



# C3V6 Návrh využití ontologických konceptuálních modelů agend v NKOD

Vytvořeno v rámci projektu

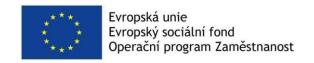
Rozvoj datových politik v oblasti zlepšování kvality a interoperability dat veřejné správy CZ.03.4.74/0.0/0.0/15 025/0013983

**Klíčová aktivita:** 04: Návrhy a realizace opatření pro zvyšování povědomí o otevřených datech

Indikátor: 8 05 00 Počet napsaných a zveřejněných analytických a strategických

dokumentů (vč. evaluačních)

Verze výstupu: 01





## Úvod

Národní katalog otevřených dat (NKOD) popisuje datové sady v podobě katalogizačních záznamů podle <u>otevřené formální normy pro rozhraní katalogů otevřených dat</u>. Katalogizační záznam datové sady v NKOD sestává z metadat popisujících datovou sadu z různých úhlů pohledu. Jedním z úhlů pohledu je tzv. sémantický (významový) úhel pohledu, který umožňuje systematickým způsobem popsat význam datové sady tak, že je datová sada označena (anotována) pojmem nebo pojmy z předpřipraveného slovníku pojmů. K tomu je možné využít dvě metadatové položky:

- <u>Téma</u> odkazuje na kategorii či téma datové sady, kdy alespoň jedno téma musí být z evropského číselníku datových témat
- Klasifikace dle EuroVoc odkazuje na kategorii či téma datové sady z evropského tezauru EuroVoc

Evropský číselník datových témat i evropský tezaurus EuroVoc jsou příklady předpřipravených slovníků pojmů. Každý pojem ve slovníku má svůj jasně definovaný a široce akceptovaný význam. Pojem použitý k označení datové sady tak jasně charakterizuje význam datové sady. Navíc dvě datové sady označené stejným tématem nebo stejnou položkou tezauru EuroVoc spolu věcně souvisejí, protože se týkají stejného tématu, tj. mají určitý společný význam. Těchto věcných souvislostí je možné využít při vyhledávání datových sad v NKOD.

Podporované řízené slovníky zmíněné výše jsou ale pouze základními slovníky. MV ČR pracuje na koncepci pokročilejšího řízeného slovníku, který nazývá <u>sémantický slovník</u> <u>pojmů veřejné správy ČR</u> (SSP), což je ontologický konceptuální model, jehož ambicí je popsat jednotlivé agendy veřejné správy ve strukturované a vzájemně propojené podobě.

Cílem tohoto dokumentu je popsat, jakým způsobem by měly být rozšířeny katalogizační záznamy o datových sadách v NKOD o možnost anotace významu datových sad pojmy z SSP a jak mohou být rozšířeny funkcionality NKOD vystavěnými nad takovým rozšířením.

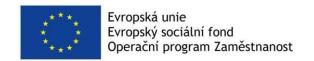
Tento dokument předpokládá znalost následujících dokumentů:

- Koncepce sémantického slovníku pojmů
- Metodika tvorby a údržby konceptuálních datových modelů agend

## Stávající řešení

## Katalogizační záznam

Dle <u>otevřené formální normy pro rozhraní katalogů otevřených dat</u> je téma datové sady z evropského číselníku datových schémat nebo z tezauru EuroVoc přiřazeno k datové sadě v jejím katalogizačním záznamu pomocí vlastnosti <u>dcat:theme</u> (katalogizační záznamy jsou vyjádřeny v modelu RDF, viz otevřená formální norma). Např. datová sada <u>Agendy</u> z registru práv a povinností je popsána následujícím způsobem:



