



# Návrh a prototypování výrobní linky pro tvorbu a údržbu konceptuálních modelů agend

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

**KA 6:** Návrhy a realizace prostředí pro ontologické konceptuální datové modelování

**Indikátor:** 8 05 00 Počet napsaných a zveřejněných analytických a strategických dokumentů (vč. evaluačních)

**Verze výstupu:** 01



# 1. Definice cíle

Typ výstupu: Dokument s návrhem výrobní linky, dokument s dokumentací vytvořeného prototypu, zdrojové kódy vytvořeného prototypu vydané též jako open source (otevřený zdrojový kód), ontologické konceptuální modely agend v podobě otevřených dat vytvořené během testování prototypu.

Bude navržena výrobní linka ontologických konceptuálních datových modelů pro zvyšování sémantické interoperability informačních systémů veřejné správy. Výrobní linka bude navržena jako kompozice existujících softwarových nástrojů (open source) a bude realizována v podobě softwarového prototypu. Prototyp bude průběžně testován. Prototyp je zásadní pro komunikaci s cílovou skupinou.

## 2. Úvod

Výrobní linkou konceptuálních modelů (nebo jen výrobní linkou) rozumíme sadu nástrojů, které lze použít k tvorbě konceptuálních modelů např. legislativy, agend, datových sad, dle architektury Sémantického slovníku pojmů viz <https://slovník.gov.cz>.

Jednotlivé části výstupu jsou strukturovány takto:

- **Dokument s návrhem výrobní linky** tvoří kapitoly 1-4 tohoto dokumentu.
- **Zdrojové kódy vytvořeného prototypu**, vč. technické dokumentace jsou k dispozici v několika GitHub repozitářích.
  - *Kontrolní panel* (nástroj pro správu uživatelů a pracovních prostorů)
    - <https://github.com/opendata-mvcr/mission-control>
  - *SGoV server* (backend nástroje kontrolní panel)
    - <https://github.com/opendata-mvcr/sgov>
  - *SGoV validátor* (služba pro kontrolu integrity slovníků)
    - <https://github.com/opendata-mvcr/sgov-validator>
  - *Autentizační služba* (služba používaná nástroji výrobní linky pro přihlašování)
    - <https://github.com/opendata-mvcr/keycloak-graphdb-user-replicator>
  - *Editor glosáře* (a.k.a. Termlt; správce odborného pojmosloví a nástroj pro jeho propojení na legislativu) - projekt je rozšířením existujícího open-source nástroje Termlt
    - back-end: <https://github.com/opendata-mvcr/termit>
    - front-end: <https://github.com/opendata-mvcr/termit-ui>
  - *Editor modelu* (a.k.a. OntoGrapher; Správce konceptuálních modelů využívajících vytvořené pojmosloví)
    - <https://github.com/opendata-mvcr/ontoGrapher>
  - repozitář sdílených komponent
    - <https://github.com/opendata-mvcr/assembly-line-shared>
- **Dokument s dokumentací vytvořeného prototypu** tvoří kapitola 5 tohoto dokumentu.
- **Scénáře uživatelského testování výrobní linky a jeho výsledky** tvoří kapitola 6 tohoto dokumentu.

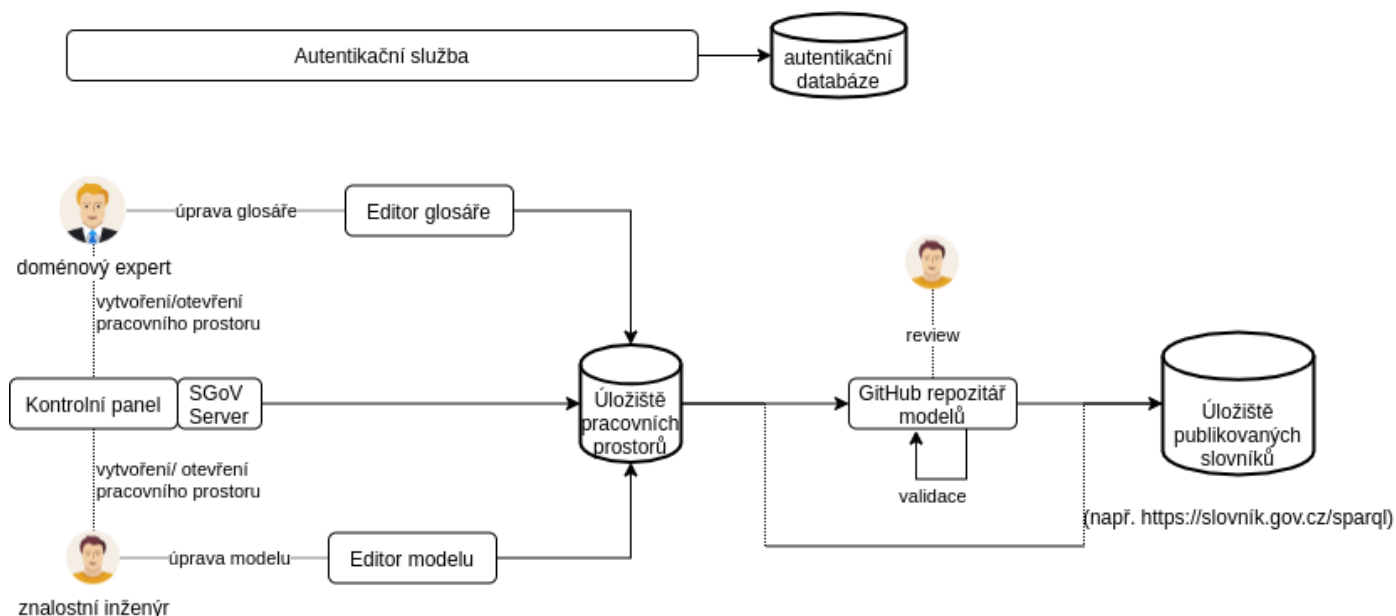
- **Ontologické konceptuální modely agend v podobě otevřených dat vytvořené během testování prototypu**
  - *GitHub repozitář modelů* (nástroj pro revizi a publikaci konceptuálních modelů)
    - <https://github.com/opendata-mvcr/ssp>

