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 179/2020 Z. z., ktorou sa ustanovuje spôsob kategorizácie a obsah bezpečnostných opatrení informačných technológií verejnej správ |
| Zákon č. 211/2000 Z. z. o slobodnom prístupe k informáciám a o zmene a doplnení niektorých zákonov (zákon o slobode informácií) v znení neskorších predpisov (úprava opakovaného použitia informácií verejného sektora) |
| Smernica Európskeho parlamentu a Rady 2019/1024 z 20. júna 2019 o otvorených dátach a opakovanom použití informácií verejného sektora |
| Zákon č. 215/2004 Z. z. o ochrane utajovaných skutočností v znení neskorších predpisov |
| Zákona č. 541/2004 Z. z. o mierovom využívaní jadrovej energie (atómový zákon) a o zmene a doplnení niektorých zákonov v znení neskorších predpisov. |
| Zákon č. 162/1995 Z. z. o katastri nehnuteľností a o zápise vlastníckych a iných práv k nehnuteľnostiam (katastrálny zákon) v znení neskorších predpisov. |
| Zákon č. 540/2001 Z. z. o štátnej štatistike v znení neskorších predpisov |
| Zákon č. 185/2015 Z. z. Autorský zákon v znení neskorších predpisov |
| Zákon č. 177/2018 Z.z. o niektorých opatreniach na znižovanie administratívnej záťaže využívaním informačných systémov verejnej správy a o zmene a doplnení niektorých zákonov (zákon proti byrokracii) v znení neskorších predpisov |
| Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 438/2019 Z.z., ktorou sa vykonávajú niektoré ustanovenia zákona o e-Governmente |
| Vyhláška Ministerstva investícií, regionálneho rozvoja a informatizácie Slovenskej republiky č. 70/2021 o zaručenej konverzii z 12. februára 2021 |
| Vyhláška Ministerstva vnútra Slovenskej republiky č. 29/2017 Z.z., ktorou sa ustanovujú podrobnosti o alternatívnom autentifikátore |
| Vyhláška Úradu podpredsedu vlády Slovenskej republiky pre investície a informatizáciu č. 85/2018 Z.z., ktorou sa ustanovujú podrobnosti o spôsobe vyhotovenia a náležitostiach listinného rovnopisu elektronického úradného dokumentu |
| Vyhláška Ministerstva financií Slovenskej republiky č. 25/2014 Z. z. o integrovaných obslužných miestach a podmienkach ich zriaďovania, označovania, prevádzky a o sadzobníku úhrad v znení neskorších predpisov |
| Zákon č. 351/2011 Z.z o elektronických komunikáciách v znení neskorších predpisov |
| Zákon č. 300/2005 Z.z. Trestný zákon v znení neskorších predpisov |
| Smernica č. 7/2019 o riešení Bezpečnostných incidentov Vládnou jednotkou CSIRT |
| Vyhláška Národného bezpečnostného úradu č. 166/2018 Z.z. o technickom, technologickom a personálnom vybavení jednotky pre riešenie kybernetických bezpečnostných incidentov |
| Vyhláška Národného bezpečnostného úradu č. 436/2019 Z.z. o audite kybernetickej bezpečnosti a znalostnom štandarde audítora |
| Zákon č. 343/2015 Z.z. o verejnom obstarávaní a o zmene a doplnení niektorých zákonov v znení neskorších predpisov |
| Zákon č. 452/2021 Z. z. o elektronických komunikáciách z 24.novembra 2021 |
| Zákon č. 45/2011 Z.z. o kritickej infraštruktúre v znení neskorších predpisov |
| Vyhláška č. 164/2018 Z. z. – Vyhláška NBÚ SR, vyhláška o IKPS, ktorou sa určujú identifikačné kritériá prevádzkovanej služby (kritériá základnej služby) v platnom znení. |
| Vyhláška č. 165/2018 Z. z. – Vyhláška NBÚ SR, vyhláška o KBI, ktorou sa určujú identifikačné kritériá pre jednotlivé kategórie závažných kybernetických bezpečnostných incidentov a podrobnosti hlásenia kybernetických bezpečnostných incidentov v platnom znení. |
| Vyhláška č. 362/2018 Z. z. – Vyhláška NBÚ SR, vyhláška o OBO, ktorou sa ustanovuje obsah bezpečnostných opatrení, obsah a štruktúra bezpečnostnej dokumentácie a rozsah všeobecných bezpečnostných opatrení v platnom znení. |
| Zákon č. 3/2010 Z.z. o národnej infraštruktúre pre priestorové údaje v znení neskorších predpisov |

Časť 1: Funkčná špecifikácia a detailný návrh riešenia

# Zámer riešenia

Otvorené údaje 2.0 jako celek

V predchádzajúcom projekte eDemokracia sa podarilo vytvoriť centrálnu platformu pre zdieľanie otvorených údajov, ktorá je založená na platforme CKAN. Projekt eDemokracia zabezpečuje služby asistencie pre poskytovateľov údajov pri sprístupňovaní otvorených údajov najmä v nasledovných oblastiach:

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

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

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

Rovnako ako ide vývoj tejto oblasti dopredu, musí aj MIRRI ako správca centrálnych komponentov reagovať na najnovšie trendy a požiadavky komunity využívajúcej otvorené údaje. Momentálne riešenie obsahuje problémové oblasti a to v súvislosti s efektívnym zabezpečením využívania mechanizmov správy a zdieľania údajov, zabezpečenia automatizovaného generovania datasetov otvorených údajov, využitím nástrojov a služieb pre publikovanie otvorených údajov, implementáciou princípov pre prepojené údaje (Linked Data). V súčasnosti medzi najvýznamnejšie obmedzenia v kvalite publikovaných otvorených údajov patrí:

* chýbajúca štandardizovaná strojovo spracovateľná reprezentácia metadát datasetov prostredníctvom štandardu DCAT,
* chýbajúce rozhranie pre dopytovanie metadát otvorených údajov používateľom alebo systémom. t.j. chýbajúci SPARQL Endpoint,
* chýbajúca funkcionalita harvestovania iných lokálnych údajov portálov s otvorenými údajmi,
* veľké množstvo otvorených údajov je zverejnených v nízkej kvalite z pohľadu strojového spracovania na úrovni 1★-2★ (PDF, HTML, XLS a podobne). Navyše dôležité otvorené údaje sú často vypublikované bez identifikátorov, kódov, alebo s inými kódmi na totožné veci na základe čoho je nie je možné s údajmi okamžite pracovať, ale je nutné ich konsolidovať,
* samotná kvalita údajov je nízka (dáta nie sú vyčistené, väzby medzi údajmi sú v mnohých prípadoch nejednoznačné, duplicitné a pod.),
* merania kvality zverejnených otvorených údajov prebiehajú manuálne, čo je extrémne náročné. Chýba napr. podpora merania kvality metadát otvorených údajov, ako napr. metodiky centrálneho dátového portálu Európskej Únie MQA, <https://www.europeandataportal.eu/mqa>.

Primárnym zámerom je využiť ekonomický potenciál otvorených údajov verejného sektora, pričom sú očakávané výsledky:

* vďaka využívaniu otvorených údajov sa výrazne zvýši miera transparentnosti verejnej správy, čím vznikne pozitívna spätná väzba smerujúca k vyššej efektivite fungovania verejnej správy,
* vznikne dostatočne silné odvetvie pracujúce s otvorenými údajmi pričom vzniknú kompetencie v progresívnych oblastiach informačných technológií, ako sú dátové analýzy,
* inovatívne použitie otvorených údajov vygeneruje pridanú hodnotu pre používateľov, ktorá sa premietne aj do ekonomických prínosov.

Národný katalóg otvorených dát NKOD – táto časť projektu

Ve stávajícím řešení data.gov.sk postaveném na platformě CKAN nejsou metadata datasetů reprezentována dle evropského standardu DCAT-AP a nelze se nad nimi volně dotazovat pomocí standardního dotazování pomocí SPARQL. Taktéž nejsou podoporovány lokální katalogy otevřených dat, takže každý jeden dataset je třeba registrovat přímo na data.gov.sk, což je nepohodlné a v důsledku vede ke snížení kvality registrovaných metadat. Nelze tak ani hovořit o měření kvality metadat, ani o pohodlném používání databáze metadatových záznamů třetími stranami, např. pro vývoj aplikací.

Nově budovaný Národní katalog otevřených dat (NKOD) bude datově orientovaný. V jeho středu je datová specifikace pro záznamy datasetů a definice rozhraní pro lokální katalogy DCAT-AP-SK. Kolem této specifikace jsou pak postavené datové procesy shromažďující metadatové záznamy v jednotném formátu a umožňující k nim volný přístup v podobě SPARQL endpointu – strojového rozhraní. To v dalších částech projektu OD2.0 umožní vznik jak uživatelského rozhraní postaveného na dnešních webových standardech, tak standardizovaný přístup k záznamům aplikacím třetích stran.

## Cieľ riešenia

Otvorené údaje 2.0 jako celek

Hlavný cieľ projektu OD2.0 je poskytovanie programového vybavenia pre verejnú správu, aby mohla plniť svoje zákonné povinnosti v oblasti zverejňovanie otvorených údajov, a to efektívnym spôsobom v súlade s aktuálnymi štandardmi, odporúčaniami a trendmi sprístupňovania otvorených údajov v Európskej Únii. Nový portál umožní strojovo spracovávať Národný katalóg otvorených údajov (čítanie, zápis, analýza, dotazovanie), pričom sa definuje jednotný model metadát otvorených údajov DCAT-AP-SK. Portál súčasne bude poskytovať funkcionalitu automatizovanej katalogizácie externých otvorených údajov a súčasne bude pravidelne vykonávať meranie kvality zverejnených otvorených údajov. Vo výsledku bude nový portál podobný ako portály iných členských krajín EU, najmä ako data.gov.uk, datos.gob.es, data.gov.cz a podobne.

Národný katalóg otvorených dát NKOD – táto časť projektu

Cíle části NKOD jsou definovat specifikaci DCAT-AP-SK standardizující podobu metadatového záznamu datasetu a rozhraní lokálních katalogu. Dále implementovat datové procesy sbírající metadatové záznamy z lokálních katalogů a umožňující registrace datasetů přímo v NKOD, které výsledný katalog uloži do RDF databáze a zpřístupní přes její SPARQL endpoint veřejnosti. Nakonec je cílem také počítat indikátory kvality metadatových záznamů a dostupnosti registrovaných zdrojů dat a zpřístupňovat výsledky v podobě reportů. Stávající metadata z data.gov.sk budou zmigrovány dle DCAT-AP-SK.

## Popis navrhovaného riešenia

Vlastníkem řešní a dat bude Datová kancelář MIRRI. Řešení bude nasazeno do komerčního Kubernetes cloudu. Řešení má následující části.

Specifikace DCAT-AP-SK

Specifikace DCAT-AP-SK bude založena na české Otevřené formální normě „Rozhraní katalogů otevřených dat: DCAT-AP-CZ“ (<https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/>), která bude upravena podle slovenských specifik. Nástrojem pro reprezentaci specifikace bude Bikeshed (<https://tabatkins.github.io/bikeshed/>). Oblast, kterou je třeba upravit zejména, se týká specifikace podmínek užití (licence) otevřených dat. Specifikace bude vystavena webu.

RDF databáze a SPARQL endpoint

Jako RDF databáze byla objednatelem vybrána implementace Ontotext GraphDB (<https://graphdb.ontotext.com/> - prozatím ve verzi Free). Přístup k jejímu SPARQL endpointu bude umožněn z Webu veřejnosti včetně aplikací třetích stran, a také z webové stránky umožňující formulaci SPARQL dotazu a zobrazení výsledků dotazu uživatelům.

Datové procesy

Procesy harvestující lokální katalogy otevřených dat, zpracovávající jednotlivé registrační záznamy a nahrávající data do RDF databáze budou implementovány jako pipeline v nástroji LinkedPipes ETL, který stejnému účelu slouží i v českém Národním katalogu otevřených dat.

Měření datové kvality

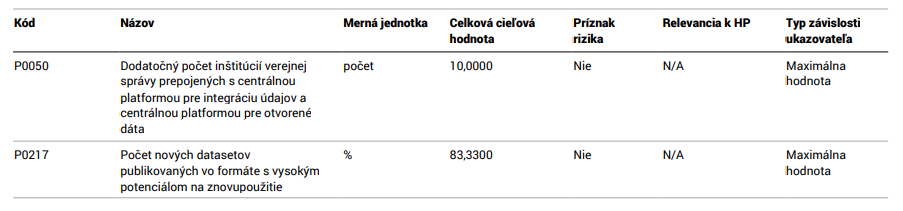
Procesy měřící datovou kvalitu budou také implementovány jako pipeline v nástroji LinkedPipes ETL, který stejnému účelu slouží i v českém Národním katalogu otevřených dat. Výsledky měření kvality metadat a dostupnosti datových zdrojů sbíraných obdobně jako v metodice MQA <https://data.europa.eu/mqa> pak budou reprezentovány dle webového standardu Data Quality Vocabulary (<https://www.w3.org/TR/vocab-dqv/>) a také ve formátu CSV.

Migrace data.gov.sk

Součástí projektu bude i migrace metadatových záznamů ze stávajícího katalogu data.gov.sk, která může zahrnovat jejich transformaci tak, aby odpovídali specifikaci DCAT-AP-SK tam, kde to bude možné. Migrace bude rovněž implementována jako pipeline v LinkedPipes ETL.

## Merateľné a výkonnostné ukazovatele (KPI)

KPI jsou stanoveny pro celý projekt OD2.0, jehož je NKOD pouze jednou součástí. Vyhodnocení KPI bude tedy možné až po dokončení všech součastí OD 2.0.



## KATALÓG POŽIADAVIEK, Akceptačné kritériá

Přílohou tohoto dokumentu je KATALOG\_POZIADAVIEK\_OD2.0-NKOD.xlsx obsahující požadavky na řešení. Návrh jejich řešení je obsahem tohoto dokumentu, a slouží také jako podklad pro plán testů.

Klasifikace vad má následující definice:

* **A - Kritická** - Kritické vady, ktoré spôsobia úplné zlyhanie diela ako celku a nie je možné používať ani jednu jeho časť.
* **B - Vysoká** - Vady a nedostatky, ktoré zapríčinia čiastočné zlyhanie diela a neumožňuje používať časť diela.
* **C - Stredná** - Vady a nedostatky, ktoré spôsobia čiastočné obmedzenia používania diela.
* **D – Nízka** - Kozmetické a drobné vady.

Z katalogu požadavků vyplývají následující akceptační kritéria:

* A - Kritická - **0 vád**.
* B – Vysoká - **maximálne 1 vada**.
* C - Stredná - **maximálne 2 vady**.¨
* D – Nízka - **nie je obmedzené**.