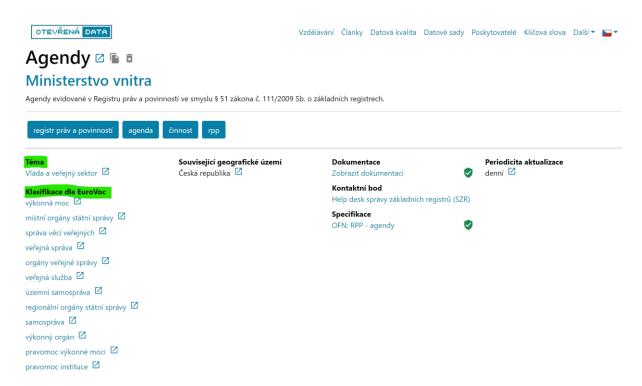


<a href="https://data.gov.cz/zdroj/datové-sady/00007064/9c73b802263c5e0ccf5542f10fbc35bb">https://data.gov.cz/zdroj/datové-sady/00007064/9c73b802263c5e0ccf5542f10fbc35bb</a> dcat:theme <a href="http://publications.europa.eu/resource/authority/data-theme/GOVE">http://eurovoc.europa.eu/2588</a>, <a href="http://eurovoc.europa.eu/39">http://eurovoc.europa.eu/39</a>.

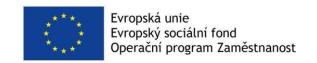
Prvním tématem je téma <u>Vláda a veřejný sektor</u> z evropského číselníku datových témat. Další dvě témata jsou témata <u>Orgány veřejné správy</u> a <u>Pravomoc výkonné moci</u> z tezauru EuroVoc. Jinými slovy témata říkají, že datová sada je o vládě a veřejném sektoru, o orgánech veřejné správy a pravomocech výkonné moci.

## Využití v NKOD

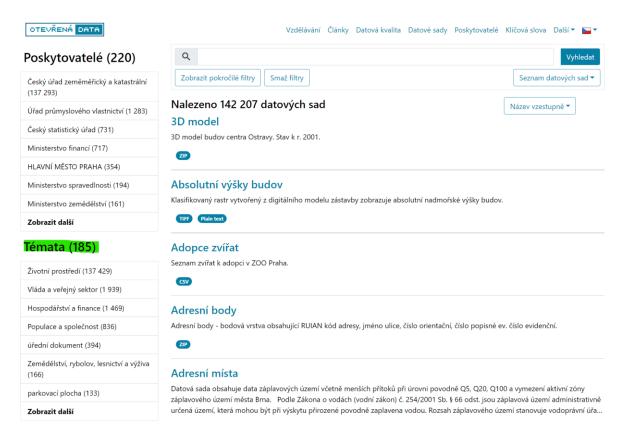
NKOD předpokládá tato témata zobrazuje v detailu datové sady, čímž potenciálnímu uživateli dat vysvětluje význam datové sady. Viz následující obrazovka z NKOD.



Dále NKOD předpokládá, že dvě datové sady označené stejným tématem jsou tematicky příbuzné a má proto smysl uživatelům umožnit podle témat vyhledávat. Filtrování datových sad dle témat (ať už témata dle evropského číselníku datových schémat tak témata dle tezauru EuroVoc) tak nabízí jako součást facetového vyhledávání. Viz následující obrazovka z NKOD.







### Návrh rozšíření o SSP

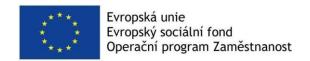
### Katalogizační záznam s pojmy ze SSP

Struktura katalogizačního záznamu dle <u>otevřené formální normy pro rozhraní katalogů otevřených dat</u> nebude rozšířena o nové položky. Bude využita existující položka <u>dcat:theme</u>, která je používána pro anotace pojmy z předpřipravených slovníků. SSP bude jedním z podporovaných slovníků. SSP sestává z pojmů, které reprezentují typy objektů, vlastností a vztahů, které popisují realitu domény, např. domény registru práv a povinností.

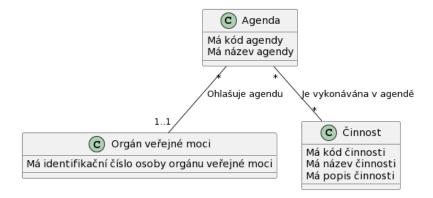
Pro popis datové sady jsou vybrány

- typy objektů, o nichž jsou v datové sadě údaje
- typy vlastností, které jsou v datové sadě o objektech vyjádřeny v podobě údajů
- typy vztahů mezi objekty, které jsou v datové sadě vyjádřeny v podobě údajů