Výrobní linka je *rozšiřitelná* o další nástroje, které mohou plnit další úlohy při vzniku konceptuálního modelu (např. alternativní editory glosářů, systémy pro podporu revize a workflows konceptuálních modelů, apod).

Předpokládanými uživateli výrobní linky jsou orgány veřejné správy při ohlašování agend, jejich aktualizaci, tvorbu datových sad, návrh informačních systémů. Tvorby konceptuálních modelů pomocí výrobní linky by se měli účastnit doménoví experti na danou agendu a experti na datové konceptuální modelování.

### 3. High-level architektura výrobní linky

Základní komponenty a workflow výrobní linky lze vysledovat na tomto diagramu:



Základní transakční jednotkou práce výrobní linky jsou **pracovní prostory**. Jsou reprezentovány několika pojmenovanými RDF<sup>1</sup> grafy, které obsahují změny upravovaných slovníků, nebo návrhy nových slovníků, které uživatel výrobní linkou edituje. Nástroj výrobní linky tedy může z pracovního prostoru číst / do pracovního prostoru ukládat:

<sup>1</sup> <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>



- aktuální stav slovníku (snapshot), ve formátu SKOS<sup>2</sup> a OWL<sup>3</sup>,
- změny,
- libovolná další metadata nutná k práci nástroje.

### 3.1. Úložiště pracovních prostorů

Současná verze výrobní linky umožňuje vytvářet všem přihlášeným uživatelům libovolné pracovní prostory, editovat je a mazat je. Aby bylo možné zajistit, že data vytvářená výrobní linkou budou co nejméně konfliktní, každý slovník je editovatelný maximálně v jednom pracovním prostoru. Ačkoliv je možné v jednom pracovním prostoru najednou upravovat více slovníků, každý pracovní prostor by měl sloužit pro samostatný účel. Pokud tedy uživatel chce měnit dva nezávislé slovníky, měl by si k tomu vytvořit dva pracovní prostory. Naopak pokud chce uživatel měnit více slovníků různých úrovní pokrývajících jeden celek (legislativní slovník, agendový slovník, datový slovník), je vhodné k tomu použít jediný pracovní prostor.

Některé slovníky vysoko v hierarchii, zejména V-SGoV, budou mít samostatné pracovní prostory, aby jejich úprava byla pod maximální kontrolou. Tyto slovníky bude možné do ostatních pracovních prostorů v souladu s tezí výše importovat pouze jako read-only.

