





**jakubklimek** Pridání odkazů na základní dokumentaci

 History

 1 contributor

 117 lines (82 sloc) | 6.02 KB

...

# Integračná příručka - Národní katalog otevřených dat

## 1. ROZSAH PLATNOSTI A ÚČEL

Tato integrační příručka slouží k popisu Národního katalogu otevřených dat (NKOD), části projektu OD2.0.

## 2. DEFINÍCIA POJMOV A SKRATIEK

### **NKOD**

Národní katalog otevřených dat

### **LKOD**

Lokální katalog otevřených dat

### **RDF**

Resource Description Framework

## 3. REFERENCIE

- Kubernetes
- Oracle Cloud Infrastructure (OCI)
- Inštalčná príručka a pokyny na inštaláciu (úvodnú/opakovanú)
- Aplikační příručka
- Uživatelská příručka

The diagram illustrates the architecture of the National Catalog of Open Data (Národní katalog otevřených dat). It is divided into several main sections:

- External Data Sources and Providers:**
  - Původní data.gov.sk data.gov.sk CKAN API:** Original data sources.
  - LKOD1 CKAN API:** Local Catalog of Open Data 1.
  - LKOD2 DCAT-AP Dokumenty:** Local Catalog of Open Data 2.
  - LKOD3 DCAT-AP SPARQL Endpoint:** Local Catalog of Open Data 3.
  - Poskytovatel dat:** Data provider.
- Central Processing Unit (Metadatový procesor LinkedPipes ETL):**
  - Frontend:** Interface for data migration and harvesting.
  - Executor:** Executes data migration and harvesting tasks.
  - Executor-monitor:** Monitors the execution of tasks.
  - Storage:** Stores data and metadata.
  - Migrace původního data.gov.sk:** Migration of original data.gov.sk data.
  - Harvestace lokálních katalogů a registrovaných zpracování datasetů:** Harvesting of local catalogs and registered dataset processing.
  - Měření dostupnosti zdrojů:** Measurement of source availability.
  - Měření kvality metadat:** Measurement of metadata quality.
  - Generování reportů kvality a dostupnosti:** Generation of quality and availability reports.
- Data Storage and Management:**
  - Úložiště registrací LKODů:** LKOD registration storage.
  - Úložiště registrací datasetů:** Dataset registration storage.
  - Úložiště registrací poskytovatelů:** Provider registration storage.
  - Manažer:** Manager of the catalog.
  - Dočasný GitHub repozitář registrací:** Temporary GitHub repository for registrations.
  - Filesystem:** File system for storing data and reports.
  - Úložiště RDF dumpů a CSV reportů:** Storage for RDF dumps and CSV reports.
- External Services and Users:**
  - Externí IS:** External information system.
  - Poskytovatel dat:** Data provider.
  - Návštěvník:** Visitor.
  - Administrátor NKOD:** NKOD administrator.

Tento popis architektury také ukazuje komunikačními rozhraní, které je většinou také integračními body. Ve zbytku této kapitoly nejprve zevrubně popíšeme jednotlivá komunikační rozhraní a následně integrační body.

Komunikační rozhraní jsou součástí rozhraní popsanych v [Aplikační příručce](#) sekce 3.5 POPIS EXISTUJÍCICH ROZHRAŇÍ V IS. Pro plné porozumění rozhraním je doporučeno přečíst jak tuto sekci tak sekci odkázanou.

- Poskytovali dat
- Externímu IS

- [Návštevník](#)

### 3.1.1 Webová stránka

Webová stránka umožňuje vybrat před připravený [SPARQL](#) dotaz nebo napsat vlastní, a nechat ho vykonat na rozhraní SPARQL endpoint nad RDF úložištěm NKOD a zobrazit výsledky. Přes komponentu webové stránky je zpřístupněn i obsah NKOD v podobě souborů ke stažení. Adresy souborů ke stažení jsou popsány v [Uživatelské příručce](#).

Webová stránka je k dispozici vícejazyčně, nyní v 5 jazycích dostupných na relativní URL:

- / - Slovenština
- /de/ - Němčina
- /en/ - Angličtina
- /hu/ - Maďarština
- /uk/ - Ukrajinština

Pro samotnou práci se SPARQL dotazem využívá knihovnu [Yasgui](#). Tato knihovna je součástí repozitáře [NKOD-SW](#) v cestě `components/website/www/libs`.

### 3.1.2 SPARQL endpoint

Vyhodnocuje dotazy v jazyce [SPARQL](#) nad RDF úložištěm NKOD. Toto rozhraní je také využíváno formulářem pro SPARQL dotaz na Webové stránce.

Endpoint je ve výchozím nastavení dostupný na URL `/api/sparql` a využívá [CORS](#). Toto nastavení je možné změnit v `configuration/nginx.conf`.

### 3.1.3 LKODY

Vstupním rozhraním pro Metadatový procesor - LinkedPipes ETL jsou registrované instance LKODů. Jejich obsah musí odpovídat specifikaci [DCAT-AP-SK 2.0](#).

V oblasti vstupní komunikace dochází k získávání dat potřebných pro vytvoření dat NKODu. Hlavním datovým vstupem jsou lokální katalogy LKOD, ze kterých jsou data harvestována dle specifikace DCAP-AP-SK 2.0.

## 4. INTEGRAČNÉ API

---

Za integrační API lze považovat i komunikační rozhraní popsaná v sekci 3.1.

Komunikačné rozhrania. Tato komunikační rozhraní slouží zejména pro integraci z venku systému, tedy třetích stran. Rozhraní popsaná v této sekci slouží pro integraci uvnitř NKOD.

## 4.1 Registrační záznamy

Struktura synchronizačních záznamů je popsána v [Uživatelské příručce](#), sekce 3.2 Registrační záznamy. Ve výchozím nastavení jsou registrační záznamy načítány z GitHub repozitáře [NKOD-REG](#). Z tohoto umístění jsou před každým spuštěním harvestování synchronizovány do `nodc-registration-pvc` definovaném v `./k8s/linkedpipes/registration-pvc.yaml`. Synchronizaci zajišťuje skript z komponenty Manažer. Konkrétně se jedná o skript v repozitáři [NKOD-SW](#) v souboru `./components/manager/entrypoint.sh`.

Ve sdíleném úložišti musí být souboru uložené v adresáři `repository`. Důvodem je nemožnost vytvoření git repository přímo v rootu PVC, neb ten obsahuje skrytý soubor pro potřeby OCI.

URL repozitáře s registračními záznamy je nastavené v konfiguračním souboru `./configuration/nodc-configuration.properties`.

Toto je integrační bod využitelný pro budoucí Portál otevřených dat, kde se registrační záznamy budou spravovat a ukládat jiným způsobem, který nahradí stávající synchronizaci z GitHub repozitáře.

## 4.2 LinkedPipes ETL pipeliney

Metadatový procesor - LinkedPipes ETL provádí harvestování na základě definovaných pipeline. Před každým spuštěním se provede stažení pipeline a template z [NKOD-PPL](#). Stažené pipeliney a templaty jsou následně zkopírovány do úložiště LinkedPipes ETL a importovány.

URL repozitáře je nastavené v konfiguračním souboru `./configuration/nodc-configuration.properties`.

[Give feedback](#)