# Popis postupu analýzy

Analýza na projektu vychází zejména ze specifikace díla, požadavků a případů užití specifikovaných již v zadání veřejné zakázky.

Podrobnosti byly dopřesněny s objednatelem v rámci řady online meetingů.

Přehled analytických produktů a jejich zohlednění v DNR:

Tabuľka 1 Analytické produkty

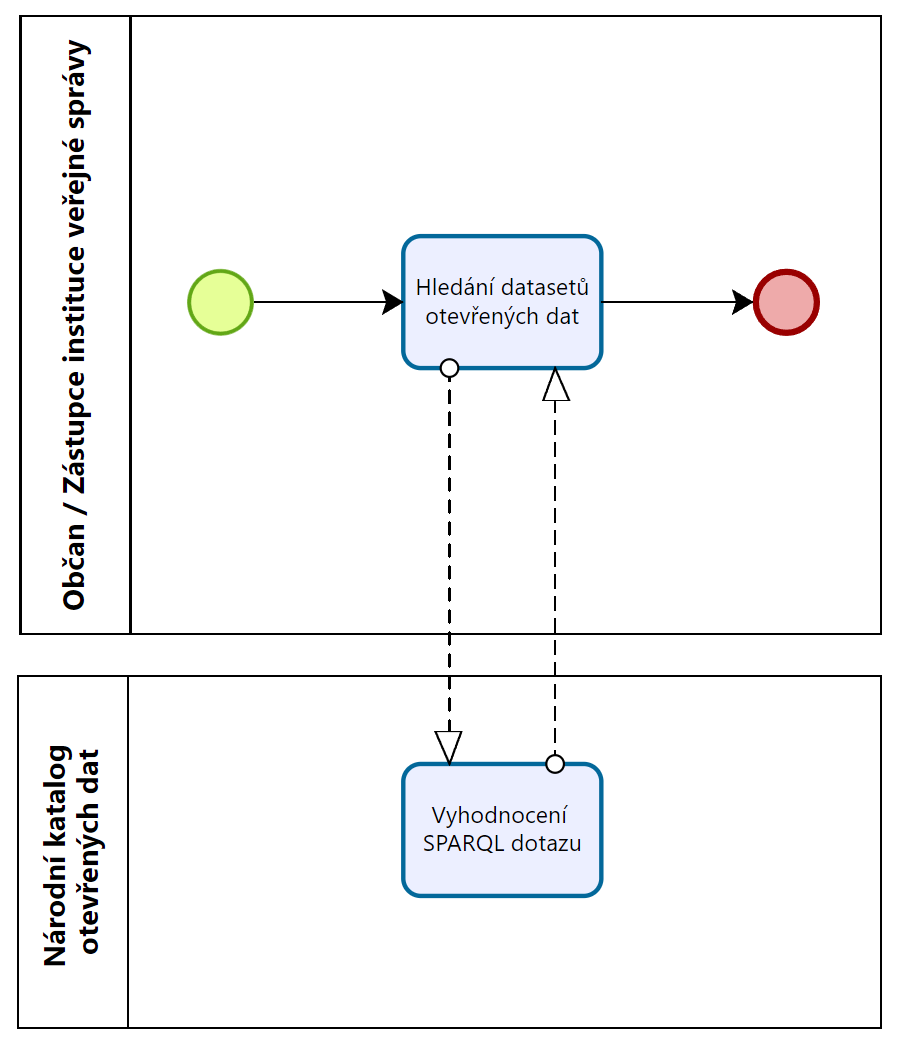
|  |  |
| --- | --- |
| Analytický produkt | Kapitola DNR |
| Specifikace byznys služeb | 6 Biznis architektúra riešenia |
| Aktéři a Případy užití | 9 Prípady použitia (Use case model) |
| Datový model | Příloha „DCAT-AP-SK“ |
| Aplikakační architektura | 10 APLIKAČNÁ ARCHITEKTÚRA |
| Technologická architektura | 11 FYZICKÁ ARCHITEKTÚRA |
| Mapa stránek | 8.3 Návrh mapy stránky (príp. toky používateľov) |
| Prototyp uživatelského rozhraní | 8.3.1 Návrh prototypov |
| Testovací scénáře | Příloha „Plán testů“ |

# Procesy podporované navhrovaným riešením

Národní katalog otevřených dat je centrálním bodem výměny informací o poskytovaných otevřených datech jak mezi veřejnou správou a občany, ale mezi jednotlivými institucemi veřejné správy navzájem.

## Hledání informací o otevřených datech

Občan (nebo zástupce instituce veřejné správy) chce najít informace o otevřených datech na základě informací odpovídajícím položkám ve specifikaci DCAT-AP-SK, např. jméno, popis, klíčová slova nebo datový formát. Použije k tomu aplikaci, která pro vykonání dotazu využije SPARQL endpoint NKOD.



Obrázok 2 Hledání informací o otevřených datech

## Hledání datasetů otevřených dat

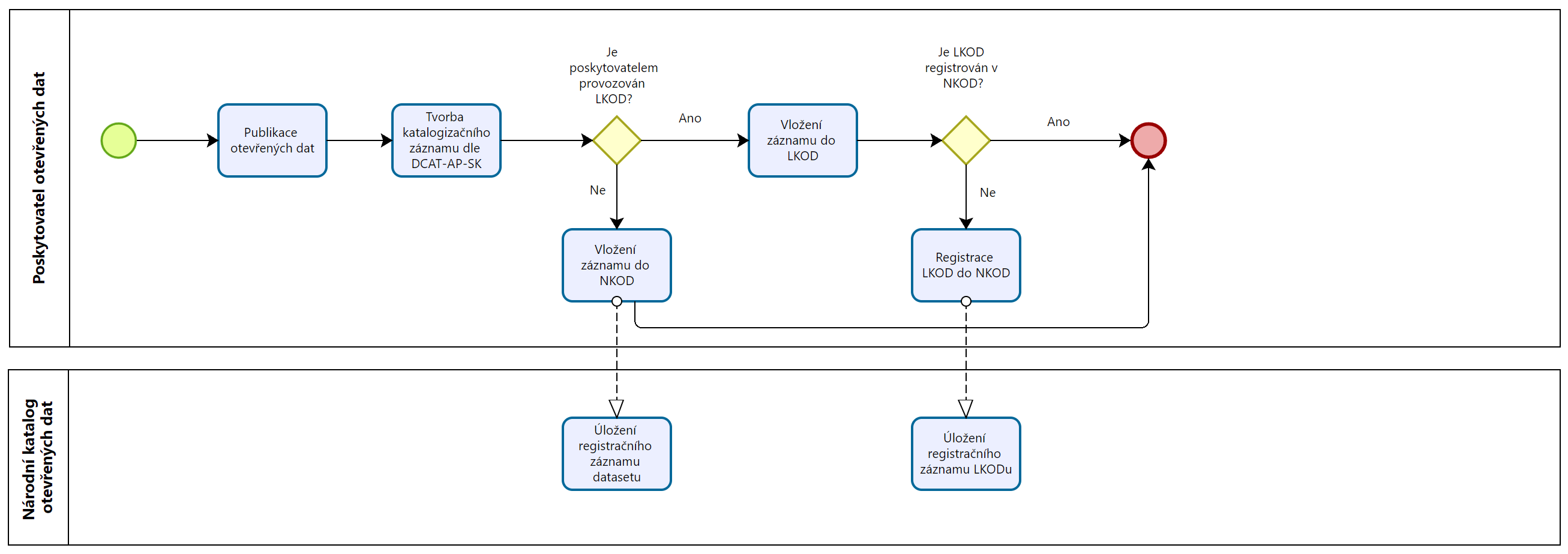
Občan / Zástupce instituce veřejné správy zformuluje (např. pomocí externí aplikace) SPARQL dotaz do NKOD a dostane výsledek.

## Vyhodnocení SPARQL dotazu

RDF databáze NKOD vyhodnotí SPARQL dotaz.

## Poskytování informací o otevřených datech

Poskytovatel dat chce zveřejnit informaci o tom, že data poskytuje, a to tak, aby potenciální uživatelé těchto dat je snadno našli. Data tedy popíše pomocí položek ze specifikace DCAT-AP-SK a zaregistruje buď přímo v NKOD, nebo v LKOD, který je registrován v NKOD.



Obrázok 3 Poskytování informací o otevřených datech

## Publikace otevřených dat

Poskytovatel otevřených dat vystaví soubory otevřených dat či zpřístupní API na svém webovém serveru.

## Tvorba katalogizačního záznamu dle DCAT-AP-SK

Poskytovatel otevřených dat popíše vystavená data pomocí položek dle specifikace DCAT-AP-SK.

## Vložení záznamu do NKOD

Pokud poskytovatel dat neprovozuje svůj LKOD, vloží záznam přímo do NKOD.

## Uložení registračního záznamu datasetu

V budoucnu toto provede poskytovatel pomocí uživatelského rozhraní, přes které bude záznam uložen do úložiště registračních záznamů datasetů. V rámci tohoto projektu NKOD však vloží registrační záznam do úložiště registračních záznamů datasetů administrátor NKOD.

## Vložení záznamu do LKOD

Pokud poskytovatel dat provozuje svůj LKOD s rozhraním dle DCAT-AP-SK, vloží připravený záznam tam.

## Registrace LKOD do NKOD

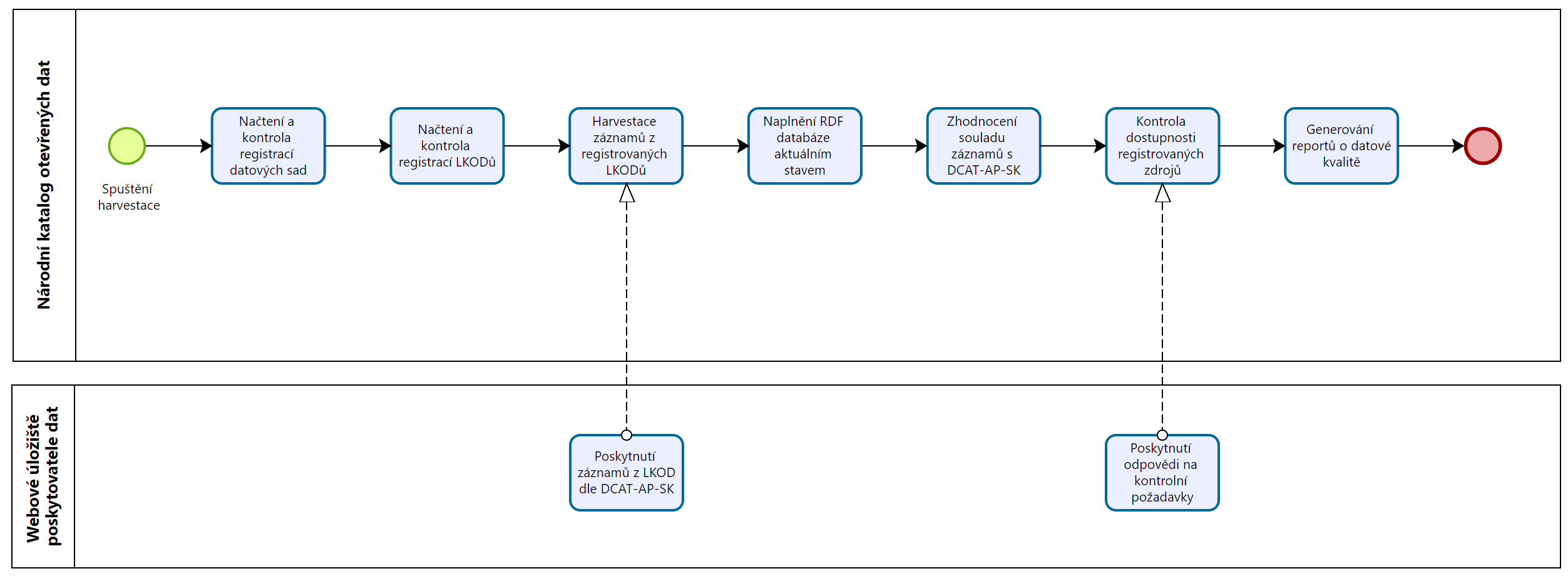
Poskytovatel otevřených dat zaregistruje svůj LKOD s rozhraním dle DCAT-AP-SK do NKOD.

## Uložení registračního záznamu LKODu

V budoucnu toto provede poskytovatel pomocí uživatelského rozhraní, přes které bude záznam uložen do úložiště registračních záznamů LKODů. V rámci tohoto projektu NKOD však vloží registrační záznam do úložiště registračních záznamů LKODů administrátor NKOD.

## Pravidelná aktualizace NKOD a měření kvality

Tento proces běží automaticky v dané periodě, např. denně. Národní katalog otevřených dat načte aktuálně platné registrace datových sad a registrace lokálních katalogů. Přes rozhraní lokálních katalogů načte registrace datových sad z těchto lokálních katalogů s využitím specifikace DCAT-AP-SK. Aktuálně platné registrace pak uloží do RDF databáze, odkud jsou přístupné přes SPARQL endpoint veřejnosti. Následuje zhodnocení souladu registrací se specifikací DCAT-AP-SK a kontrola dostupnosti registrovaných zdrojů. Ze zjištění jsou vygenerovány a publikovány reporty.



Obrázok 4 Pravidelná aktualizace NKOD a měření kvality

## Načtení a kontrola registrací datových sad

NKOD z úložiště registračních záznamů datasetů načte validní záznamy a zařadí je do nově tvořeného obsahu NKOD.

## Načtení a kontrola registrací LKODů

NKOD z úložiště registračních záznamů LKODů načte validní záznamy.

## Harvestace záznamů z registrovaných LKODů

Na základě záznamů z minulého kroku NKOD přistoupí k rozhraní LKOD dle specifikace DCAT-AP-SK a načte registrační záznamy datasetů z LKOD. Zařadí je do nově tvořeného obsahu NKOD.

## Poskytnutí záznamů z LKOD dle DCAT-AP-SK

Poskytovatel otevřených dat provozující LKOD poskytuje registrační záznamy datasetů skrz rozhraní LKOD dle specifikace DCAT-AP-SK.

## Naplnění RDF databáze aktuálním stavem

Registrační záznamy z prvního kroku a registrační záznamy z LKOD vytvoří nový obsah NKOD. Tím je naplněna RDF databáze, jejíž minulý obsah je přepsán aktuálním stavem. Nový obsah NKOD je k dispozici veřejnosti k dotazování.

## Zhodnocení souladu záznamů s DCAT-AP-SK

NKOD projde registrační záznamy datasetů v aktuálním stavu a zhodnotí jejich soulad se specifikací DCAT-AP-SK.