#### 3.1.1. Struktura pracovního prostoru

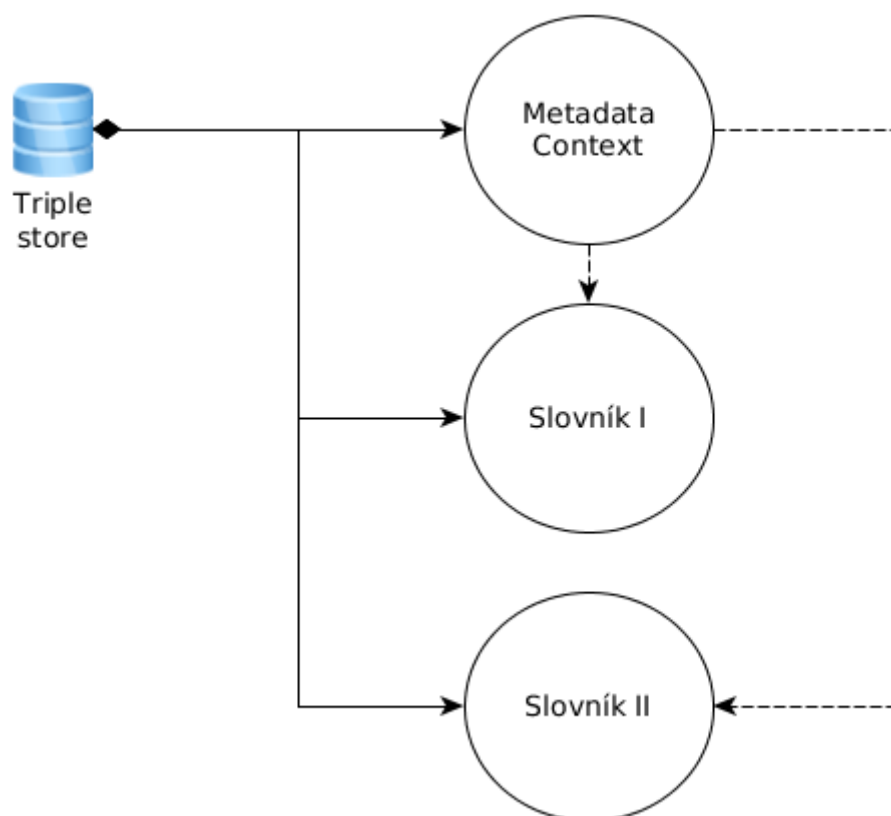
**Pracovním prostorem** se rozumí RDF dataset obsahující alespoň 2 pojmenované grafy (kontexty), kde jeden z kontextů je *metadata kontext* a ostatní kontexty obsahují buď slovníky či metadata týkající se slovníků. Pracovní prostor je identifikován IRI, která je shodná s IRI metadata kontextu. Pracovní prostor je uložen v triple store se SPARQL<sup>4</sup> rozhraním.

---

<sup>2</sup> <https://www.w3.org/2009/08/skos-reference/skos.html#>

<sup>3</sup> <https://www.w3.org/TR/owl2-overview/>

<sup>4</sup> <https://www.w3.org/TR/sparql11-query/>



**Metadata kontext** - pojmenovaný RDF graf (kontext v terminologii triple store) obsahuje data odkazující na kontexty slovníků dostupných v rámci tohoto pracovního prostoru. Stačí IRI příslušných kontextů, neboť požadujeme, že se nachází ve stejném triple store jako metadata kontext.

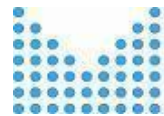
Metadata kontext zároveň slouží jako vyhledávací mapa - odkazuje-li slovník na jiné slovníky pomocí importů, metadata kontext obsahuje informaci o tom, v jakém kontextu v tomto pracovním prostoru je odkazovaný slovník dostupný (daný slovník může být v rámci pracovního prostoru pouze ke čtení). Slovník může být v pracovním prostoru maximálně jednou. Jelikož se ale slovníky mohou vyskytovat v rámci různých pracovních prostorů ve stejném úložišti víckrát, je identifikátor kontextu slovníku v rámci pracovního prostoru parametrizován i identifikátorem pracovního prostoru, aby se zajistila jeho unikátnost.

Nástroje výrobní linky dostanou při importu slovníků pro editaci URL příslušného triple store a IRI pracovního prostoru (tj. metadata kontextu). Z něj získají všechna potřebná data pro práci se slovníky.

### 3.2. Správa pracovních prostorů

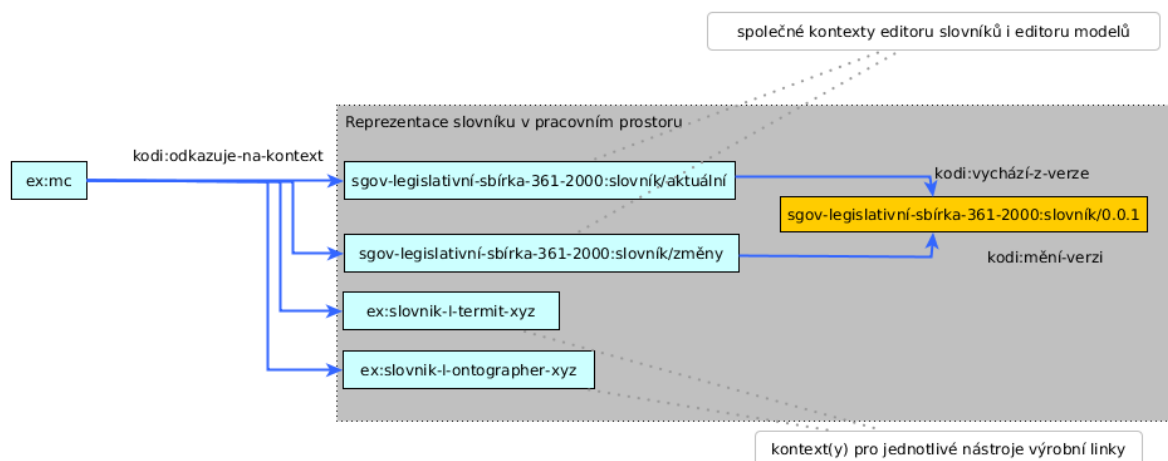
Pracovní prostory vytváří a spravuje kontrolní panel výrobní linky. Stejně tak export do primárního úložiště (a případně na GitHub) provádí kontrolní panel.