Např. pro popis datové sady s agendami z registru práv a povinností mohou být vybrány pojmy ze SSP zobrazené na následujícím diagramu:







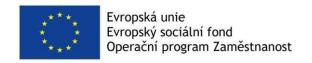
- Typy objektů Agenda, Orgán veřejné moci a Činnost vyjadřují, že datová sada obsahuje údaje o objektech, které jsou agendami, agendovými činnostmi a orgány veřejné moci.
- Typy vlastností Má kód agendy, Má název agendy, Má identifikační číslo osoby orgánu veřejné moci, Má kód činnosti, Má název činnosti a Má popis činnosti vyjadřují, pro jaké vlastnosti objektů z předchozího bodu lze v datové sadě získat údaje.
- Typy vztahů Ohlašuje agendu a je vykonávána v agendě vyjadřují, jaké vztahy mezi agendami, orgány veřejné moci a činnostmi jsou v datové sadě reprezentovány jako údaje.

Technicky je ve strojově čitelné podobě datová sada pojmy označena následujícím způsobem:

### Využití SSP v NKOD

SSP sestává ze dvou částí. První částí je glosář pojmů agend veřejné správy. Z tohoto pohledu je podobný jiným řízeným slovníkům, se kterými již NKOD pracuje, tj. výše zmíněný evropský číselník datových témat nebo tezaurus EuroVoc. V glosáři má každý pojem svůj preferovaný název, alternativní názvy, definici a popis.

Druhou částí je pak ontologický konceptuální model, který pojmy glosáře propojuje do vzájemných souvislostí následujícími vazbami:





- Specializace pojmů propojující speciální pojem s obecnějším pojmem
- Přiřazení typu vlastnosti k typu objektu
- Přiřazení účastnících se typů objektů k typu vztahu

Ontologický konceptuální model tak pro každý pojem definuje strukturovaným způsobem jeho významový kontext tvořený dalšími pojmy připojenými k pojmu vazbami.

- Kontextem typu vlastnosti je typ objektu, ke které je typ vlastnosti přiřazen.
- Kontextem typu vztahu jsou vztažené typy objektů.
- Kontextem typu objektu jsou k ní přiřazené typy vlastností a typy vztahů, kterých se typ objektů účastní a další typy objektů, které se účastní těchto typů vztahů.

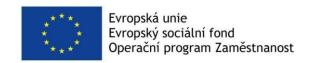
#### SSP je pak možné v NKOD využít 3 způsoby:

- 1. Zobrazení charakteristiky datové sady v podobě seznamu pojmů, kterými je datová sada označena.
- 2. Zobrazení detailu pojmu, kterými je datová sada označena.
- 3. Zobrazení konceptuálního modelu datové sady v podobě vizuálního diagramu.

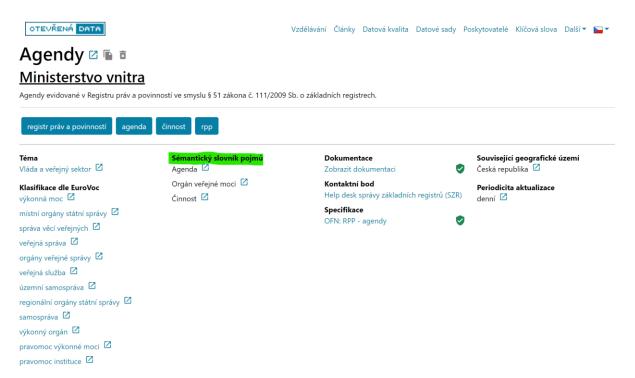
#### Zobrazení charakteristiky datové sady

Obrazovka s detailem datové sady, na které jsou zobrazovány pojmy z evropského číselníku datových témat a z tezauru EuroVoc, kterými je datová sada označena, bude rozšířena o sekci, ve které budou zobrazeny pojmy ze SSP, kterými je datová sada označena.

Takových pojmů ale může být poměrně hodně. Seznam pojmů nesmí být příliš velký, protože by byl pro uživatele nepřehledný. Proto budou zobrazeny pouze typy objektů, kterými je datová sada označena, a pro které navíc existuje alespoň jeden jim přiřazený typ vlastnosti, kterým je datová sada označena. Typy vlastností a typy vztahů nebudou zobrazeny.