## Kontrola dostupnosti registrovaných zdrojů

NKOD projde registrační záznamy datasetů v aktuálním stavu a zkontroluje dostupnost registrovaných zdrojů, jako jsou URL souborů ke stažení, URL dokumentace, URL schématu apod. a dostupnost techniky CORS na nich.

## Poskytnutí odpovědi na kontrolní požadavky

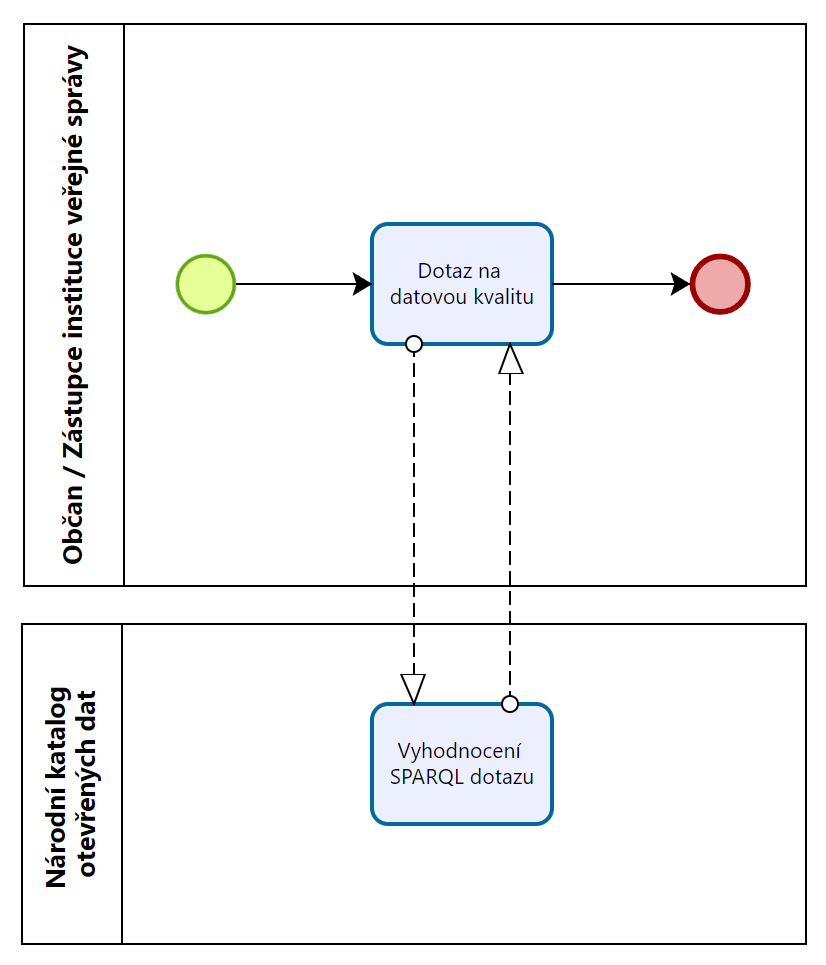
Webový server poskytovatele otevřených dat musí odpovídat na kontrolní požadavky NKOD ve formě požadavků protokolu HTTP HEAD a OPTIONS.

## Generování reportů o datové kvalitě

Ze zjištění kontrol kvality a dostupnosti vygeneruje NKOD report o datové kvalitě.

## Monitorování kvality poskytování informací o otevřených datch

Občan (nebo zástupce instituce veřejné správy), nebo i přímo poskytovatel dat chce zjistit, jak kvalitně vzhledem ke specifikaci DCAT-AP-SK je popsána konkrétní datová sada, jak jsou dostupné v ní registrované zdroje, jak je na tom s kvalitou popisu datových sad a dostupností zdrojů daný poskytovatel, nebo i celý NKOD. Využije k tomu reporty generované v rámci měření datové kvality.



Obrázok 5 Monitorování kvality poskytování informací o otevřených datech

## Dotaz na datovou kvalitu

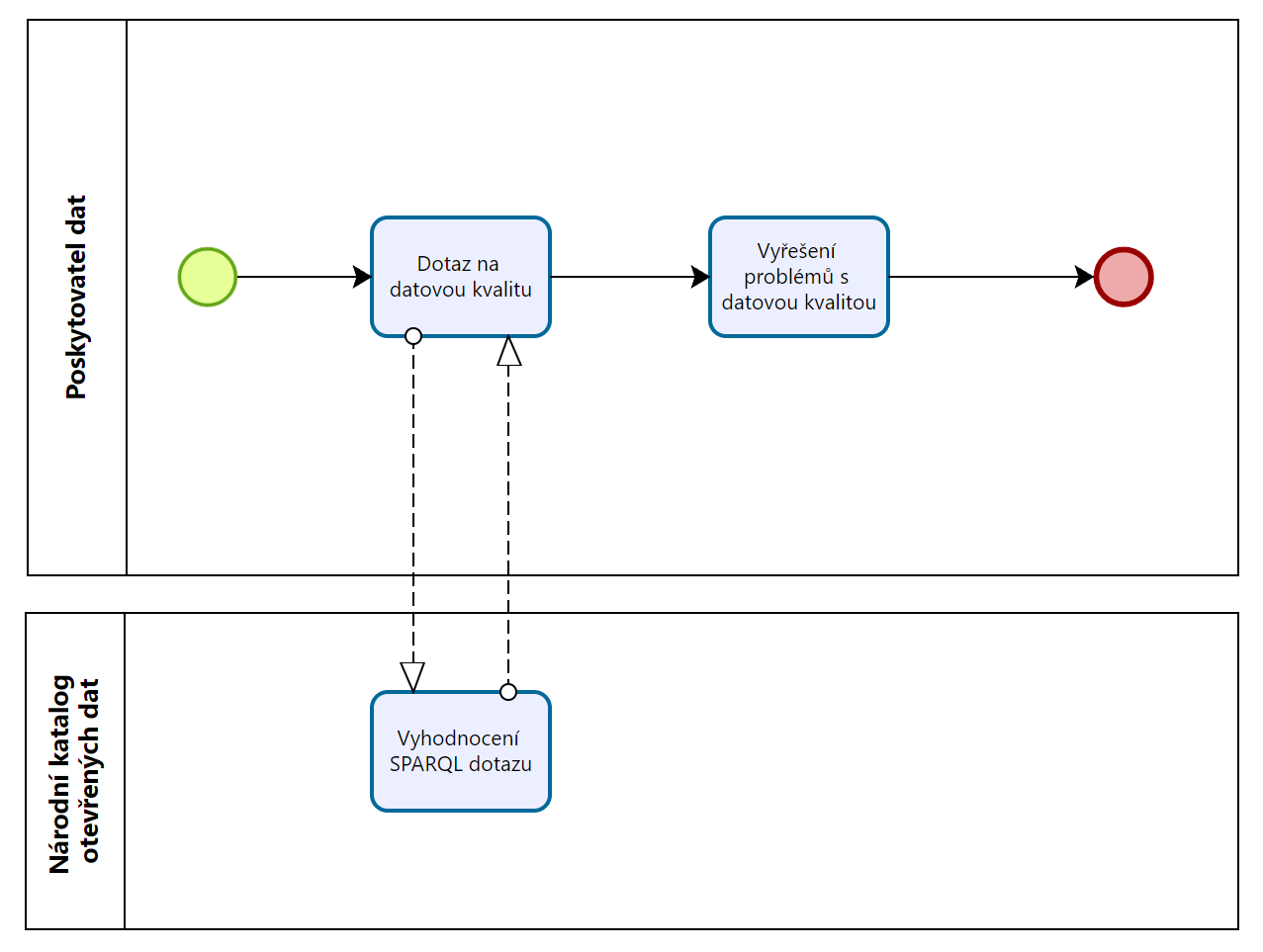
Občan / Zástupce instituce veřejné správy zformuluje (např. pomocí externí aplikace) SPARQL dotaz do NKOD a dostane výsledek.

## Vyhodnocení SPARQL dotazu

RDF databáze NKOD vyhodnotí SPARQL dotaz.

## Zlepšování kvality poskytování otevřených dat

Poskytovatel dat na základě zjištění nedostatků z procesu „Monitorování kvality poskytování informací o otevřených datech“ chce kvalitu poskytování svých datových sad zlepšit. V dokumentaci jednotlivých indikátorů kvality zjistí, co pro zvýšení kvality poskytování dat může udělat.



Obrázok 6 Zlepšování kvality poskytování otevřených dat

## Dotaz na datovou kvalitu

Poskytovatel dat zformuluje (např. pomocí externí aplikace) SPARQL dotaz do NKOD a dostane výsledek.

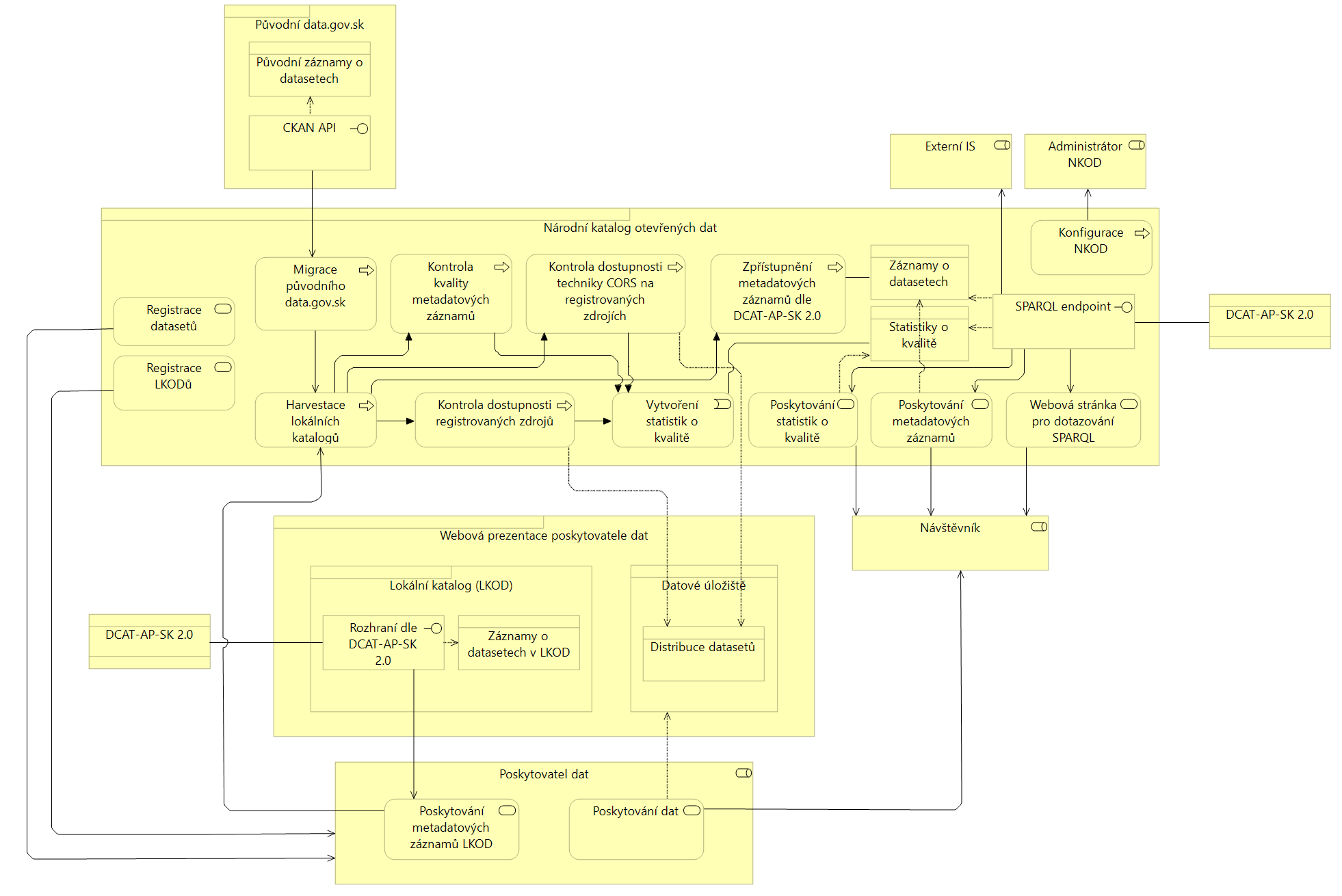
## Vyhodnocení SPARQL dotazu

RDF databáze NKOD vyhodnotí SPARQL dotaz.

## Vyřešení problémů s datovou kvalitou

Na základě problémů přítomných v reportu kvality dat poskytovatel dat provede opatření k jejich nápravě.

# Biznis architektúra riešenia



Obrázok 7 Byznys architektura NKOD

* **DCAT-AP-SK** je standard, který je vytvořen v rámci projektu. Slouží pro definici položek metadatového záznamu datové sady a zároveň definici rozhraní Lokálních katalogů otevřených dat jednotlivých poskytovatelů otevřených dat. To používá NKOD pro harvestaci metadatových záznamů z LKOD.
* **Původní data.gov.sk** bude dočasně harvestován skrz jeho CKAN API jako další LKOD s tím rozdílem, že jeho rozhraní neodpovídá specifikaci a je tedy třeba ho řešit samostatně.
* **Webová prezentace poskytovatele dat** zahrnuje datové úložiště, skrz které jsou poskytovány distribuce datasetů v podobě souborů ke stažení a datových služeb. Dále může obsahovat Lokální katalog (LKOD), který bude harvestován do NKOD přes rozhraní kompatibilní s DCAT-AP-SK.
* **Národní katalog otevřených dat** je hlavním produktem tohoto projektu. Zahrnuje proces harvestace metadat z jednotlivých LKODů a z původního data.gov.sk, čímž vznikají záznamy o datasetech v NKOD. Následně NKOD provádí kontrolu kvality metadatových záznamů, kontrolu dostupnosti registrovaných zdrojů a kontrolu dostupnosti techniky CORS na registrovaných zdrojích. Z těchto procesů vzniknou statistiky o kvalitě, které jsou, spolu s metadatovými záznamy, poskytovány uživatelům NKOD. Samotná data jednotlivých poskytovatelů dat zůstávají v jejich datových úložištích. NKOD obsahuje pouze odkazy na tato data, nikoliv data samotná.

## Aktéři

### Poskytovatel dat

Poskytovatel otevřených dat, který vystavuje data na své webové prezentaci a registruje metadata buď přímo v NKOD, nebo ve svém LKOD, který je registrován v NKOD.

### Návštevník

Libovolná fyzická osoba v roli návštevníka webové stránky. Může vybrat SPARQL dotaz nebo napsat vlastní, a nechat si ho vyhodnotit.

### Externí IS

Externí informační systém, který se dotazuje Národního katalogu otevřených dat přímo přes SPARQL Endpoint, bez webové stránky.