Nástroje výrobní linky (správce slovníků, online modelovací nástroj) umožňují uživateli zadat URL úložiště pracovního prostoru a IRI metadata kontextu. Na základě těchto dat jsou nástroje schopny natáhnout si nutná data. Předpokladem je, že nástroje mohou měnit data přímo v pracovním kontextu a nemusí si vytvořit lokální kopii. Samozřejmě si kopii vytvářet mohou, v takovém případě ale jsou zodpovědné za zpětnou synchronizaci do pracovního kontextu.



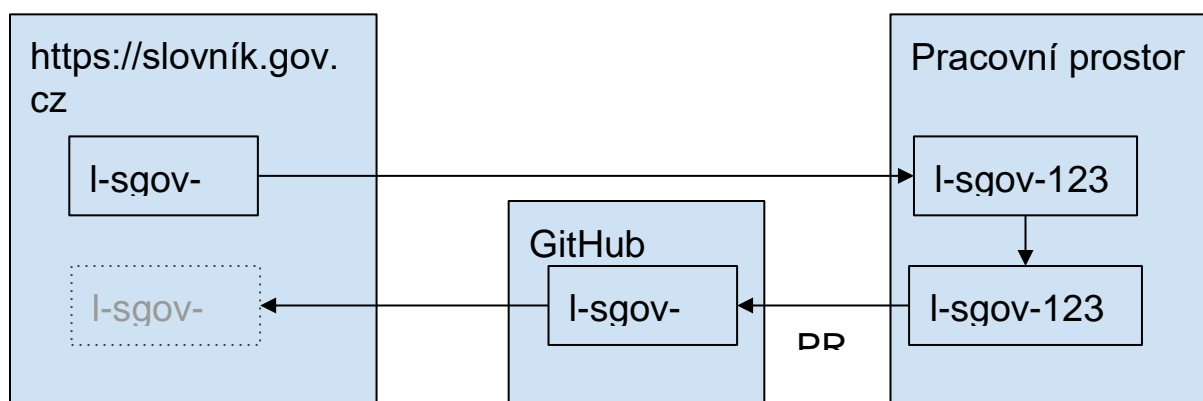
Offline nástroje musí umožňovat import RDF exportu pracovního kontextu a naopak export editovaných dat do RDF kompatibilního s obsahem v pracovním kontextu.

Příklad reprezentace slovníku v pracovním prostoru je zobrazen zde (návrh na mírný update reprezentace vzhledem k odkazu výše).



Každý pracovní prostor obsahuje data k zobrazování a editaci slovníků pro účely jejich změny pomocí editorů glosářů a editorů modelů. Např. pracovní prostor `ex:mc` pro editaci slovníku zákon 361-2000 obsahuje tyto kontexty:

- **sgov-legislativní-sbírka-361-2000:slovník/aktuální** obsahuje aktuální snapshot slovníku po editaci některými z nástrojů výrobní linky
  - tento kontext (nebo dva kontexty pro glosář/model) obsahuje glosář ve formátu SKOS a model ve formátu OWL
- **sgov-legislativní-sbírka-361-2000:slovník/změny**
  - tento kontext (nebo dva kontexty pro glosář/model) obsahuje změnové vektory oproti původní verzi slovníku (`sgov-legislativní-sbírka-361-2000:slovník/0.0.1` na obrázku).
- další kontexty, kde si jednotlivé nástroje drží libovolné další informace (např. layout)
- Cesta "mc" v daném příkladu reprezentuje parametrizaci identifikátoru kontextu slovníku identifikátorem pracovního prostoru jako takového, aby se zajistila unikátnost identifikátorů v případě, že více pracovních prostorů importuje ten samý slovník.