Seznam typů objektů pro zobrazení v detailu datové sady lze získat následujícími SPARQL dotazy. První dotaz získá ze <u>SPARQL endpointu NKOD</u> seznam všech pojmů, kterými je datová sada označena. Za #datováSada musí být dosazeno IRI dané datové sady. (<u>příklad</u>)

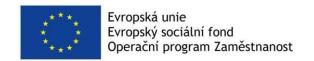
```
PREFIX dcat: <http://www.w3.org/ns/dcat#>

SELECT ?pojem
WHERE {
    #datováSada dcat:theme ?pojem .
}
```

Druhý dotaz pak vybere ze seznamu pojmů, které jsou výsledkem předchozího dotazu, ty pojmy, které mají být zobrazeny v detailu datové sady. Za #pojmy musí být dosazen seznam IRI pojmů, které jsou výsledkem předchozího dotazu, a tento dotaz musí být spuštěn na SPARQL endpointu SSP. (příklad)

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX z-sgov: <https://slovník.gov.cz/základní/pojem/>

SELECT ?typObjektu ?typObjektuLabel
WHERE {
   VALUES ?typObjektu {
        #pojmy
      }
      ?typObjektu a z-sgov:typ-objektu;
        skos:prefLabel ?typObjektuLabel .
```





```
FILTER EXISTS {
    ?typVlastnosti a z-sgov:typ-vlastnosti ;
    rdfs:subClassOf [
        a owl:Restriction ;
        owl:onProperty z-sgov:je-vlastností ;
        owl:someValuesFrom ?typObjektu
    ] .
}
```

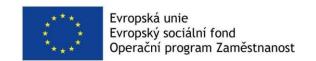
#### Zobrazení detailu pojmu

Zobrazení typů objektů v detailu datové sady poskytuje základní informaci o sémantice datové sady. NKOD umožní zobrazit detail každého jednotlivého zobrazeného pojmu tím, že pojmy ze SSP zobrazí jako odkazy do <u>prohlížečky SSP</u> ve tvaru

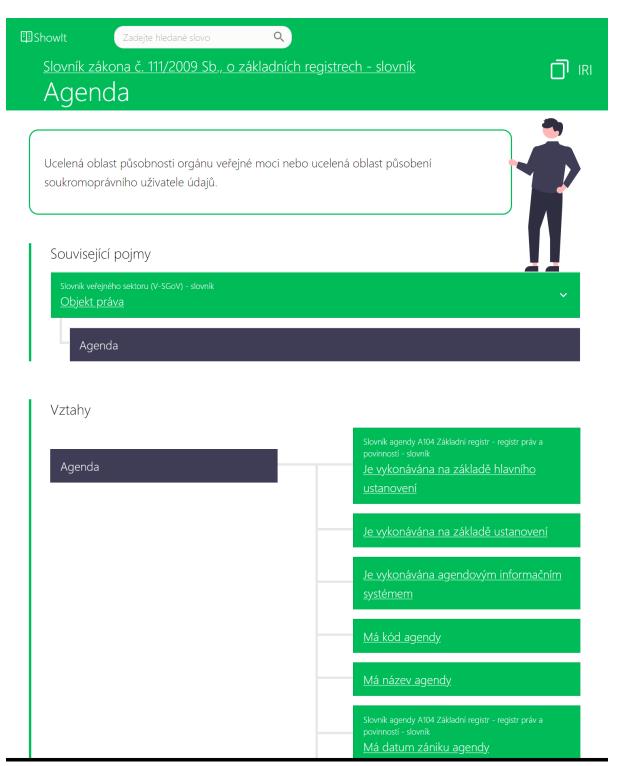
https://slovník.gov.cz/prohlížíme/pojem?iri=[*IRI pojmu*] (příklad)

Pro detailní specifikaci prohlížečky SSP viz příslušný <u>GitHub repozitář</u> se zdrojovými kódy a dokumentací.

Následující obrázek ukazuje výsledek zobrazení detailu pojmu agenda.

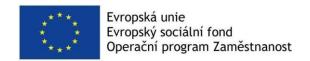






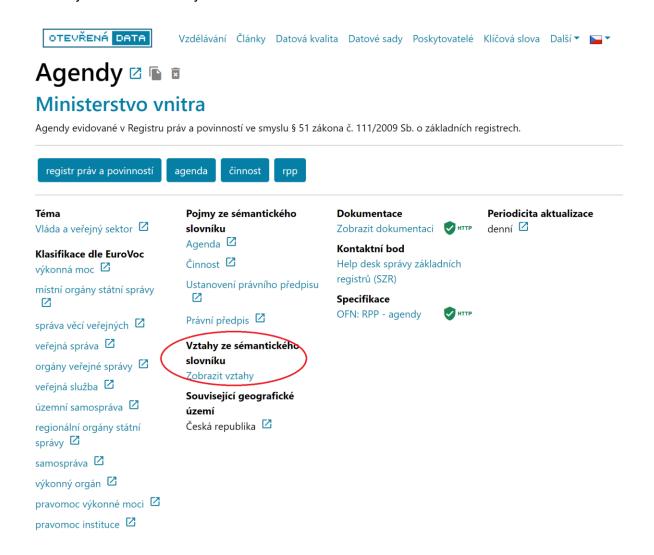