### Administrátor NKOD (Správca NKOD)

Administrátor Národního katalogu otevřených dat. Konfiguruje předpřipravené SPARQL dotazy na webové stránce, a dokud nebude uživatelské rozhraní, registruje datasety a LKODy k harvestaci.

# MapovAnIE Funkčných požIadaviek

V této tabulce jsou funkční požadavky na řešení ze zadání mapovány na byznys procesy, kterých se týkají.

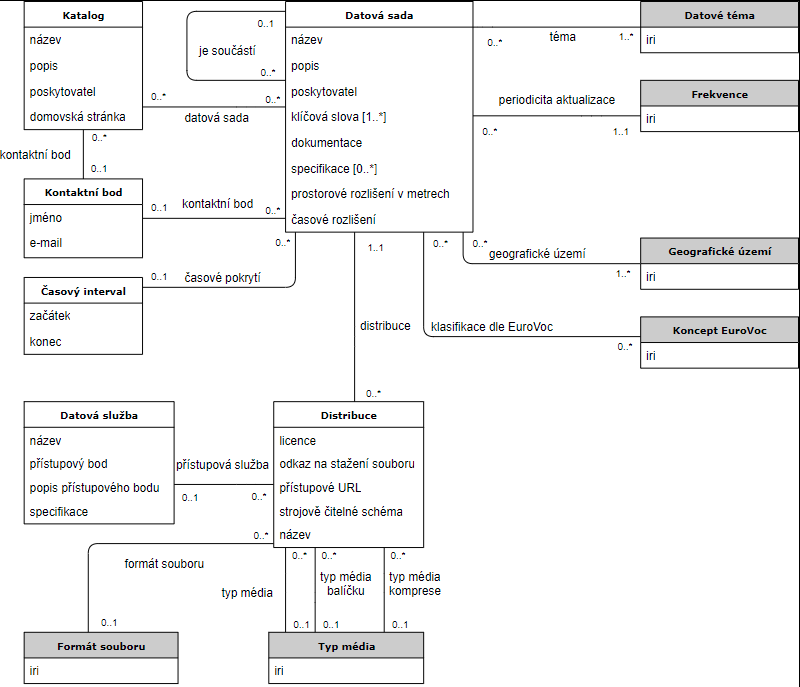
Tabuľka 2 Mapování funkčních požadavků na byznys procesy

|  |  |  |
| --- | --- | --- |
| ID | Požadavek | Byznys artefakt |
| 2 | Vykonanie dotazu na metadáta otvorených údajov používateľským rozhraním | Poskytování metadatových záznamů |
| 3 | Spôsob dotazovania prostredníctvom SPARQL Endpointu | Webová stránka pro dotazování SPARQL |
| 4 | Vykonanie preddefinovaných dotazov | Webová stránka pro dotazování SPARQL |
| 5 | Konfigurácia preddefinovaných dotazov | Webová stránka pro dotazování SPARQL |
| 6 | Konfigurácia externého dátového portálu | Konfigurace NKOD |
| 7 | Automatické harvestovanie externých portálov | Harvestace lokálních katalogů, Zpřístupnění metadatových záznamů dle DCAT-AP-SK |
| 9 | Spôsob harvestovania externých portálov | Harvestace lokálních katalogů |
| 10 | Vytvorenie základných štatistík metadát | Vytvoření statistik o kvalitě |
| 11 | Vytvorenie štatistík kvality metadát | Kontrola kvality metadatových záznamů, Kontrola dostupnosti registrovaných zdrojů, Kontrola dostupnosti techniky CORS na registrovaných zdrojích |
| 12 | Migrácia metadát otvorených údajov portálu OD1.0 | Migrace původního data.gov.sk |
| 14 | URL portálu | Webová stránka pro dotazování SPARQL |
| 18 | Harvestovanie súčasného portálu data.gov.sk | Migrace původního data.gov.sk |
| 21 | Viacjazyčnosť používateľského rozhrania | Webová stránka pro dotazování SPARQL |

# Popis dátových entít

## Diagram entít – Záznam dÁtové sady

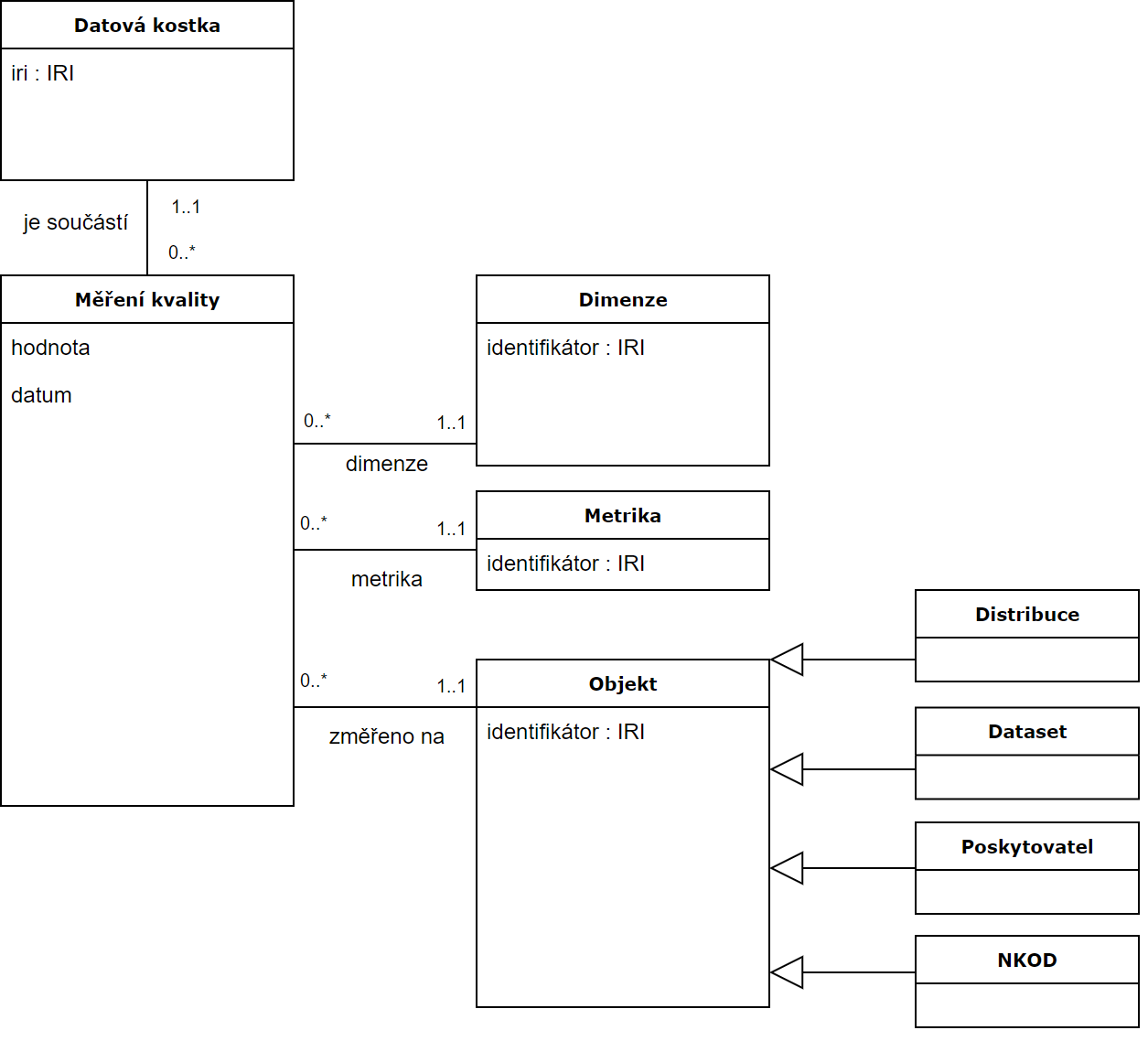
Datová struktura záznamu o datasetu se bude řídit standardem DCAT-AP-SK (v příloze).



Obrázok 8 Datový model pro záznam datasetu

## Diagram entít – MERANIA KVALITY

Záznam o měření kvality vychází se standardu Data Quality Vocabulary (<https://www.w3.org/TR/vocab-dqv/>).



Obrázok 9 Datový model pro měření datové kvality

### Popis datových entit

* **Měření kvality:** reprezentuje jednu naměřenou hodnotu v dané datum na daném objektu, v konkrétní dimenzi kvality a metrice, patřící do příslučné datové kostky.
* **Datová kostka:** obsahuje sadu měření na daném typu objektu.
* **Dimenze:** Určuje dimenzi kvality dle <https://www.w3.org/TR/vocab-dqv/#DimensionsofZaveri>
* **Metrika:** Určuje, která konkrétní metrika kvality je měřena.
* **Objekt:** Objekt, na kterém je měření provedeno. Může to být NKOD jako takový, Poskytovatel, Dastaset či Distribuce.

### Měřené metriky kvality

### Dostupnost a dostupnost CORS

V této sekci je přehled měřených metrik dostupnosti registrovaných zdrojů. Počítány jsou pak i jejich agregace na úroveň datové sady, poskytovatele a NKOD, například počet datových sad v NKOD s nedostupným souborem ke stažení.

1. Dostupnost dokumentace datové sady
2. Dostupnost specifikace datové sady
3. Dostupnost souboru ke stažení
4. Dostupnost CORS na souboru ke stažení
5. Dostupnost schématu souboru ke stažení
6. Dostupnost CORS na schématu souboru ke stažení
7. Dostupnost podmínek užití distribuce datové sady
8. Dostupnost specifikace datové služby
9. Dostupnost přístupového bodu datové služby
10. Dostupnost CORS na přístupovém bodu datové služby
11. Dostupnost popisu přístupového bodu datové služby
12. Dostupnost CORS na popisu přístupového bodu datové služby

### Kvalita metadat

V této sekci je přehled měřených metrik kvality metadatových záznamů.

Tabuľka 3 Přehled metrik kvality metadat

|  |  |
| --- | --- |
| ID Metriky | Definice |
| PočetDatovýchSad | Indikuje, kolik datových sad (dcat:Dataset) daný poskytovatel má. |
| PočetDatovýchSadDatováSlužba | Indikuje, kolik datových sad (dcat:Dataset) mající distribuci reprezentujících datovou službu daný poskytovatel má. |
| PočetDatovýchSadDatováSlužbaSPopisem | Indikuje, kolik datových sad (dcat:Dataset) mající distribuci reprezentujících datovou službu s popisem přístupového bodu daný poskytovatel má. |
| PočetDatovýchSadDatováSlužbaSPopisemVNKOD | Indikuje, kolik datových sad v NKOD obsahuje distribuci ve formě datové služby s popisem přístupového bodu. |
| PočetDatovýchSadDatováSlužbaSeSpecifikací | Indikuje, kolik datových sad (dcat:Dataset) mající distribuci reprezentujících datovou službu se specifikací daný poskytovatel má. |
| PočetDatovýchSadDatováSlužbaSeSpecifikacíVNKOD | Indikuje, kolik datových sad v NKOD obsahuje distribuci ve formě datové služby se specifikací. |
| PočetDatovýchSadDatováSlužbaVNKOD | Indikuje, kolik datových sad v NKOD obsahuje distribuci ve formě datové služby. |
| PočetDatovýchSadSDistribucíBezPodmínekUžití | Indikuje, kolik datových sad (dcat:Dataset) má distribuce (dcat:Distribution), jejichž podmínky užití nejsou specifikovány. |
| PočetDatovýchSadSDistribucíBezPodmínekUžitíVNKOD | Indikuje, kolik datových sad (dcat:Dataset) v NKOD má distribuce (dcat:Distribution), jejichž podmínky užití nejsou specifikovány. |
| PočetDatovýchSadSDokumentací | Indikuje, kolik datových sad (dcat:Dataset) s dokumentací (foaf:page) daný poskytovatel má. |
| PočetDatovýchSadSDokumentacíVNKOD | Indikuje, kolik datových sad v NKOD má dokumentaci. |
| PočetDatovýchSadSeSpecifikací | Indikuje, kolik datových sad (dcat:Dataset) se specifikací (dcterms:conformsTo) daný poskytovatel má. |
| PočetDatovýchSadSeSpecifikacíVNKOD | Indikuje, kolik datových sad v NKOD má specifikaci. |
| PočetDatovýchSadSouborKeStažení | Indikuje, kolik datových sad (dcat:Dataset) mající distribuci reprezentujících soubor ke stažení daný poskytovatel má. |
| PočetDatovýchSadSouborKeStaženíSeSchématem | Indikuje, kolik datových sad (dcat:Dataset) mající distribuci reprezentujících soubor ke stažení se schématem daný poskytovatel má. |
| PočetDatovýchSadSouborKeStaženíSeSchématemVNKOD | Indikuje, kolik datových sad v NKOD obsahuje distribuci ve formě souboru ke stažení se schématem. |
| PočetDatovýchSadSouborKeStaženíVNKOD | Indikuje, kolik datových sad v NKOD obsahuje distribuci ve formě souboru ke stažení. |
| PočetDatovýchSadVNKOD | Indikuje, kolik datových sad (dcat:Dataset) je v NKOD. |
| PočetDistribucíDatovýchSad | Indikuje, kolik distribucí datových sad (dcat:Distribution) daný poskytovatel má. |
| PočetDistribucíDatovýchSadBezPodmínekUžití | Indikuje, kolik distribucí datových sad (dcat:Distribution) nemá specifikovány podmínky užití. |
| PočetDistribucíDatovýchSadDatováSlužba | Indikuje, kolik distribucí datových sad (dcat:Distribution) reprezentujících datovou službu daný poskytovatel má. |
| PočetDistribucíDatovýchSadDatováSlužbaVNKOD | Indikuje počet distribucí datových sad (dcat:Distribution) reprezentujících datovou službu v NKOD. |
| PočetDistribucíDatovýchSadSouborKeStažení | Indikuje, kolik distribucí datových sad (dcat:Distribution) reprezentujících soubor ke stažení daný poskytovatel má. |
| PočetDistribucíDatovýchSadSouborKeStaženíVNKOD | Indikuje počet distribucí datových sad (dcat:Distribution) reprezentujících soubor ke stažení v NKOD. |
| PočetDistribucíDatovýchSadVNKOD | Indikuje, kolik distribucí datových sad (dcat:Distribution) je v NKOD. |
| PočetKlíčovýchSlovVNKOD | Indikuje, kolik klíčových slov je použito v NKOD. |
| PočetPoskytovatelů | Indikuje, kolik je v NKOD posktovatelů. |
| PočetTématVNKOD | Indikuje, kolik témat je použito v NKOD. |
| PočetZáznamůDatovýchSadNesplňujícíchPovinnéAtributy | Indikuje, kolik záznamů datových sad neobsahuje povinné atributy. |
| PočetZáznamůDatovýchSadNesplňujícíchPovinnéAtributyVNKOD | Indikuje, kolik záznamů datových sad v NKOD neobsahuje povinné atributy. |
| ProcentoDatovýchSadSDistribucemiSPodmínkamiUžití | Indikuje, kolik procent datových sad (dcat:Dataset) poskytovatele má distribuce (dcat:Distribution) popsané podmínkami užití. |
| ProcentoDatovýchSadSDistribucíBezPodmínekUžitíVNKOD | Indikuje, kolik procent datových sad (dcat:Dataset) v NKOD má distribuce (dcat:Distribution) se specifikovanými podmínkami užití. |
| ProcentoDistribucíDatovýchSadSPodmínkamiUžití | Indikuje, kolik procent distribucí datových sad poskytovatele (dcat:Distribution) je popsaných podmínkami užití. |
| ProcentoZáznamůDatovýchSadSplňujícíchPovinnéAtributy | Indikuje, kolik procent záznamů datových sad poskytovatele obsahuje povinné atributy. |
| ProcentoZáznamůDatovýchSadSplňujícíchPovinnéAtributyVNKOD | Indikuje, kolik procent záznamů datových sad v NKOD obsahuje povinné atributy. |
| PrůměrProcentDatovýchSadSDistribucemiSPodmínkamiUžitíPřesPoskytovatele | Indikuje průměr procent datových sad (dcat:Dataset) má distribuce (dcat:Distribution) popsané podmínkami užití přes poskytovatele. |
| PrůměrProcentZáznamůDatovýchSadSplňujícíchPovinnéAtributyPřesPoskytovatele | Indikuje průměr procent záznamů datových sad v NKOD obsahujících povinné atributy přes poskytovatele. |
| PřítomnostPovinnýchAtributů | Indikuje, zda záznam datové sady obsahuje povinné atributy. |

# Návrh riešenia dizajnu koncových služieb

Pro Návštěvníka (občana či zástupce instituce veřejné správy) poskytuje NKOD jedinou webovou stránku umožňující dotazovat se na SPARQL. Jedná se o formulář umožňující zaslání vlastního či předpřipraveného SPARQL dotazu a zobrazení výsledku dotazu.

