



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

C2V7

Návrh zajištění podpory katalogizace dat ve veřejném datovém fondu v NKOD

Vytvořeno v rámci projektu

Rozvoj datových politik v oblasti zlepšování kvality
a interoperability dat veřejné správy

CZ.03.4.74/0.0/0.0/15_025/0013983

Klíčová aktivita: 03 Návrhy a realizace opatření pro zlepšování kvality
otevřených dat

Verze výstupu: 01



Navrhované rozšíření funkcionalit NKOD

V rámci tohoto výstupu jsou navrženy nové funkcionality rozšiřující stávající stav Národního katalogu otevřených dat (NKOD) o podporu dat publikovaných ve veřejném datovém fondu (VDF). Tyto funkcionality závisí na existenci podkladových otevřených dat publikovaných z Registru práv a povinností (RPP). V době psaní tohoto výstupu potřebná data z RPP ještě publikována nebyla, a tak byla připravena umělá data, dle předpokládané podoby, pro účely ukázání funkcionalit ve vyvinutém prototypu.

Označení datové sady náležející do VDF

Pro tuto funkcionality není třeba rozšiřovat katalogizační záznam NKOD ani vstupní registrační formulář. Fakt, že datová sada náleží do VDF, bude zjištěn z dat RPP v průběhu harvestace NKOD, a příslušné datové sady budou označeny jako náležející do VDF pomocí jejich zařazení do třídy <https://data.gov.cz/slovník/nkod/vdf/datová-sada-publikující-veřejné-údaje>.

Pokud je datová sada číselníkem, bude navíc označena třídou <https://data.gov.cz/slovník/nkod/vdf/číselník>.

Tento fakt pak bude zohledněn v uživatelském rozhraní NKOD v detailu datové sady pomocí odznaku VDF pro datové sady publikující veřejné údaje:

OTEVŘENÁ DATA Vzdělávání Články Datová kvalita Datové sady Poskytovatelé Klíčová slova Další 🇨🇪

Agendy VDF: veřejné údaje 🔗 📄 🗑️

Ministerstvo vnitra

Agendy evidované v Registru práv a povinností ve smyslu § 51 zákona č. 111/2009 Sb. o základních registrech.

agenda registr práv a povinností rpp činnost

Téma Vláda a veřejný sektor 🔗	Pojmy ze sémantického slovníku Agenda 🔗 Činnost 🔗 Ustanovení právního předpisu 🔗 Právní předpis 🔗	Dokumentace Zobrazit dokumentaci 🟢 HTTP	Periodicita aktualizace denní 🔗
Klasifikace dle EuroVoc orgány veřejné správy 🔗 územní samospráva 🔗 regionální orgány státní správy 🔗	Vztahy ze sémantického slovníku	Kontaktní bod Help desk správy základních registrů (SZR)	
		Specifikace OFN: RPP - agendy 🟢 HTTP	

Návrh kódu pro odznak: `VDF: veřejné údaje`



OTEVŘENÁ DATA

Vzdělávání Články Datová kvalita Datové sady Poskytovatelé Klíčová slova Další

Číselník povinností sběru adres VDF: číselník

Ministerstvo vnitra

Původce dat: Český statistický úřad

Číselník povinností sběru adres použitý v Registru práv a povinností ve smyslu § 51 zákona č. 111/2009 Sb. o základních registrech.

číselník

povinnost sběru adresy

registru práv a povinností

rpp

Téma

Vláda a veřejný sektor

Související geografické území

Česká republika

Kontaktní bod

Help desk správy základních registrů (SZR)

Periodicita aktualizace

nepravidelný

Klasifikace dle EuroVoc

výkonná moc

Návrh kódu pro odznak: `VDF: číselník`

Dále bude zobrazena informace o původci dat v případě, že je odlišný od poskytovatele dat. Data s touto informací by měla být dostupná v otevřených datech RPP.

Vyhledávání datových sad přes náležitost do VDF

V uživatelském rozhraní NKOD bude umožněno vybrat datové sady náležející do VDF pomocí pokročilého filtru "Veřejný datový fond".