### Zobrazení konceptuálního modelu datové sady

Uživatel může také potřebovat zobrazit kompletní konceptuální model. K tomu NKOD nabídne pod seznamem pojmů ze SSP v detailu datové sady tlačítko, které povede na zobrazení diagramu kompletního konceptuálního modelu. Tento diagram musí být automaticky vygenerován na základě seznamu pojmů, kterými je datová sada označena. K vygenerování lze použít volně dostupná řešení, jako např. <a href="PlantUML">PlantUML</a>.

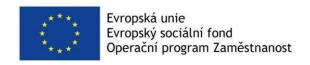




Následující obrázek ukazuje návrh umístění tlačítka.



Po kliknutí na tlačítko se v novém okně zobrazí kompletní ontologický konceptuální model vizualizovaný v podobě UML diagramu tříd podobnému tomu, uvedenému v příkladu výše. Doporučujeme jej maximálně zjednodušit, což by mělo určitě zahrnovat neuvedení kardinalit. Příklad zobrazení:



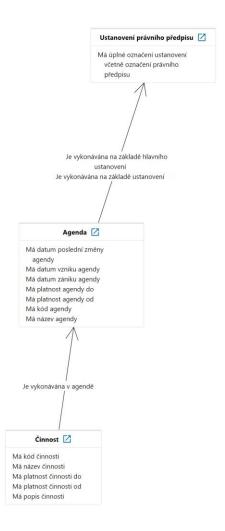


OTEVŘENÁ DATA

English

## Agendy 🛚

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

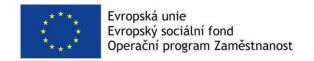


Pro konstrukci UML diagramu je nejprve nutno získat seznam pojmů, kterými je datová sada označena pomocí následujícího SPARQL dotazu do NKOD. Za #datováSada musí být dosazeno IRI dané datové sady. (příklad)

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>

SELECT ?pojem
WHERE {
    #datováSada dcat:theme ?pojem .
}
```

Dále je potřeba vybrat pouze ty pojmy, které jsou typy objektů, typy vlastností a typy vztahů pomocí následujícího dotazu do SPARQL endpointu SSP. Za #pojmy musí být dosazen seznam IRI pojmů, které jsou výsledkem předchozího dotazu. (příklad)



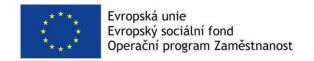


```
PREFIX skos: <a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <a href="http://www.w3.org/2002/07/owl">http://www.w3.org/2002/07/owl</a>
PREFIX z-sgov: <a href="mailto://slovník.gov.cz/základní/pojem/">
SELECT ?pojem ?typ
WHERE {
     VALUES ?pojem {
       #pojmy
     ?pojem a ?typ .
     FILTER (
        ?typ IN (
          z-sgov:typ-objektu,
          z-sgov:typ-vlastnosti,
          z-sgov:typ-vztahu
     )
}
```

Dotaz vrací ze všech pojmů, kterými je datová sada popsána pouze pojmy ze SSP, které jsou typy objektů, typy vlastností nebo typy vztahů. Pro každý je nutno získat detail pro diagram.