Uživatelské rozhraní bude vybudováno v jiných částech projektu OD2.0.

## Identifikácia vlastníka dizajnu koncových služieb (UX)

### Pomenovanie vlastníka

Ing. Miroslav Líška, PhD., Dátová kancelária, MIRRI

### Kontakt na vlastníka

e-mail: miroslav.liska@mirri.gov.sk

## Špecifikácia potrieb koncového používateľa

Potřeby koncových uživatelů týkající se přímo části NKOD:

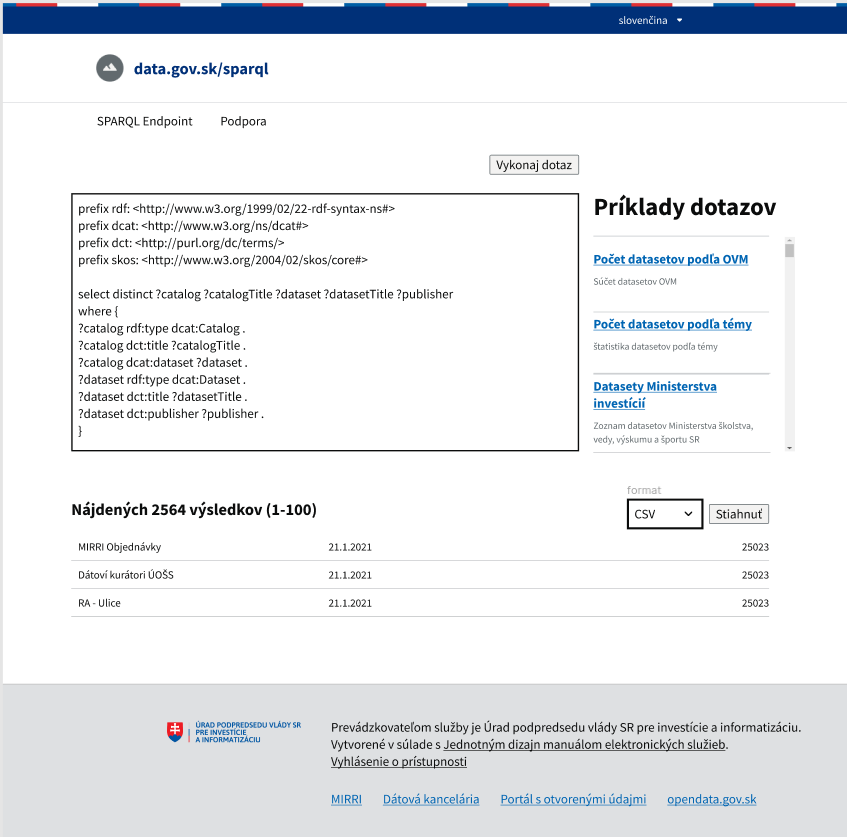
1. Jako občan (veřejnost) se potřebuji dotázat do NKOD a dostat výsledek dotazu.
   * **Návrh řešení:** Implementace SPARQL endpointu a specifikace dat v něm v podobě DCAT-AP-SK
2. Jako poskytovatel dat potřebuji, aby mnou zarigstrovaný lokální katalog otevřených dat byl harvestován do NKOD.
   * **Návrh řešení:** Specifikace rozhraní LKOD v podobě DCAT-AP-SK a umožnění administrátorovi NKOD toto rozhraní LKOD zaregistrovat k harvestaci.
3. Jako poskytovatel dat potřebuji vidět report hodnocení metadatové kvality a dostupnosti registrovaných zdrojů, abych mohl zlepšit způsob poskytování otevřených dat
   * **Návrh řešení:** Vytvoření reportů dle webového standardu Data Catalog Vocabulary a v datovém formátu CSV.
4. Jako občan (veřejnost) chci zjistit, jak je na tom konkrétná poskytovatel otevřených dat s kvalitou metadat a dostupností registrovaných zdrojů.
   * **Návrh řešení:** Vytvoření reportů dle webového standardu Data Catalog Vocabulary a v datovém formátu CSV.

## Návrh mapy stránky (príp. toky používateľov)

Výstupem projektu je jediná obrazovka – zadání SPARQL dotazu k vyhodnocení. V rámci této stránky se zobrazují i výsledky vyhodnocení dotazu.

### Návrh prototypov

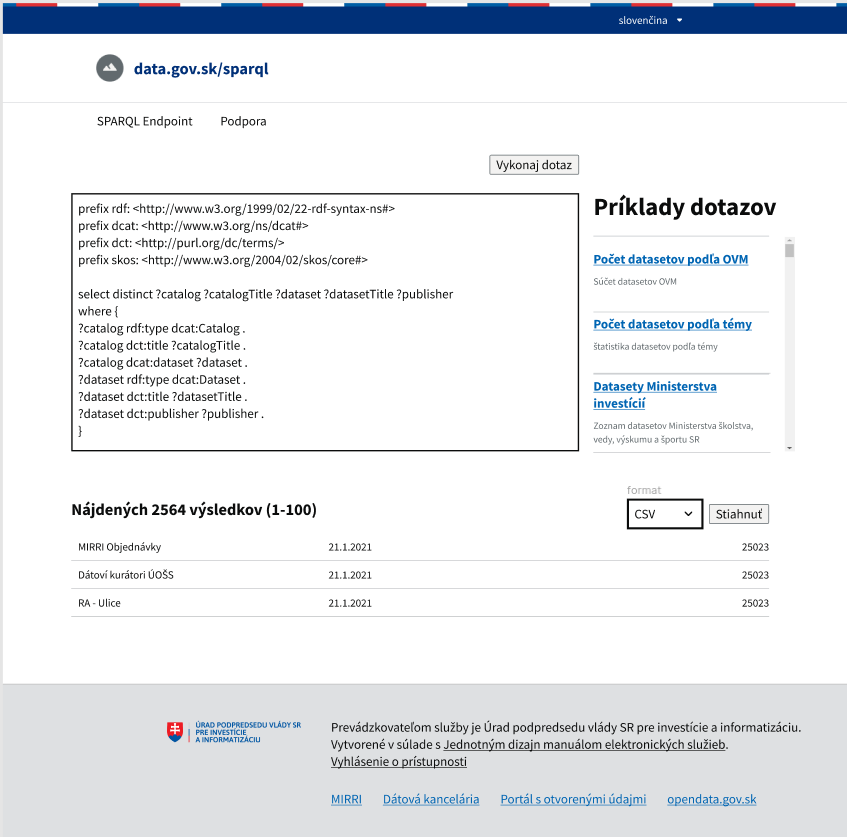
Na obrazovce je místo pro zadání vlastního SPARQL dotazu a také seznam předpřipravených SPARQL dotazů na výběr. Po spuštění dotazu jsou vidět výsledky dotazu.



Obrázok 10 Obrazovka pro zadání SPARQL dotazu

# Prípady použitia (Use case model)

## Návrh vizuálov obrazoviek, podporených E2E procesom



Obrázok 11 Obrazovka pro zadání SPARQL dotazu

|  |  |  |
| --- | --- | --- |
| **ID obrazovky** | - | 1 |
| **Názov** | - | Zadání a vyhodnocení SPARQL dotazu |
| **Význam** | - | Slouží k zadání SPARQL dotazu či výběru předpřipraveného dotazu a jeho vykonání |

## Používateľské roly – aktéri

### Poskytovatel dat

Poskytovatelé dat aktuálně registrovaní na data.gov.sk.

### Návštevník

Libovolná fyzická osoba v roli návštevníka webové stránky.

### Externí IS

Externí informační systém, který se dotazuje Národního katalogu otevřených dat přes SPARQL Endpoint.

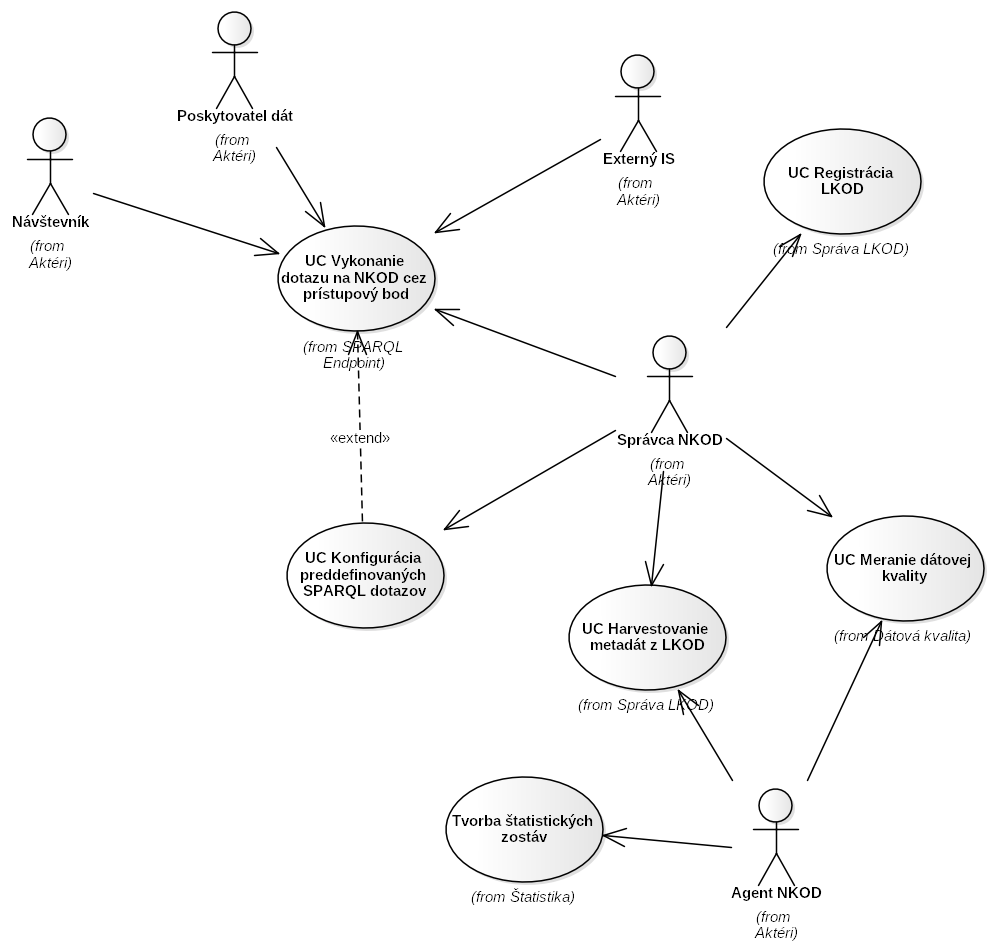
### Administrátor NKOD (Správca NKOD)

Administrátor Národního katalogu otevřených dat.

### Agent NKOD

Komponent systému NKOD zodpovedný za automatizované harvestovanie metadát lokálnych katalógov (LKOD) spolu s automatizovanou tvorbou štatistík systému.

## Všebecné prípady použitia



### UC Vykonanie dotazu cez prístupový bod

**Vykonanie vlastného dotazu**

1. Používateľ si zobrazí používateľské rozhranie SPARQL Endpointu a napíše vlastný dotaz a zvolí možnosť vykonať
2. Systém vykoná daný dotaz a zobrazí zoznam výsledkov
3. Používateľ má možnosť exportovať zoznam výsledkov do zoznamu poskytovaných formátov

**A1 - Vykonanie preddefinovaného dotazu {Alternatívny tok}**

1. Používateľ má možnosť vybrať pripravený dotaz zo zoznamu, pričom ho môže editovať a zvolí možnosť vykonať
2. Následne pokračuje bod 2. v základnom toku

**A1.1 – Vykonanie dotazov pre registrované LKOD {Alternatívny tok}**

**A1.2 – Vykonanie dotazov harverstovania LKOD {Alternatívny tok}**

**A1.3 – Vykonanie dotazov merania kvality metadát {Alternatívny tok}**

**A1.4 – Vykonanie predefinovaných štatistických dotazov {Alternatívny tok}**

### UC Konfigurácia preddefinovaných dotazov

**Konfigurácia preddefinovaných dotazov**

1. Správca NKOD má možnosť pridať nový preddefinovaný dotaz
2. Systém následne zobrazuje daný dotaz v zozname preddefinovaných dotazov

### UC Registrácia LKOD

**Registrácia LKOD**

1. Správca NKOD má možnosť registrovať lokálny katalóg a nastaviť spôsob jeho harvestovania.

### UC Harvestovanie metadát z lokálnych katalógov

**Harvestovanie lokálneho katalógu**

1. Systém automatizovane harvestuje katalógy zo zoznamu registrovaných LKOD a vytvára štatistické dáta harvestovania lokálnych katalógov

### UC Meranie dátovej kvality

**Meranie kvality metaúdajov**

1. Systém automatizovane vykonáva meranie dátovej kvality metaúdajov a vytvára štatistické dáta merania dátovej kvality

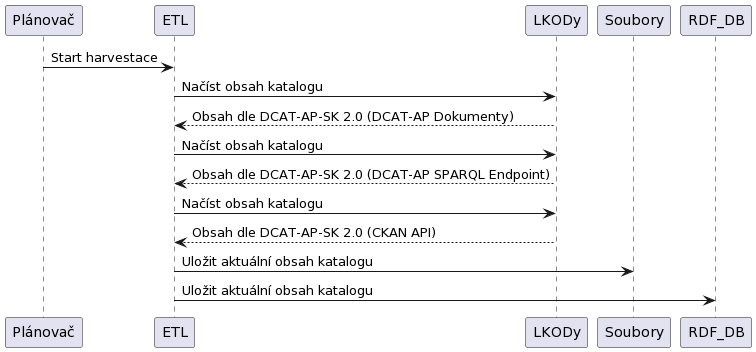
### UC Tvorba štatistických zostáv

**Tvorba štatistických zostáv**

1. Systém automatizovane vytvára štatistické zostavy systému

## Sekvenčné diagramy (model správania systému)

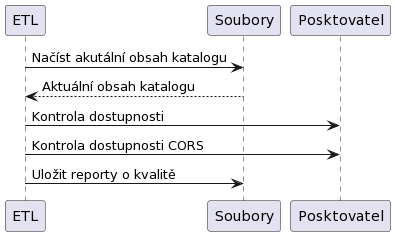
### Harvestace lokálních katalogů

Tento sekvenční diagram popisuje princip fungování automatické harvestace a měření dostupnosti registrovaných zdrojů.