OTEVŘENÁ DATA

Vzdělávání Články Datová kvalita Datové sady Poskytovatelé Klíčová slova Další

Poskytovatelé (30)

Ministerstvo vnitra (8 063)
Český statistický úřad (849)
Úřad průmyslového vlastnictví (762)
Ministerstvo financí (736)

Časové pokrytí od do
Veřejný datový fond: Veřejné údaje Číselník

Nalezeno 11 182 datových sad

Výpis OVM publikujících ve VDF

Zjistit poskytovatele (OVM) publikující ve VDF bude možno dvěma způsoby.

První způsob využívá funkcionalitu "Vyhledávání datových sad přes náležitost do VDF". Po zaškrtnutí typu datové sady ve VDF filtru budou vyfiltrováni právě ti poskytovatelé dat, kteří publikují datové sady do VDF. Tento způsob však neumožňuje filtrovat dle původce dat.

Druhý způsob zjištění, zda poskytovatel poskytuje datové sady VDF, bude označení poskytovatele dat VDF v seznamu poskytovatelů pomocí odznaku VDF, jež lze prokliknout na seznam datových sad VDF daného poskytovatele. Pokud původce dat není poskytovatelem dat do NKOD, tak se v tomto seznamu neobjeví.



Nalezeno 30 poskytovatelů

Ministerstvo vnitra 8 063 datových sad	Český statistický úřad 849 datových sad	Úřad průmyslového vlastnictví 762 datových sad	Ministerstvo financí 736 datových sad
Statutární město Brno 144 datových sad	Ministerstvo zdravotnictví VDF poskytovatel původce 124 datových sad	Ministerstvo zemědělství 107 datových sad	Česká správa sociálního zabezpečení 84 datových sad

Návrh kódu pro odznaky:

```
<li class="list-group-item">  
  <span class="h6 px-1"><span class="badge bg-info">VDF</span></span><span  
class="h6">  
  <span class="badge bg-info">poskytovatel</span></span>  
  <span class="h6"><span class="badge bg-info">původce</span></span>  
</li>
```

Zobrazení číselníků z VDF využívaných v datových sadách publikujících ve VDF veřejné údaje a naopak, zobrazení datových sad publikujících veřejné údaje využívajících číselník ve VDF v zobrazení jeho detailu

Tyto informace se dotáhnou online ze SPARQL endpointu RPP. Dokud v něm potřebné údaje nebudou, prototyp je bude získávat z předpřipravených dat.

Datové sady typu číselník

Pro datovou sadu publikující číselník bude přidán seznam datových sad obsahujících údaje kódované tímto číselníkem:



Datové sady VDF publikující údaje kódované tímto číselníkem

Seznam říčních plavidel

Poskytovatel: Ministerstvo vnitra, Původce: Český statistický úřad

Datová sada obsahuje říční plavidla

JSON-LD

Seznam ponorek provozovaných na Vltavě

Poskytovatel: Ministerstvo vnitra, Původce: Český statistický úřad

Datová sada obsahuje ponorky na Vltavě

JSON-LD

SPARQL dotaz pro získání seznamu datových sad:

```
PREFIX pojmy: <https://slovník.gov.cz/legislativní/sbírka/111/2009/pojem/>

SELECT DISTINCT ?ds WHERE {
  [] a pojmy:údaj ;
    pojmy:je-kódovaný-číselníkem/pojmy:iri-číselníku-v-nkod
  <https://data.gov.cz/zdroj/datové-
  sady/00007064/7e1c097e0b500f25d0d5a33832dd3dff> ; #toto je aktuálně zobrazený
  číselník
    pojmy:iri-datové-sady-publikující-veřejný-údaj ?ds .
}
```

Datové sady publikující veřejný údaj

Pro datovou sadu publikující veřejné údaje seznam datových sad s použitými číselníky:

Číselníky VDF kódující údaje této datové sady

Barva plavidla

Poskytovatel: Ministerstvo vnitra, Původce: Český statistický úřad

Datová sada obsahuje číselník barev plavidel

JSON-LD

Velikost plavidla

Poskytovatel: Ministerstvo vnitra, Původce: Český statistický úřad

Datová sada obsahuje číselník velikostí plavidel

JSON-LD

SPARQL dotaz pro získání seznamu číselníků:

```
PREFIX pojmy: <https://slovník.gov.cz/legislativní/sbírka/111/2009/pojem/>

SELECT DISTINCT ?ciselnik WHERE {
```



```
[ ] a pojmy:údaj ;  
    pojmy:je-kódovaný-číselníkem/pojmy:iri-číselníku-v-nkod ?ciselnik ;  
    pojmy:iri-datové-sady-publikující-veřejný-údaj  
<https://data.gov.cz/zdroj/datové-  
sady/00007064/9c73b802263c5e0ccf5542f10fbc35bb> . #toto je aktuálně  
zobrazená datová sada publikující veřejný údaj kódovaný číslníkem  
}
```

Dokumentace prototypu implementujícího návrh nových funkcionalit NKOD

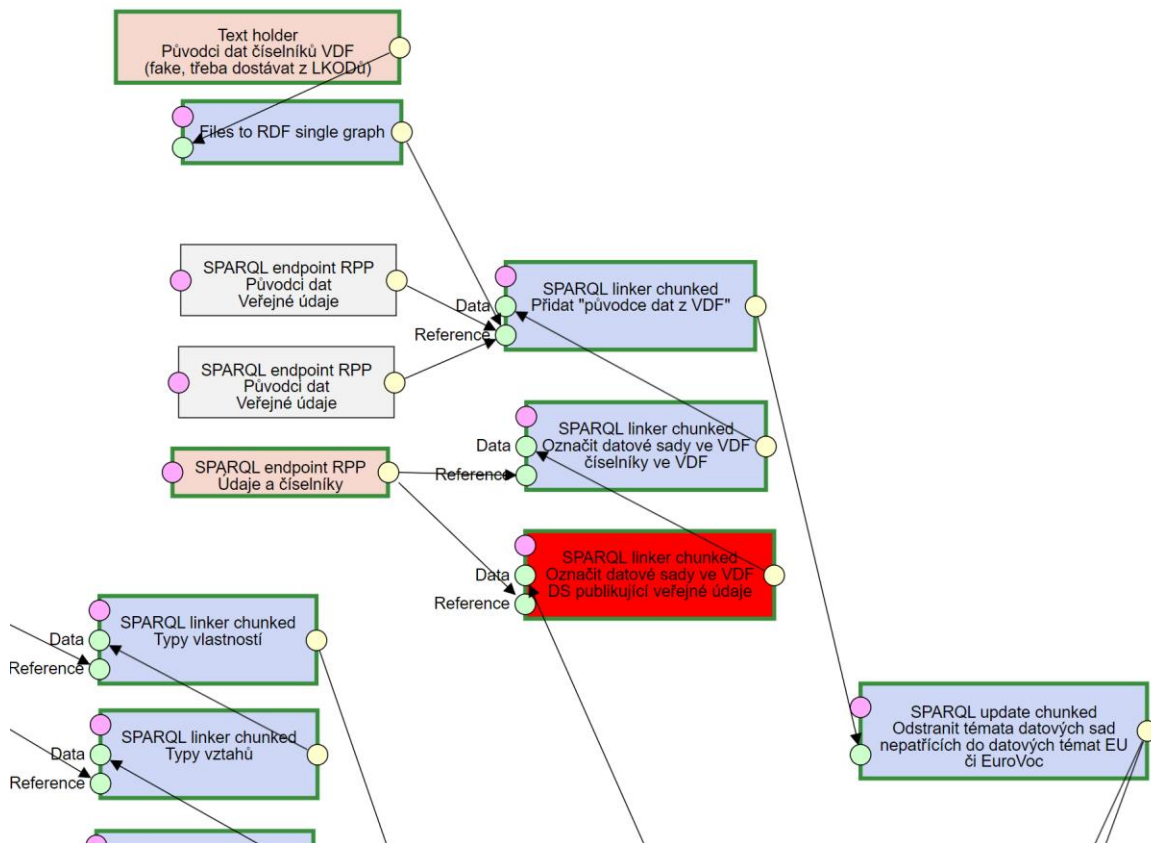
Předpoklady, které je nutné splnit před nasazením prototypu do produkce