### 3.2.1. Validace

Konzistenci slovníků v pracovním prostoru budou nástroje výrobní linky zajišťovat pomocí REST služby validující aktuální verzi slovníků v pracovním prostoru. Služba spouští sadu předdefinovaných pravidel, které definuje komponenta **SGoV validátor**. Stejná validační pravidla budou rovněž řídit kvalitu uvnitř Github repozitáře, do kterého mohou přicházet Pull requesty i mimo výrobní linku. Slovník se skládá z **glosáře** a **modelu**. Validace probíhá nad glosářem, nad modelem i nad celým slovníkem.

### 3.2.2. Struktura GitHub repozitáře

Struktura GitHub repozitáře SSP vychází z toho, že jeden slovník může upravovat více uživatelů.

1. Každý slovník je reprezentován v repozitáři jediným adresářem:

z-sgov

v-sgov

o-sgov-<DOMENA>

<DOMENA> je doména daného obecného slovníku, např. *silniční-doprava*

l-sgov-sbírka-<C>-<R>

<C> je částka, např. 111

<R> je rok, např. 2009

a-sgov-<A>

<A> je číslo agendy, např. 101

d-sgov-<D>

<D> je identifikátor datové sady/datového rozhraní/aplikace

2. V každém adresáři je jeden nebo více souborů ve tvaru:

<SLOVNIK>-glosář(-<SPECIFICKY-IDENTIFIKATOR>).ttl

<SLOVNIK>-model(-<SPECIFICKY-IDENTIFIKATOR>).ttl

<SLOVNIK>-slovník(-<SPECIFICKY-IDENTIFIKATOR>).ttl

kde <SLOVNIK> je identifikátor slovníku, např. l-sgov-sbírka-111-2009 a (-<SPECIFICKY-IDENTIFIKATOR>) je libovolný identifikátor, kterým rozlišujeme jednotlivé části slovníku (např. podle uživatelů, kteří tyto části tvoří).

3. Volitelně adresář rovněž obsahuje soubory tvaru

<SLOVNIK>-mapování(-<SPECIFICKY-IDENTIFIKATOR>).ttl

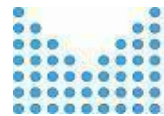
<SLOVNIK>-diagram(-<SPECIFICKY-IDENTIFIKATOR>).ttl

4. V každém adresáři obsahují všechny soubory stejného typu (např. glosář) odkaz na stejný RDF zdroj, např.

l-sgov-sbírka-111-2009-glosář-MaNe.ttl

l-sgov-sbírka-111-2009-glosář-PeKr.ttl





budou oba referencovat stejný glosář v RDF:

<https://slovník.gov.cz/legislativní/sbírka/111/2009/glosář>

## 4. Nasazení výrobní linky

Výrobní linka je plně dockerizovaná - její nasazení je tedy možné jak na bare-metal serverech, tak prostřednictvím cloudu. Aktuální instrukce k sestavení docker obrazů, vč. orchestračního skriptu je k dispozici na:

<https://github.com/opendata-mvcr/sgov-assembly-line/commit/d9eba251ad03ff9dcf7704ac0ea759d0b8b53968>

## 5. Dokumentace vytvořeného prototypu

Tato kapitola popisuje základní scénář práce s vytvořeným prototypem. Výrobní linka je nasazena (výstup C5V7) ve dvou prostředích:

- vývojové - <https://xn--slovnk-test-scb.mvcr.gov.cz/modelujeme> - pro testování nových funkcí
- produkční - <https://slovník.gov.cz/modelujeme> - pro vývoj modelů

Přístup do výrobní linky je zájemcům o testování na požádání vytvořen.

### 1. Přihlášení uživatele do výrobní linky

### 2. Vytvoření pracovního prostoru

Po přihlášení se uživatel dostane do nástroje Kontrolní panel, ve kterém se mu





zobrazí seznam aktuálních pracovních prostorů, které dříve vytvořil:

Kontrolní panel

NAPRAVIT CHYBU

ČESKY

### Moje pracovní prostory

+ VYTVOŘIT PRACOVNÍ PROSTOR

Název	Poslední změna	Datum poslední změny	Akce
čas	Petr Křemen	26. dubna 2021 v 17:28	<button>EDITOVAT POJMY</button> <button>EDITOVAT VZTAHY</button>
eSbírka v AJ	Petr Křemen	24. května 2021 v 17:24	<button>EDITOVAT POJMY</button> <button>EDITOVAT VZTAHY</button>

10 rows |< 1 of 2 >|

[Zobrazit všechny pracovní prostory](#)

Tato aplikace je udržována v rámci projektu  
OPZ č. CZ.03.4.74/0.0/0.0/15.025/0013983.

Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

Tlačítkem v pravém horním rohu seznamu pracovních prostorů je možné vytvořit nový pracovní prostor s daným názvem:

### Moje pracovní prostory

+ VYTVOŘIT PRACOVNÍ PROSTOR

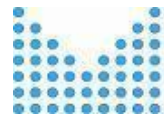
Název	Poslední změna	Datum poslední změny	Akce
čas	Petr Křemen	26. dubna 2021 v 17:28	<button>EDITOVAT POJMY</button> <button>EDITOVAT VZTAHY</button>
eSbírka v AJ	Petr Křemen	24. května 2021 v 17:24	<button>EDITOVAT POJMY</button> <button>EDITOVAT VZTAHY</button>

Vytvořit pracovní prostor

Název

VYTVOŘIT PRACOVNÍ PROSTOR

Detail pracovního prostoru zobrazuje kromě základních informací a editačních metadat rovněž seznam slovníků. Ten je na počátku prázdný.



Kontrolní panel NAHLÁSIT CHYBU ČESKY

[-- Pracovní prostory](#)

Pracovní prostor zákona JVF DTM EDITOVAT POJMY EDITOVAT VZTAHY PUBLIKOVAT POKROČILÉ

Základní informace

Název	Pracovní prostor zákona JVF DTM
Vlastník	<span>PK</span> Petr Křemen
Poslední změna	<span>PK</span> Petr Křemen
Datum poslední změny	26. května 2021 v 17:35

Slovníky + PŘIDAT SLOVNÍK

Název	Akce
Nic nenalezeno	

Závislosti mezi slovníky

Uživatel si může do pracovního prostoru přidat slovníky existující, které tak bude moci upravovat, nebo vytvořit slovník nový.

[-- Pracovní prostory](#)

Pracovní prostor zákona JVF DTM EDITOVAT POJMY EDITOVAT VZTAHY PUBLIKOVAT POKROČILÉ

Základní informace

Název	Pracovní prostor zákona JVF DTM
Vlastník	<span>PK</span> Petr Křemen
Poslední změna	<span>PK</span> Petr Křemen
Datum poslední změny	26. května 2021 v 17:35

Slovníky + PŘIDAT SLOVNÍK

Název	Akce
Slovník vyhlášky č. 393/2020 Sb. <a href="https://slovník.gov.cz/legislativni/sbirka/393/2020">https://slovník.gov.cz/legislativni/sbirka/393/2020</a>	<span>AKTUALIZOVAT</span> <span>SMAZAT</span>

### 3. Vytvoření nového pojmu

Po vytvoření nového pracovního prostoru (nebo otevření stávajícího) je možné vytvářet glosář/tezaurus a to i na základě existujícího dokumentu (např. legislativy). Tlačítkem "Editovat pojmy" je možné přejít do Editoru glosářů (systém TermIt):



**TermIt** ☰  Pracovní prostor: Pracovní prostor zákona JVF DTM Nahlásit chybu 👤

Hlavní strana / Slovníky

### Správa slovníků

NÁZEV IS

Filtrovat 1 záznam...

Slovník vyhlášky č. 393/2020 Sb.

Zobrazit 10 záznamů na stránku

Čestina ▼ © KBSS FEL ČVUT v Praze, MVČR, 2021 Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost ⚡ Verze 2.4.1

Po otevření slovníku je možné tlačítkem “Nový pojem” vytvořit nový pojem.

Vytvořit nový pojem

čestina 🌐

Název 📎

Povinné

Synonyma 📎

+ Přidat

• nádrž 📌 🔍

Definice

Text 📎

Zdroj 📎

Dopřítující poznámka 📎

Nadřazené pojmy 📎

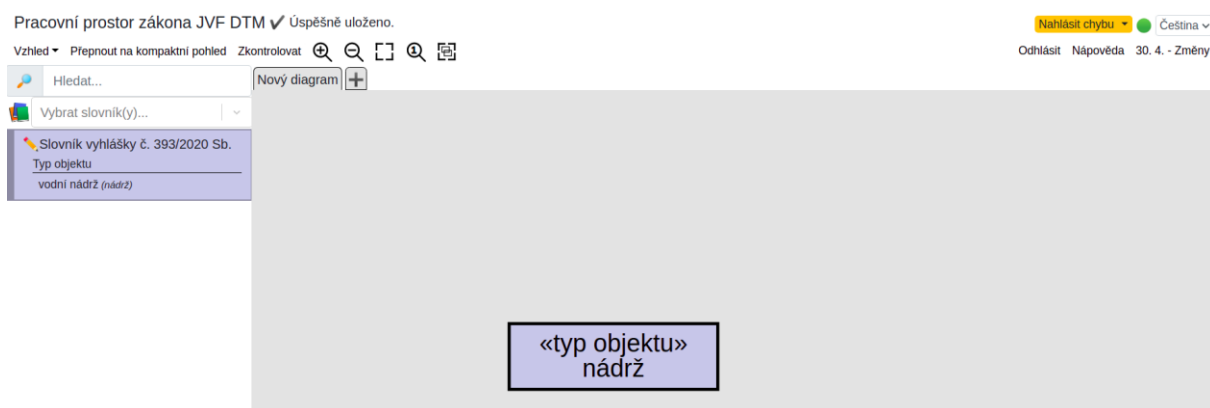
[Zobrazit pokročilé možnosti](#)