Proces harvestace je spuštěn z plánovače, např. denně

1. ETL – nástroj implementující datové procesy NKOD – načte obsahy registrovaných katalogů dle specifikace DCAT-AP-SK – v diagramu jsou znázorněny 3 typy rozhraní LKOD
2. ETL zpracuje záznamy a uloží je do souborového systému. Ten může být např. zveřejněn přes webový server
3. ETL záznamy uloží i do grafové databáze

### Měření dostupnosti registrovaných zdrojů



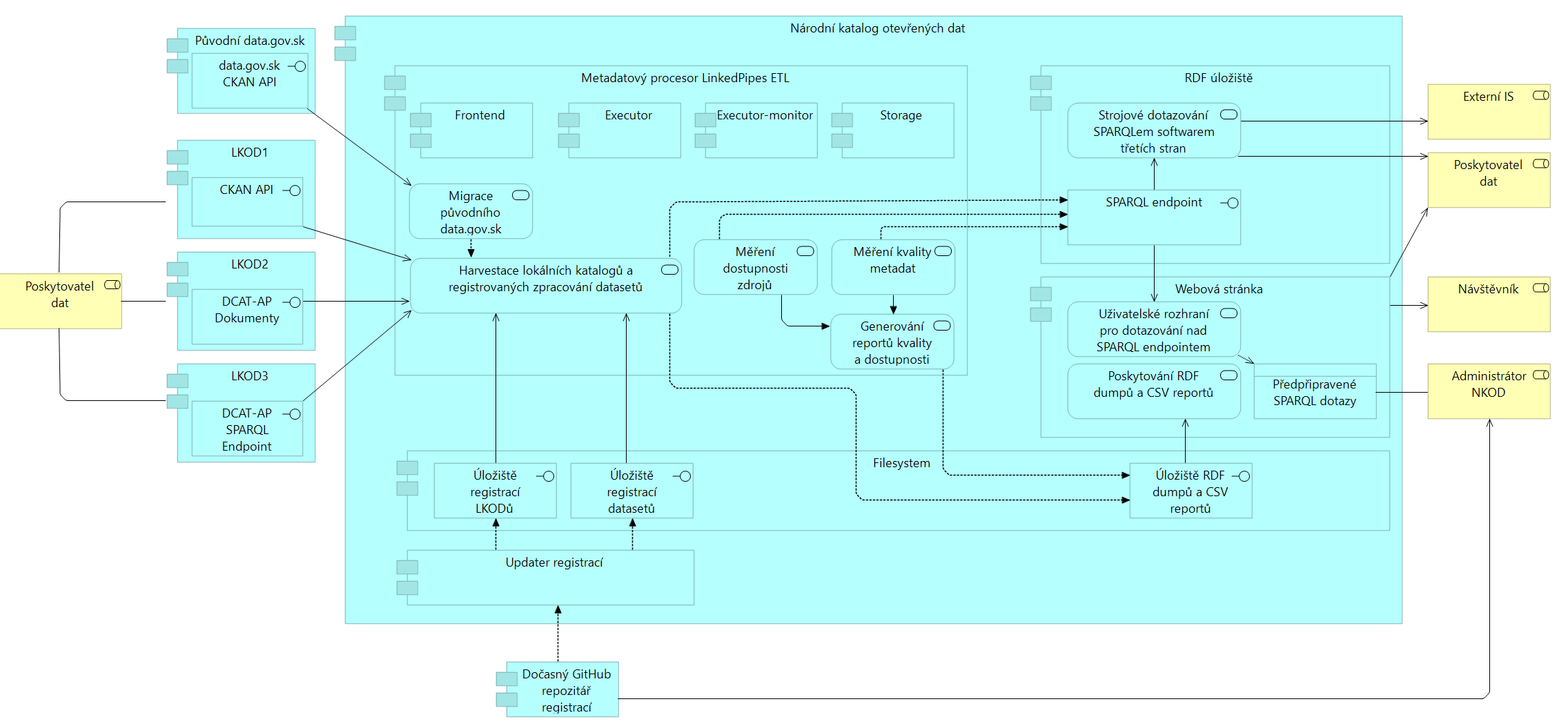
Tento sekvenční diagram zobrazuje proces měření dostupnosti a dostupnosti techniky CORS.

1. ETL načte aktuální obsah katalogu ze souborového systému
2. ETL kontaktuje webové servery poskytovatelů hostující registrované zdroje a zjišťuje jejich dostupnost
3. ETL kontaktuje webové servery poskytovatelů hostující registrované zdroje a zjišťuje dostupnost techniky CORS
4. ETL ukládá výsledky měření do souborového systému

Časť 2: Technická špecifikácia a detailný návrh riešenia

# APLIKAČNÁ ARCHITEKTÚRA

## prehľad



Obrázok 12 Aplikační architektura

### SPARQL Endpoint

SPARQL endpoint je webová služba pro dotazování v RDF úložišti. Lze k němu přistupovat strojově, což využijí Externí IS – aplikace třetích stran. Pro uživatele je dotazování zprostředkováno přes Webovou stránku.

### DCAT-AP-SK

Specifikace metadat popisujících datasety otevřených dat a rozhraní lokálních katalogů (v příloze).

### RDF úložiště

RDF úložiště obsahuje aktuální obsah NKOD a měření kvality metadat a dostupnosti zdrojů.

### Národní katalog otevřených dat (NKOD)

Národní katalog otevřených dat. Z datového pohledu je tvořen aktuálním stavem registrovaných datových sad a měření kvality, které jsou uloženy jednak ve Filesystemu, a také v RDF úložišti. Z pohledu aplikačního se jedná o centrální produkt tohoto projektu, který se skládá z Metadatového procesoru (LinkedPipes ETL), RDF úložiště a jeho SPARQL endpointu, webové stránky a dočasného updateru registrací datasetů a LKODů.

### LKOD (1,2,3)

Lokální katalog otevřených dat daného poskytovatele registrovaný v NKOD pro pravidelné harvestování metadat.

### Metadatový procesor (LinkedPipes ETL)

Sada datových procesů implementovaných v nástroji LinkedPipes ETL, zajišťující pravidelné harvestování metadat z LKODů a měření kvality metadat a dostupnosti registrovaných zdrojů.

### Původní data.gov.sk

Původní portál otevřených dat. Poskytuje metadata upraveným rozhraním CKAN API. Při jeho migraci se v metadatovém procesoru provádí věcné úpravy metadat – mapování číselníků, sdružování do datových sérií apod.

### Filesystem

Souborový systém obsahující data NKOD, která jsou také nahrána v RDF úložišti, v podobě souborů. Slouží také pro úložiště registrací datasetů a LKODů.

### Webová stránka

Webová stránka zpřístupňuje SPARQL dotazování lidským uživatelům. Obsahuje možnost v konfiguraci přidat předpřipravené dotazy a umožňuje vícejazyčné zobrazení rozhraní.

### Updater registrací

Jedná se o dočasnou komponentu, která reaguje na změnu GitHub repozitáře s registracemi datasetů a LKODů a zrcadlí jeho obsah do Filesystému. Po dokončení frontendu zajišťujícího tvorbu a ukládání registračních záznamů (jiná část OD2.0) může být nahrazena.

## KOMUNIKÁCIA MEDZI KOMPONENTAMI

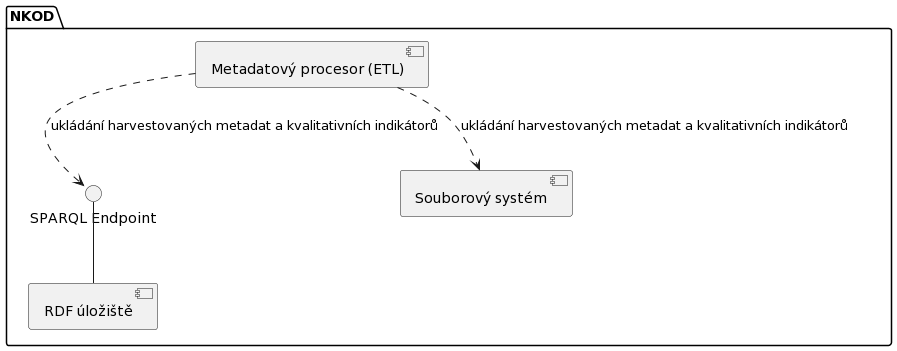
### Komunikácia medzi UI a servisnou vrstvou

RDF úložiště NKOD poskytuje SPARQL endpoint, který využívá webová stránka tvořící UI. SPARQL endpoint je přímo datová vrstva, servisní vrstva není potřeba.

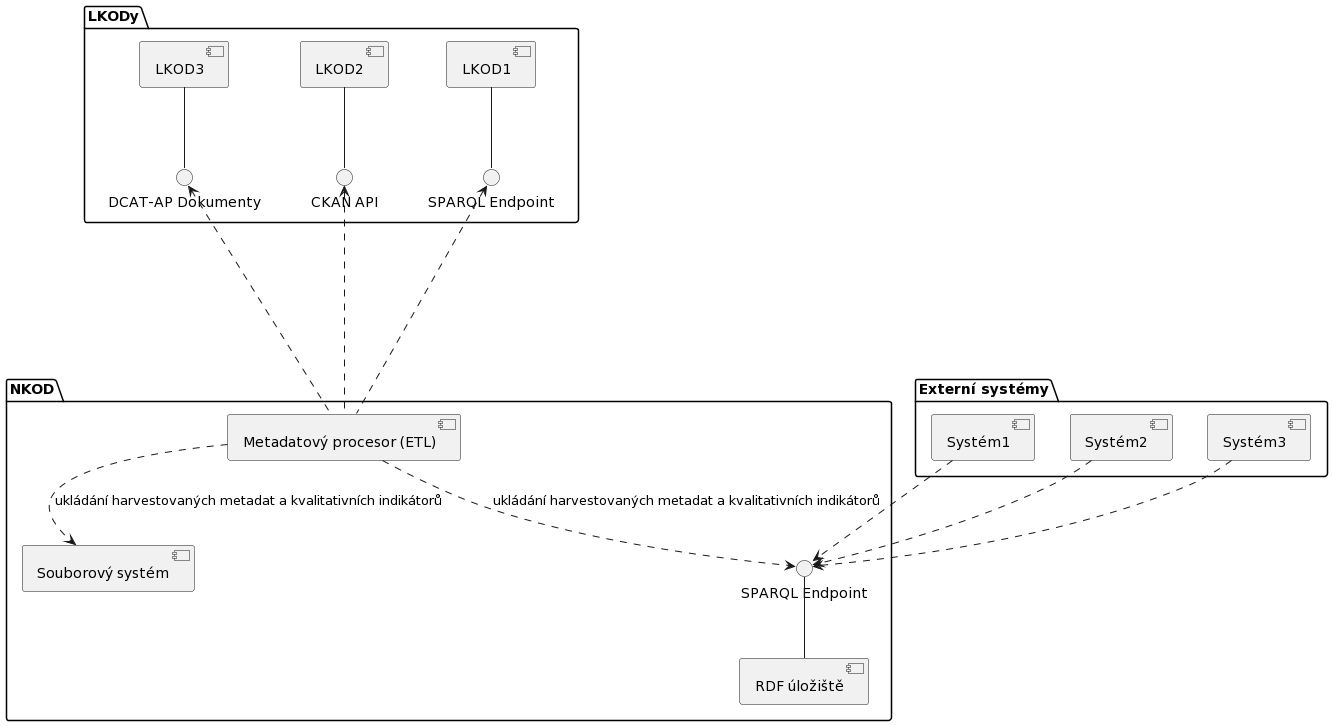
### Komunikácia medzi servisnou vrstvou a dátovou vrstvou

RDF úložiště NKOD poskytuje SPARQL endpoint, který využívá webová stránka tvořící UI. SPARQL endpoint je přímo datová vrstva, servisní vrstva není potřeba.

### Iné formu komunikácie v ráci riešenia



## KOMUNIKÁCIA MEDZI INTERNÝMI A EXTERNÝMI SYSTÉMAMI



### Zoznam a popis existujúcich integrácií

Nejsou žádné stávající integrace.