Detail pro typy objektů je získán následujícím SPARQL dotazem nad SPARQL endpointem SSP. Za #typy\_objektů musí být dosazen seznam IRI typů objektů, které jsou výsledkem předchozího dotazu. Dotaz vrátí pro každý dosazený typ objektu jeho název a seznam všech typů objektů, které specializuje, a jejich názvy. Nevrací specializované pojmy ze slovníku veřejného sektoru a ze základního slovníku. (příklad)

```
PREFIX z-sgov: <a href="mailto://slovník.gov.cz/základní/pojem/">https://slovník.gov.cz/základní/pojem/>
PREFIX skos: <a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <a href="http://www.w3.org/2000/01/rdf-schema">http://www.w3.org/2000/01/rdf-schema">http://www.w3.org/2000/01/rdf-schema</a>
CONSTRUCT {
  ?typObjektu a z-sgov:typ-objektu ;
     rdfs:subClassOf ?předek ;
     skos:prefLabel ?typObjektuLabel .
  ?předek a z-sgov:typ-objektu ;
     skos:prefLabel ?předekLabel .
} WHERE {
  VALUES ?typObjektu {
     #typy objektů
  ?typObjektu a z-sgov:typ-objektu ;
     skos:prefLabel ?typObjektuLabel .
  OPTIONAL {
     ?typObjektu rdfs:subClassOf ?předek .
     ?předek a z-sgov:typ-objektu ;
        skos:prefLabel ?předekLabel .
```

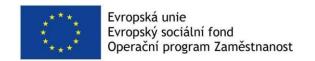




Detail pro typy vlastností je získán následujícím SPARQL dotazem nad SPARQL endpointem SSP. Za #typy\_vlastností musí být dosazen seznam IRI typů vlastností, které jsou výsledkem předchozího dotazu. Dotaz vrátí pro každý dosazený typ vlastnosti její název a typů objektů, kterému patří. Pro typ objektu také vrací jeho název. Pokud by jím nebyla datová sada označena, tj. byla by označena pouze typem vlastnosti, nebyl by název ve výsledku předchozích dotazů výše. (příklad)

```
PREFIX z-sgov: <a href="mailto://slovnik.gov.cz/základni/pojem/">https://slovnik.gov.cz/základni/pojem/>
PREFIX skos: <a href="http://www.w3.org/2004/02/skos/core">http://www.w3.org/2004/02/skos/core">
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
CONSTRUCT {
  ?typVlastnosti a z-sgov:typ-vlastnosti ;
    skos:prefLabel ?typVlastnostiLabel ;
    rdfs:domain ?typObjektu .
  ?typObjektu a z-sgov:typ-objektu;
    skos:prefLabel ?typObjektuLabel .
} WHERE {
  VALUES ?typVlastnosti {
    #typy_vlastností
  ?typVlastnosti a z-sgov:typ-vlastnosti ;
    skos:prefLabel ?typVlastnostiLabel .
    ?typVlastnosti rdfs:subClassOf [
      owl:allValuesFrom ?typObjektu ;
      owl:onProperty z-sgov:je-vlastností
    ?typObjektu skos:prefLabel ?typObjektuLabel .
  } UNION {
    ?typObjektu rdfs:subClassOf [
      owl:allValuesFrom ?tvpVlastnosti ;
      owl:onProperty z-sgov:má-vlastnost
    ?typObjektu skos:prefLabel ?typObjektuLabel .
```

A nakonec detail pro typy vztahů je získán následujícím SPARQL dotazem nad SPARQL endpointem SSP. Za #typy\_vztahů musí být dosazen seznam IRI typů vztahů, které jsou





výsledkem předchozího dotazu. Dotaz vrátí pro každý dosazený typ vztahu jeho název a typy objektů, které jsou účastníky vztahu. Pro účastníky vztahu také vrací jejich názvy. Pokud by jimi nebyla datová sada označena, tj. byla by např. označena pouze typem vztahu, nebyl by název ve výsledku předchozích dotazů výše. (příklad)