Vytvořit Vytvořit a začít nový Zrušit

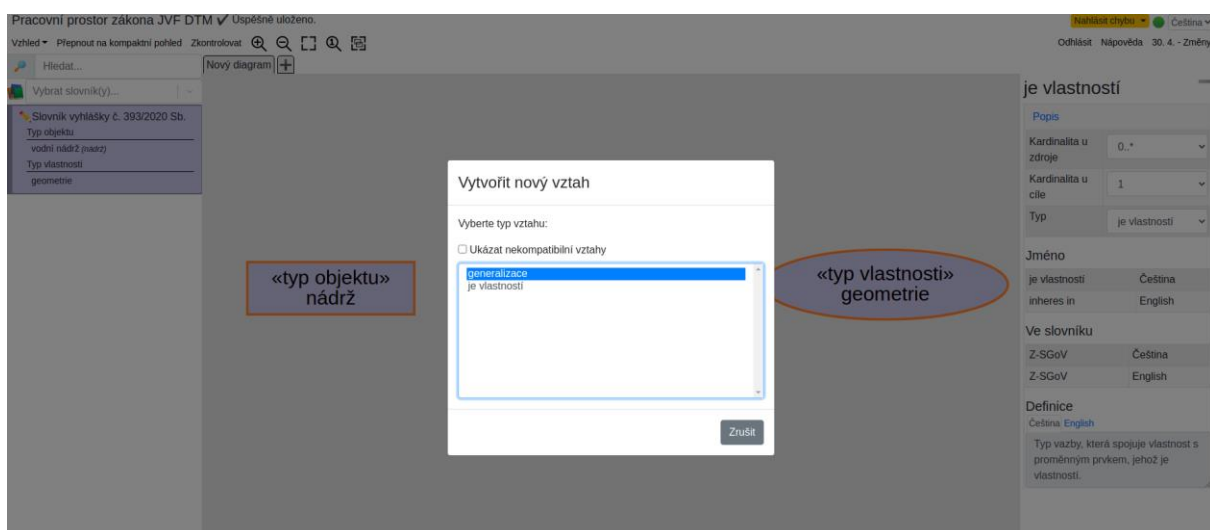
a potvrdit tlačítkem “Vytvořit”.

#### 4. Vytvoření ontologické vazby

Po dokončení práce se editorem glosáře je možné vytvářet ontologické vazby v editoru modelů (systém OntoGrapher), který lze otevřít tlačítkem “Editovat vztahy” v kontrolním panelu. Po otevření je možné ze seznamu pojmů slovníku v levém panelu přetáhnout na plátno pojmy, které chce uživatel propojovat.

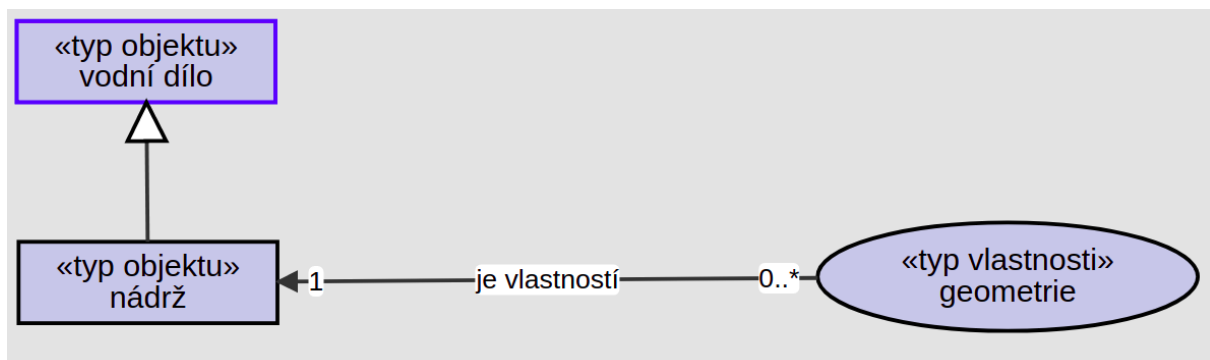


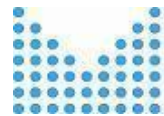
Pokud některé pojmy modeláři chybí, může je vytvořit pomocí kontextového menu. Při selekci některého uzlu v diagramu se otevře pravý panel s detaily vybraného pojmu. Zde je možné vybírat z předdefinovaných typů pojmů (zde v příkladu “Typ objektu” a “Typ vlastnosti”). Typy vlastností je pak možné propojovat s typy objektů pomocí vazby “je vlastností”:



Výsledný diagram pak modeluje fakt, že nádrž má geometrii jako svoji vlastnost. Analogicky je možné vytvořit i specializační vazbu a modelovat např. že vodní nádrž je speciální případ vodního díla:

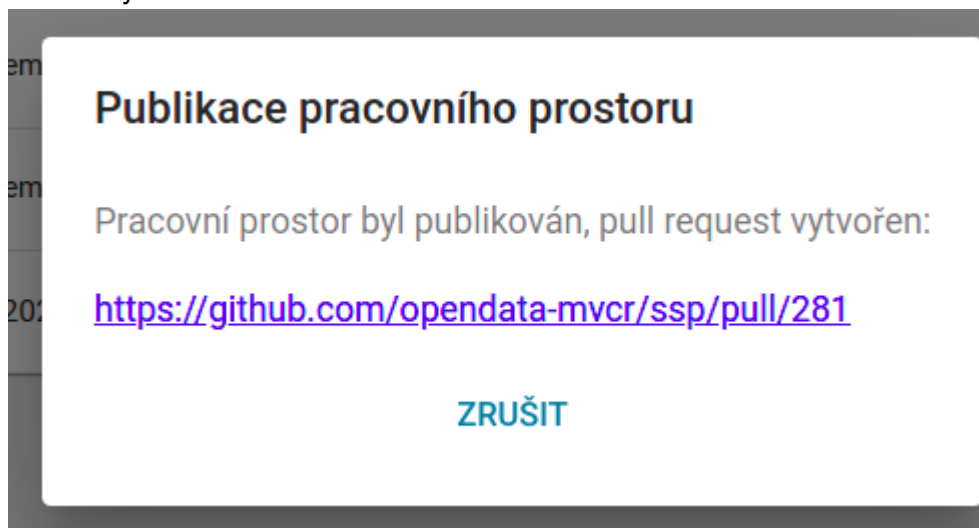
Výsledný model pak má podobu:





## 5. Publikace slovníku k revizi

Po dokončení editace v pracovním prostoru je možné změny v pracovním prostoru publikovat v kontrolním panelu pomocí tlačítka “Publikovat” v detailu pracovního prostoru. Samotná publikace trvá několik desítek sekund a poté je vytvořena žádost o revizi v systému Github:



## 6. Revize a publikace slovníku

Správci repozitáře <https://github.com/opendata-mvcr/ssp> revidují změny ve slovníku pomocí prostředků systému GitHub a po potvrzení změn jsou změny publikovány do veřejného SPARQL endpointu.

# 6. Testování výrobní linky

K otestování výrobní linky byla použita metoda System usability scale (SUS). Ta je založena na uživatelském hodnocení produktu odpovědí na deset otázek ve škále 1 - 5 (silně nesouhlasím, spíše nesouhlasím, neutrální stanovisko, spíše souhlasím, silně souhlasím) a následném výpočtu celkového skóre produktu. Systém nehodnotí to, jak uživatel v testovaném produktu něco udělal, ale jak se mu s produktem pracovalo. Metodika testu je popsána zde.

## 6.1 Testovací scénáře výrobní linky

V první fázi je definováno pět uživatelských scénářů, které by měly otestovat všechny části výrobní linky. Testovací scénáře jsou popsány v příloze C5V3a, obsahují podrobné zadání scénářů, náповědu k jejich vypracování a ukázkové řešení. Jedná se o tyto scénáře:

- 1) Import existujícího slovníku do výrobní linky a doplnění pojmů,
- 2) Doplnění existujícího modelu slovníku, propojení objektů a vlastností v rámci jednoho slovníku,
- 3) Propojení pojmů ze slovníku s pojmy z jiných slovníků,



- 4) Vytvoření slovníku a glosáře,
- 5) Odvozování a interpretace závislostí mezi pojmy, pojmy se stejným nebo podobným významem.

## 6.2 Vyhodnocení SUS

Po absolvování testovacích scénářů uživatelé odpověděli na otázky SUS a bylo vypočteno celkové skóre. Testování se zúčastnilo celkem 13 uživatelů, výsledné skóre SUS pro testování bylo 67,5. Průměrné hodnocení produktů dle SUS je 68, ale výrazně se liší podle zaměření aplikace a podle očekávané uživatelské skupiny. Pro zaměření aplikace a cílovou skupinu uživatelů je hodnocení 68 dobré.