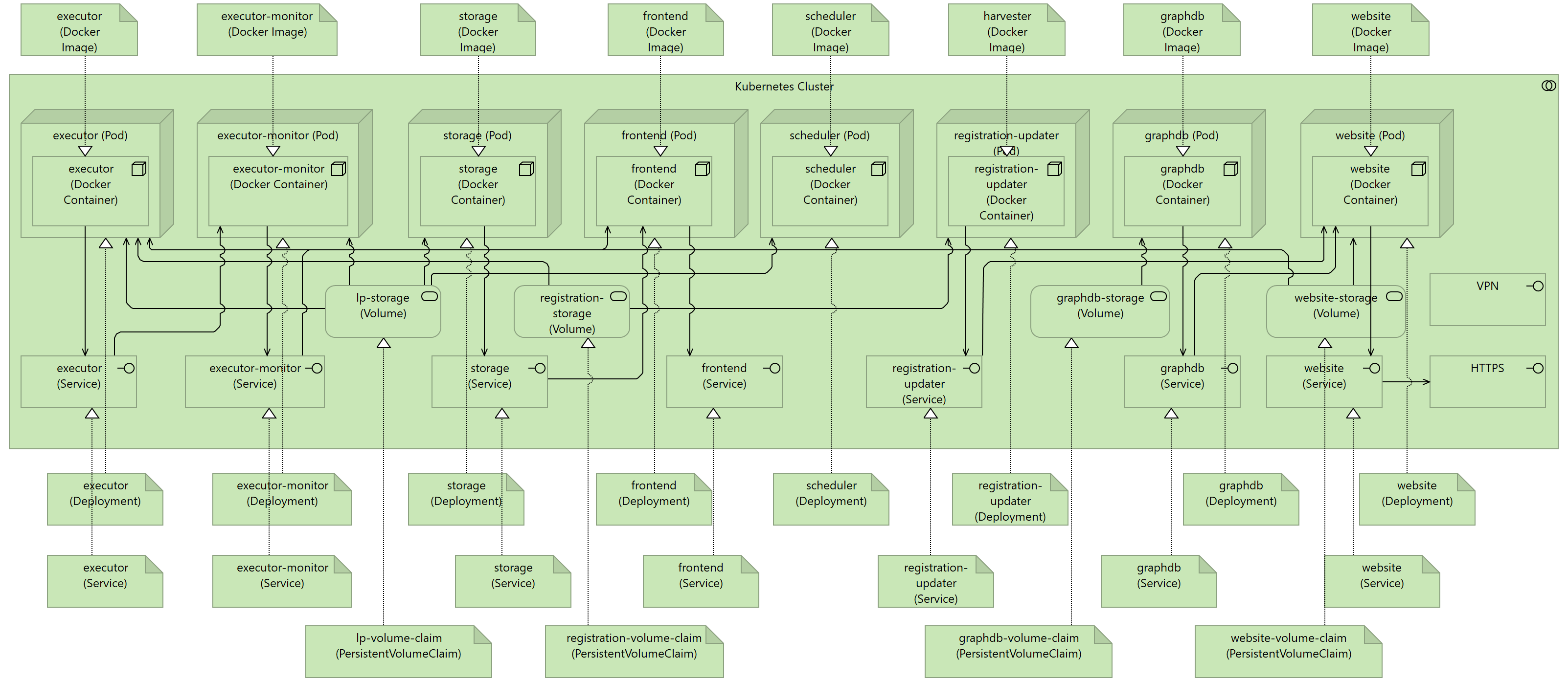
### Zoznam a popis navrhovaných integrácií

Metadatový procesor bude přistupovat k rozhraním LKODů jednotlivých poskytovatelů a k původním data.gov.sk za účelem harvestace metadat.

Libovolné externí systémy se mohou dotazovat SPARQL endpointu RDF úložiště NKOD nebo stáhnout RDF dumpy či CSV reporty přes Webovou stránku.

# FYZICKÁ ARCHITEKTÚRA

## prehľad



Obrázok 13 Technologická architektura

Komponenty aplikace budou nasazeny do prostředí Kubernetes (*Kubernetes Cluster*). Toto prostředí a jeho správa není součástí projektu a je zde použito pouze jako běhové prostředí, se kterým je komunikováno skrze standartní API. Jednou z možností je využití nástroje kubectl z příkazového řádku, který umožňuje nahrávat definice požadovaných zdrojů (resources). Fyzická architektura je tedy většinově definována definicemi zdrojů (yaml soubory), které jsou v Kubernetes vytvořeny.

### Požadavky na běhové prostředí

Aplikace vyžaduje dostupnost PersistentVolume o dostatečné velikosti. Dostatečná velikost je určena požadavky z definic na *lp-volume-claim (PersistentVolumeClaim)*, *registration-volume-claim (PersistentVolumeClaim), website-volume-claim (PersistentVolumeClaim)* a *graphdb-volume-claim (PersistentVolumeClaim)*.

PersistentVolume pro *lp-volume-claim (PersistentVolumeClaim), website-volume-claim (PersistentVolumeClaim)* a *registration-volume-claim* *(PersistentVolumeClaim)* musí být dostupné z více podů. Konkrétně pak *lp-volume-claim (PersistentVolumeClaim)* musí být dostupný z podů *executor (Pod)*, *executor-monitor (Pod)*, *storage (Pod)*, *frontend (Pod)* a *scheduler (Pod)*. *registration-volume-claim (PersistentVolumeClaim)*musí být dostupný z podů *executor (Pod)* a *registration-updater (Pod)*. *website-volume-claim (PersistentVolumeClaim)* musí být dostupný z podů *executor (Pod) a website (Pod).*

V případě *lp-volume-claim (PersistentVolumeClaim)* pak musí být změny synchronizovány v intervalu kratším než 1 vteřina.

Pro přístup z veřejné sítě internet, je dále třeba dostupnosti zvoleného portu z Kubetnetes úzlů (nodes). V případě potřeby je možné využít i alternativní řešení za předpokladu vhodné změny definice *website (Service)*.

### Konfigurace zdrojů (resources)

Konfigurace komponent je realizována skrze proměnné prostředí a úpravu definičních YAML souborů. Tato konfigurace zahrnuje zejména přístupové údaje a čísla portů.

### Definice nasazení (Deployments)

Z hlediska nasazení je definováno osm komponent nasazení (Deployment): *executor (Deployment)*, *executor-monitor* *(Deployment)*, *storage* *(Deployment)*, *frontend* *(Deployment)*, *scheduler* *(Deployment)*, *registration-updater* *(Deployment)*, *graphdb* *(Deployment)* a *website* *(Deployment)*. Každá tato komponenta odpovídá aplikační komponentě. Každá definice je zodpovědná za vytvoření podu odpovídajícího jména s odpovídajícím docker image. Mapování je následující:

|  |  |  |  |
| --- | --- | --- | --- |
| Deployment | Pod | Docker Container | Docker Image |
| *executor (Deployment)* | *executor (Pod)* | *executor (Docker Container)* | *executor (Docker Image)* |
| *executor-monitor* *(Deployment)* | *executor-monitor (Pod)* | *executor-monitor (Docker Container)* | *executor-monitor (Docker Image)* |
| *storage* *(Deployment)* | *storage (Pod)* | *storage (Docker Container)* | *storage (Docker Image)* |
| *frontend* *(Deployment)* | *frontend (Pod)* | *frontend (Docker Container)* | *frontend (Docker Image)* |
| *scheduler* *(Deployment)* | *scheduler (Pod)* | *scheduler (Docker Container)* | *scheduler (Docker Image)* |
| *registration-updater* *(Deployment)* | *registration-updater (Pod)* | *registration-updater (Docker Container)* | *registration-updater(Docker Image)* |
| *graphdb* *(Deployment)* | *graphdb (Pod)* | *graphdb (Docker Container)* | *graphdb (Docker Image)* |
| *website* *(Deployment)* | *website (Pod)* | *website (Docker Container)* | *website (Docker Image)* |

### Definice sekundárního úložiště (PersistentVolumeClaim)

Data jsou ukládata do sekundární paměti (souborový systém) do definovaných oddílu (PersistentVolumeClaim). Tyto definice pak tvoří odpovídající zdroje (resources) v Kubernetes.

|  |  |
| --- | --- |
| PersistentVolumeClaim | Definovaný volume |
| *lp-volume-claim (PersistentVolumeClaim)* | *lp-storage (Volume)* |
| *registration-volume-claim (PersistentVolumeClaim)* | *registration-storage (Volume)* |
| *graphdb-volume-claim (PersistentVolumeClaim)* | *graphdb-storage (Volume)* |
| *website-volume-claim (PersistentVolumeClaim)* | *website-storage (Volume)* |

### Definice služeb (Service)

Pro potřeby komunikace mezi Pody jsou vytvořeny služby (Services) dle odpovídajích definice. Pro každou komponentu, respektive Pod, který obsahuje síťové rozhraní je definována odpovídající služba (Service). Příslušnost služeb (Service) k Podům je popsána v následující tabulce.

|  |  |  |
| --- | --- | --- |
| Služba (Service) | Definovaná Služba | Příslušný Pod |
| *executor (Service)* | *executor (Service)* | *executor (Pod)* |
| *executor-monitor (Service)* | *executor-monitor (Service)* | *executor-monitor (Pod)* |
| *storage (Service)* | *storage (Service)* | *storage (Pod)* |
| *frontend (Service)* | *frontend (Service)* | *frontend (Pod)* |
| *registration-updater (Service)* | *registration-updater (Service)* | *registration-updater (Pod)* |
| *graphdb (Service)* | *graphdb (Service)* | *graphdb (Pod)* |
| *website (Service)* | *website (Service)* | *website (Pod)* |

Služba *website (Service)* slouží ke zpřístupnění aplikace z vnější sítě.

### Docker (Docker Image, Docker Container)

Docker container vzniká spuštěním Docker Image. Mapování mezi definicí Docker Image a běžícím Docker Container je následující:

|  |  |
| --- | --- |
| Docker Image | Docker Container |
| *executor (Docker Image)* | *executor (Docker Container)* |
| *executor-monitor (Docker Image)* | *executor-monitor (Docker Container)* |
| *storage (Docker Image)* | *storage (Docker Container)* |
| *frontend (Docker Image)* | *frontend (Docker Container)* |
| *scheduler (Docker Image)* | *scheduler (Docker Container)* |
| *harvester (Docker Image)* | *registration-updater (Docker Container)* |
| *graphdb (Docker Image)* | *graphdb (Docker Container)* |
| *website (Docker Image)* | *website (Docker Container)* |

Jednotlivé Docker Image odpovídají aplikačním komponentám. Samotné Docker Images jsou uložené v repozitáři (repository) přístupném z *Kubernetes Cluster*. Verze použitých Docker Images jsou definovány jako součást definice nasazení (Deployments). Při publikaci nové verze je tedy třeba tyto definice upravit.

Pro potřebu tvorby Docker Images se předpokládá využití prostředí poskytnutého zadavatelem.

### Příchozí komunikace

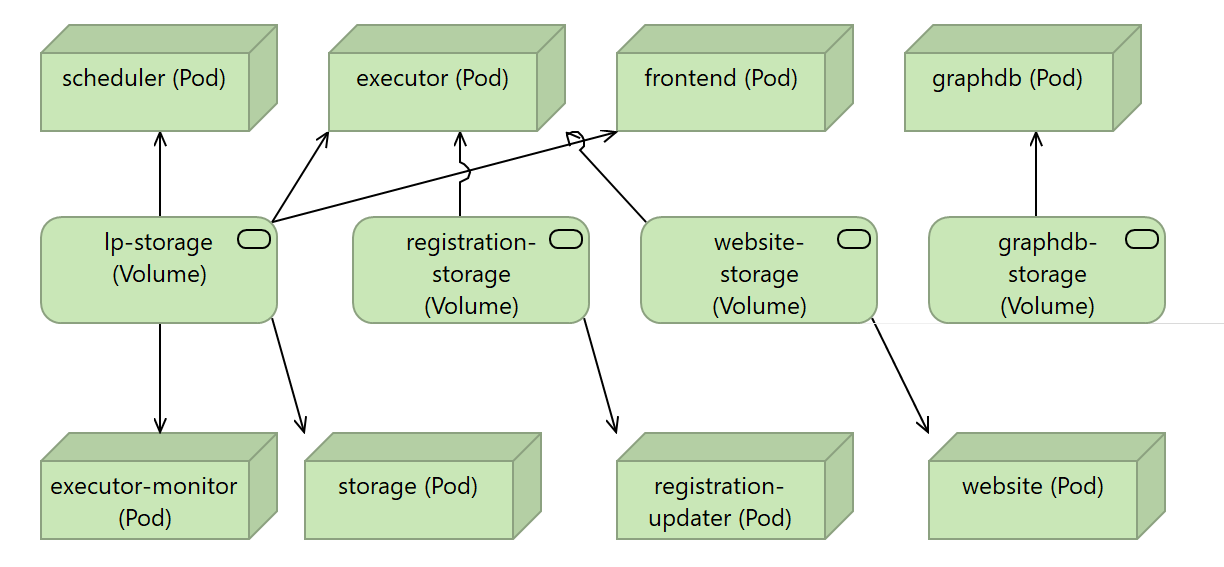
Vstupní bránou do aplikace je služba *website (Service)*. Tato služba je obsluhována *website (Pod)* který funguje jak proxy a zároveň poskytuje statické webové stránky. Mezi tyto stránky patří zejména (*Webová Stránka*). Tato skutečnost je v diagramu (sekce 11.1 PREHĽAD) zachycena jako obsluha *HTTPS*.

Pro potřeby administrace je možné se připojit na jednotlivé Pody pomocí SSH. Pro tyto potřeby se počítá s připojeného skrze *VPN* do sítě *Kubernetes Cluster.*

## KOMUNIKÁCIA

Pro potřeby komunikace, sdílení dat, mezi Pody je využito sekundární paměti a síťového rozhraní.