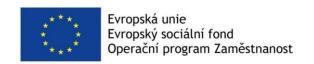
```
PREFIX z-sgov: <a href="mailto://slovník.gov.cz/základní/pojem/">https://slovník.gov.cz/základní/pojem/>
PREFIX skos: <a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <a href="http://www.w3.org/2000/01/rdf-schema">http://www.w3.org/2000/01/rdf-schema">
CONSTRUCT {
  ?typVztahu a z-sgov:typ-vztahu ;
    skos:prefLabel ?typVztahuLabel ;
    rdfs:domain ?zdrojovýTypObjektu ;
    rdfs:range ?cílovýTypObjektu .
  ?zdrojovýTypObjektu a z-sgov:typ-objektu ;
    skos:prefLabel ?zdrojovýTypObjektuLabel .
  ?cílovýTypObjektu a z-sgov:typ-objektu;
    skos:prefLabel ?cílovýTypObjektuLabel .
} WHERE {
  VALUES ?typVztahu {
    #typy_vztahů
  ?typVztahu a z-sgov:typ-vztahu ;
    skos:prefLabel ?typVztahuLabel ;
    rdfs:subClassOf [
       owl:allValuesFrom ?zdrojovýTvpObjektu;
       owl:onProperty z-sgov:má-vztažený-prvek-1
    rdfs:subClassOf [
       owl:allValuesFrom ?cílovýTypObjektu;
       owl:onProperty z-sgov:má-vztažený-prvek-2
  ?zdrojovýTypObjektu skos:prefLabel ?zdrojovýTypObjektuLabel .
  ?cílovýTypObjektu skos:prefLabel ?cílovýTypObjektuLabel .
```

## Testování použitelnosti

V této sekci se věnujeme testování použitelnosti prototypu pomocí standardní metodiky <u>System Usability Scale (SUS)</u>. Testovací uživatelé nejprve prototyp vyzkoušeli na základě scénáře, a pak vyplnili dotazník založený na SUS.

#### Scénář

 V uživatelském rozhraní prototypu přistupte k detailu datové sady Agendy od Ministerstva vnitra





- 2. V sekci "Pojmy ze sémantického slovníku" vidíte pojmy, kterými je tato datová sada anotována.
- 3. Kliknutím na pojem "Právní předpis" uvidíte ostatní datové sady v NKOD anotované tímto pojmem. Vraťte se zpět na datovou sadu Agendy.
- 4. Kliknutím na šipku za pojmem "Právní předpis" nebo jiným uvidíte jeho definici přímo v Sémantickém slovníku pojmů v nástroji ShowIT (ten není předmětem tohoto testování).
- 5. Kliknutím na odkaz "Zobrazit vztahy" v sekci "Vztahy ze sémantického slovníku" zobrazíte vztahy mezi pojmy, kterými je datová sada anotována.
- 6. Kliknutím na šipku za názvem objektu, vztahu nebo atributu, např. vztah "Je vykonávána na základě hlavního ustanovení", uvidíte jeho definici v Sémantickém slovníku pojmů v nástroji ShowIT (ten není předmětem tohoto testování).

### Výsledky dle metodiky SUS

Počet testovacích uživatelů: 3

Výsledné skóre a jeho interpretace: 85.83 - nadprůměrné skóre

#### Jednotlivé výsledky

Už./O.	01	O2	О3	O4	O5	<b>O</b> 6	07	08	О9	O10	Skóre
1	3	4	4	3	4	4	4	4	4	3	92.5
2	3	3	4	4	3	2	3	3	3	4	80
3	3	4	4	3	4	4	3	4	3	2	85

## Přílohy

- 1. Naplnění SKOD.jsonld LinkedPipes ETL pipeline pro naplnění testovacích pojmů dp triplestore prototypu.
- 2. dcat-ap-viewer-develop.zip Rozšířený nástroj LinkedPipes DCAT-AP Viewer o prototyp zobrazování pojmů ze sémantického slovníku pojmů
- 3. <a href="nkod-graph-browser-master.zip">nkod-graph-browser-master.zip</a> Prototyp zobrazování vztahů mezi pojmy ze SSP