1. Národní katalog otevřených dat (NKOD) má k dispozici SPARQL endpoint Registru práv a povinností (RPP), který je aktuálně dostupný na <https://rpp-opendata.egon.gov.cz/odrpp/sparql/>
2. Ve SPARQL endpointu jsou nahrána data o veřejnosti a neveřejnosti údajů
3. Pro veřejné údaje je indikováno, které datové sady z NKOD veřejný údaj obsahují
4. Pro údaje kódované číslníkem je indikováno, které datové sady z NKOD číslník obsahují
5. Pro číslníky je evidováno, kdo je poskytovatelem číslníku, tj. původcem dat, což je něco jiného, než kdo je poskytovatelem datové sady do NKOD, což je DIA.
6. Pro agendové údaje je indikováno, kdo je ohlašovatelem dané agendy



Klíčové funkcionality v prototypu NKOD

LinkedPipes ETL a pipeline tvořící NKOD



V rámci harvestace je nyní stahován seznam údajů evidovaných v RPP pomocí simulovaného SPARQL endpointu RPP spolu s informacemi, v jakých datových sadách v NKOD je daný veřejný údaj publikován, a jakým číselníkem v NKOD jsou kódovány údaje kódované číselníkem. Tato informace je pak použita v indexu pro Apache Solr, kam přibyly položky popisující datovou sadu:

```
"vdf_public_data" : "false",  
"vdf_codelist" : "false",
```

Informace o tom, kdo je původcem datové sady číselníku, je v NKOD reprezentována dle standardu DCAT 2 jako vlastnost datové sady `dcterms:creator`, například:

```
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix ovm: <https://rpp-opendata.egon.gov.cz/odrpp/zdroj/orgán-veřejné-moci/> .  
  
<https://data.gov.cz/zdroj/datové-sady/1>
```



```
dcterms:creator ovm:00025593 .
```

LinkedPipes DCAT-AP Viewer (uživatelské rozhraní NKOD)

Do konfigurace je nyní třeba přidat SPARQL endpoint RPP. Pokud nebude zadán, nové funkcionality se neprojeví.

Detail datové sady

Do detailu datové sady byly přidány funkcionality dle návrhu.

1. Odznaky indikující veřejné údaje a číselníky
2. Původce dat
3. Pro datovou sadu obsahující veřejné údaje z VDF seznam datových sad s číselníky
4. Pro datovou sadu obsahující číselník z VDF seznam datových sad obsahujících údaje kódované tímto číselníkem

Seznam datových sad

Do seznamu datových sad přibyla možnost filtrovat datové sady publikující veřejné údaje z VDF a číselníky VDF. Do Apache Solr bylo třeba přidat příslušné označení datové sady.

```
"vdf_public_data" : "false",  
"vdf_codelist" : "false",
```

Seznam poskytovatelů dat

V seznamu poskytovatelů dat nyní rozlišujeme původce a poskytovatele dat. Předpokládáme, že původci dat a číselníků ve VDF taktéž tato data publikují jako otevřená data, a tedy jsou v seznamu poskytovatelů otevřených dat. Pokud by tomu tak nebylo, jména původců dat, pokud to jsou OVM, lze také získat z RPP.

Přílohy

Použitý open-source software:

1. LinkedPipes ETL (fork v <https://github.com/datagov-cz/etl>)
 - o soubor `etl.zip`
2. LinkedPipes DCAT-AP Viewer (fork v <https://github.com/datagov-cz/dcat-ap-viewer>) - Zdrojový kód prototypu uživatelského rozhraní
 - o soubor `dcat-ap-viewer.zip`
3. Upravená pipeline LinkedPipes ETL pro harvestaci v rámci NKOD
 - o soubor `07 Harvestace LKOD a formulářů, aktualizace uživatelského rozhraní.jsonld`
4. Umělá testovací data simulující obsah RPP
 - o soubor `testovací data.ttl`
5. Umělá testovací data simulující označení původců dat
 - o soubor `test původce dat z VDF.ttl`