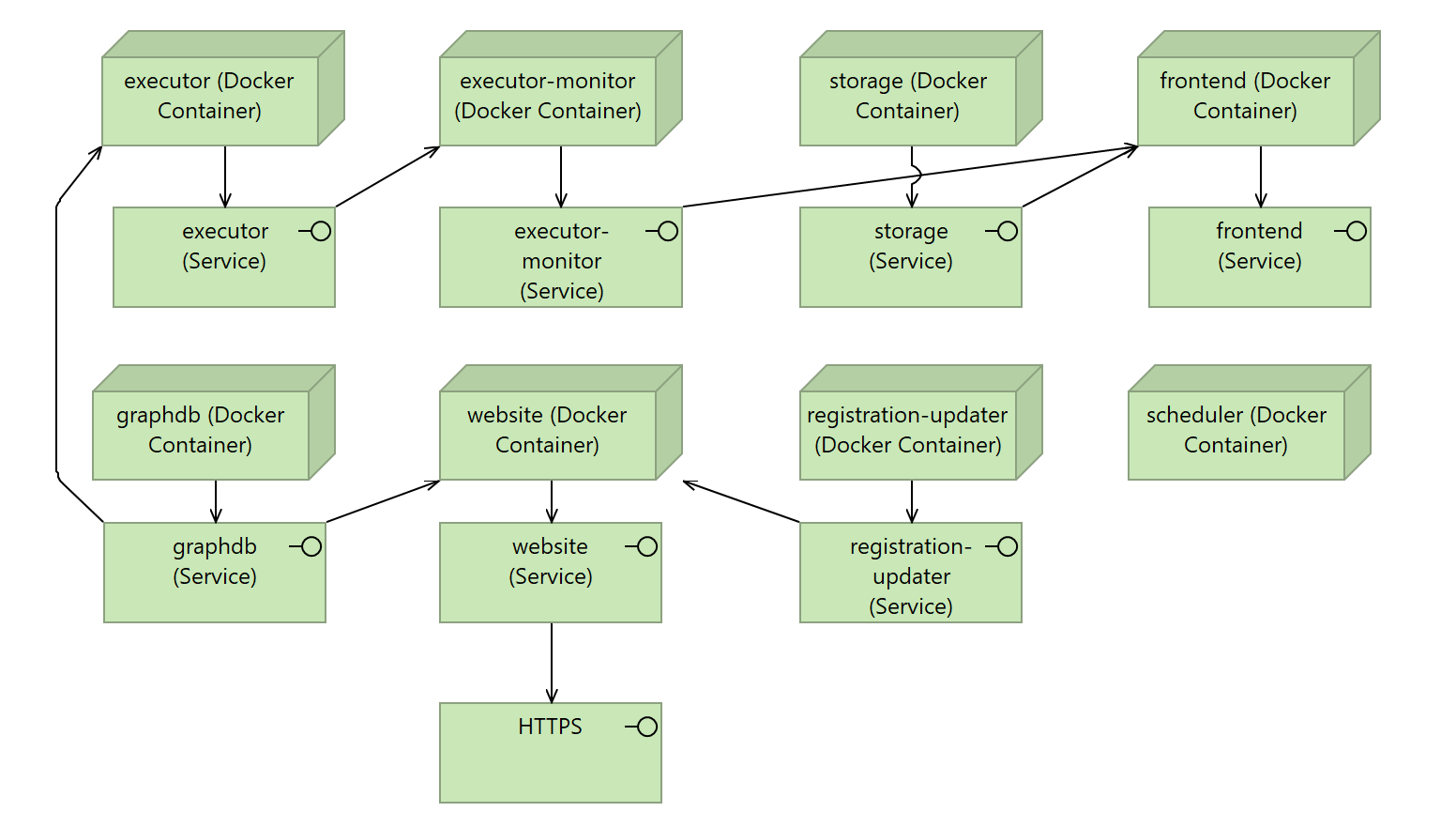
### Komunikace pomocí sekundární paměti



Obrázok 14 Úložiště na technologické vrstvě

Pro potřeby komunikace je použito zejména *lp-storage (Volume)*, *registration-storage (Volume)* a *website-storage (Volume)*. První slouží pro uložení konfigurace LinkedPipes ETL, definic pipeline a jejich běhů. Toto úložiště je využíváno pouze pro potřeby komponent v rámci LinkedPipes ETL. Úložiště *registration-storage (Volume)* slouží pro zrcadlení registrovaných katalogizačních záznamů. Úložiště *website-storage (Volume)* slouží pro uložení dumpů pro jejich zpřístupnení skrze *website (Pod)*.

### Komunikace pomocí síťového rozhraní



Obrázok 15 Komunikace na technologické vrstvě

Pro síťovou komunikaci je téměř výlučně využíváno protokolu HTTP. Pro odpovídající komponenty jsou definovány služby (Services), které umožňují přístup k Podům a tím i Docker Containerům. Komunikace probíhá na úrovni dvou komunikačních celků: LinkedPipes ETL a veřejný přístup skrze webovou stránku a SPARQL endpoint.

Pro potřeby nahrání dat do RDF databáze je využito komunikace mezi *graphdb (Docker Container)* a *executor (Docker Container)*. Z hlediska přísupu po sít nejsou komponenty LinkedPipes ETL dosažitelné z vnějšího prostředí.

Pro přístup z vnějšího, prostředí, zobrazeno jako HTTPS, slouží služba *website (Service)*. Ta poskytuje statické soubory. Komponenta dále funguje jako Proxy pro *graphdb (Service)*, čímž zpřístupňuje SPARQL endpoint. Komponenta dále funguje, dočasně, jako proxy pro *registration-updater (Service)*.

## CI/CD PROCESS

Použit je převážně již existující software. Projekt spočívá v jeho konfiguraci a nasazení. CI/CD proces tedy není požadován. Výjimku tvoří jen webová stránka pro dotazování, pro kterou by implementace CI/CD procesů byla neadekvátní její nízké složitosti.

## KONTAJNERIZÁCIA A ORCHESTRÁCIA

Jednotlivé části produktu, tj. LinkedPipes ETL, Updater registrací,  Ontotext GraphDB a webová stránka budou dodány k nasazení jako Docker image.

Nasazení proběhne do prostředí Kubernetes. Definice zdrojů (resources) bude dodáva v podobě YAML souborů.

## VYUŽITIE VLÁDNEHO CLOUDU (EZ)

Bude využit komerční cloud.

## POPIS TESTOVACIEHO PROSTREDIA

Bude využit komerční cloud s podporou Kubernetes.

## POPIS PRODUKČNÉHO PROSTREDIA

Bude využit komerční cloud s podporou Kubernetes.

# SOFTVÉROVÉ LICENCIE

1. Licence Ontotext GraphDB Free: <https://graphdb.ontotext.com/LICENSE-GraphDB-Free.txt>
2. Licence LinkedPipes ETL: [MIT License](https://github.com/linkedpipes/etl/blob/develop/LICENSE)

# ZÁLHOVANIE A OBNOVA RIEŠENIA

## POPIS RIEŠENIA PRE ZÁLOHOVANIE a OBNOVu

Celé prostředí bude nasaditelné jako sada Docker images, včetně konfigurace, které budou uloženy mimo produkční prostředí. V případě jeho havárie tedy lze celé prostředí znovu nasadit.

V případě havárie RDF úložiětě lze obsah databáze kdykoliv obnovit z datových souborů, které jsou uloženy v souborovém systému. Databázi samotnou tedy není třeba zálohovat. V LinkedPipes ETL k tomu bude speciální pipeline.

Národní katalog otevřených dat samotný je tvořen daty, která jsou pravidelně harvestována z lokálních katalogů otevřených dat. Harvestaci lze provést v libovolný okamžik, tedy i při obnově po havárii.

Jediná data, která je třeba zálohovat, jsou registrační záznamy jednotlivých datasetů a LKODů. V rámci tohoto projektu NKOD je registrace zabezpečená Administrátorem NKOD – registrační záznamy jsou primárně uloženy do GitHub repozitáře, odkud jsou do prostředí NKOD kopírovány při změně pomocí komponenty „updater registrací“. Tedy v případě havárie prostředí NKOD nebudou ztraceny.

V rámci řešení uživatelského rozhraní pro registraci datasetů a LKODů v dalších částech OD2.0 bude třeba vyřešit, jak se záznamy dostanou do prostředí NKOD místo stávajícího způsobu, a v rámci tohoto bude třeba vyřešit spolehlivé zálohování těchto záznamů.

# ZABEZPEČENIE DOSTUPNOSTI

## POPIS RIEŠENIA NA DOSTUPNOSŤ SYSTÉMU IS

Požadavek na dostupnost:

|  |  |  |
| --- | --- | --- |
| **70** | Dostupnosť služieb | Požiadavka na dostupnosť služieb: - Dostupnosť nových služieb okrem plánovaných výpadkov je 90% v režime 12/5, - Plánovaný výpadok je oznámený minimálne 7 dní vopred, - Plánovaný výpadok nie je dlhší ako 8 hodín a je prioritne medzi 20:00 – 08:00 počas pracovných dní, sobota, nedeľa, štátny sviatok, deň pracovného pokoja. |

Dostupnost služeb je zajištěna jejich nasazením do cloudového prostředí s požadovaným SLA a včasným plánováním údržby.

# BEZPEČNOSŤ, AUTORITY MANAŽMENT

## ZABEZPEČENIE KOMUNIKÁCIE A INTEGRITY

Jednotlivé komponenty (LinkedPipes ETL a RDF úložiště) mezi sebou komunikují po vnitřní síti, čímž je bezpečnost a integrita jejich komunikace zajištěna.

Veřejnosti bude přístupný SPARQL endpoint provozovaný RDF úložištěm. Ten bude zabezpečen tak, že veřejnosti bude poskytovat přístup pouze pro čtení, což je realizováno přímo konfigurací zvolené implementace Ontotext GraphDB.

Jak ke SPARQL endpointu, tak ke statické webové stránce s formulářem pro zaslání SPARQL dotazu na endpoint bude přístup zajištěn pomocí protokolu HTTPS konfigurovaného dle doporučení <https://ssl-config.mozilla.org/> na úrovni „Modern“, tj. pouze skrz TLS 1.3.

Metadatový procesor (LinkedPipes ETL) pak komunikuje s webovými servery poskytovatelů otevřených dat, taktéž výhradně pomocí HTTPS.

## RIADENIE PRÍSTUPU (AUDIT TRAILS – SLEDOVANIE LOGOVANIA DO SYSTÉMU)

Systém neobsahuje části, do kterých by měl být řízen uživatelský přístup pro veřejnost. SPARQL endpoint je veřejnosti k dispozici pouze v režimu read-only.

Přístup pro zápis do RDF úložiště bude řešen vytvořením systémového účtu pro zápis do databáze NKOD pomocí vnitřního mechanismu správy uživatelů Ontotext GraphDB.

Administrátorský přístup do celého prostředí NKOD pak bude možný pouze přes VPN cloudového prostředí, která poskytne i logování těchto přístupů.

## DEFINOVANIE POUŽÍVATEĽSKÝCH ROLÍ

K read-only SPARQL endpointu a webové stránce má přístup veřejnost. Uživatelské role není třeba definovat.

## METODIKA MANAŽMENTU POUŽÍVATEĽOV

K read-only SPARQL endpointu a webové stránce má přístup veřejnost. Uživatelské role není třeba definovat.

# KONVERZIE DÁT, MIGRÁCIA DÁT, DÁTOVÝ MODEL A NÁVRH RIEŠENIA

## POSTUP VYKONANIA MIGRÁCIE DÁT

Migrace dat z původního katalogu data.gov.sk bude provedena pomocí separátní pipeline v LinkedPipes ETL tak, aby výsledek odpovídal specifikaci DCAT-AP-SK. Vstupem migrace budou i informace, které datasety se mají sdružit do datových sérií, a které naopak rozpadnout do datové série. Výstup této pipeline bude začleněn do obsahu NKOD při pravidelné tvorbě aktuálního obsahu NKOD.

## OVERENIE MIGRÁCIE DÁT

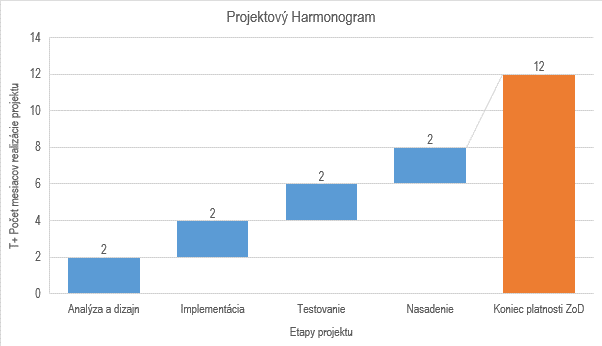
Migrace dat bude ověřena

1. namátkovou kontrolou vybraných záznamů a jejich zmigrovaných ekvivalentů
2. kontrolou počtu datových sad jednotlivých poskytovatelů s přihlédnutím k rozpadu/sloučení některých datových sad do datových sérií

Časť 3: Detailný návrh postupov a produktov pre fázy Testovanie a Nasadenie do produkcie.

## FÁZY PROJEKTU A ČASOVÝ HARMONOGRAM – KONSOLIDOVANÝ PREHĽAD

Následující harmonogram zobrazuje nejdelší možné trvání projektu. Odhady reálných trvání jednotlivých fází jsou v následujících kapitolách.



### Fáza - Analýzy a Dizajnu

Odhad: 20.4.2022 – 20.6.2022

### Fáza - Implementácie a Testovania

Odhad: 20.6.2022 – 20.8.2022

### Fáza - Nasadenia a Post-Implementačnej podpory – príprava a nasadenie do prevádzky

Odhad: 20.8.2022 – 20.9.2022

### Ukončenie projektu

Odhad: 20.9.2022 – 20.10.2022

# ZÁVISLOSTI A NÁVRH RIEŠENIA

## ZÁVISLOSTI (PROJEKTOVÉ/MEDZIPROJEKTOVÉ)

Projekt NKOD je jednou ze součástí projektu OD2.0, které probíhají jak paralelně, tak po projektu NKOD. Styčnými body jsou

1. **Vstup:** Úložiště registračních záznamů datasetů a LKODů. V aktuálním projektu jsou vyřešeny GitHub repozitářem, jehož obsah se synchronizuje do prostředí NKOD a který spravuje Administrátor NKOD. To bude nahrazeno řádným uživatelským rozhraním pro poskytovatele dat, které bude vyvinuto v jiných částech OD2.0.

2. **Výstup:** Výstupní data jsou uložena v RDF úložišti a přístupná přes SPARQL endpoint. Metadata datasetů v úložišti odpovídají specifikaci DCAT-AP-SK. Data o měření kvality odpovídají specifikaci v tomto dokumetnu.

# DOKUMENTÁCIA

Budou dodány následující dokumenty:

|  |  |
| --- | --- |
| Dokument | Bude vytvorené v rámci projektu? [Áno/Nie] |
| Aplikačná príručka | Áno |
| Inštalačná príručka | Áno |
| Konfiguračná príručka | Áno |
| Integračná príručka | Áno |
| Používateľská príručka | Áno |
| Prevádzkový popis | Áno |
| Bezpečnostný projekt (podľa metodiky zabezpečenia) | Áno |
| Popis produktu a celého riešenia z pohľadu nasadenia do produkčného prostredia | Áno |
| Pokyny pre inštaláciu (úvodnú aj opakovanú) | Áno |
| Pokyny pre prevádzku a administráciu | Áno |
| Pokyny pre servis a údržbu | Áno |
| Pokyny pre používanie | Áno |
| Pokyny pre diagnostiku (konfiguračná príručka) | Áno |
| Pokyny pre obnovu v prípade výpadku alebo havárie (disaster recovery plan / havarijný plán) | Áno |
| Ďalšie (doplňte podľa potreby, napr. Príručka pre školenie) |  |
|  |  |
|  |  |
|  |  |
|  |  |

# OTVORENÉ OTÁZKY

|  |  |  |
| --- | --- | --- |
| ID | Otázka | Dopad |
| 1 | Dostupnost komerčního cloudu | Testovací a produkční prostředí má běžen v komerčním cloudu. Pokud dojde ke zpoždění jeho dostupnosti, může to ovlivnit harmnogoram nasazování a testování. |

# PRÍLOHY

1. Katalóg požiadaviek - KATALOG\_POZIADAVIEK\_OD2.0-NKOD.xlsx
2. Specifikace DCAT-AP-SK
3. Plán testů