



Metodika tvorby SK GOV URI identifikátorov

Sémantické dátové štandardy SR



data.gov.sk-semanticweb-0.9.3

Päťhviezdičkové otvorené údaje v SR, 24.7.2015

Miroslav Líška, Datalan

Obsah

I. Úvod

1. Sémantické štandardy

II. Bezproblémové časti

2. Ontológie

3. Individuá

4. Katalógy/Datasety

III. Nové návrhy na schválenie

5. Spôsob verzionovania (distribúcie dát)

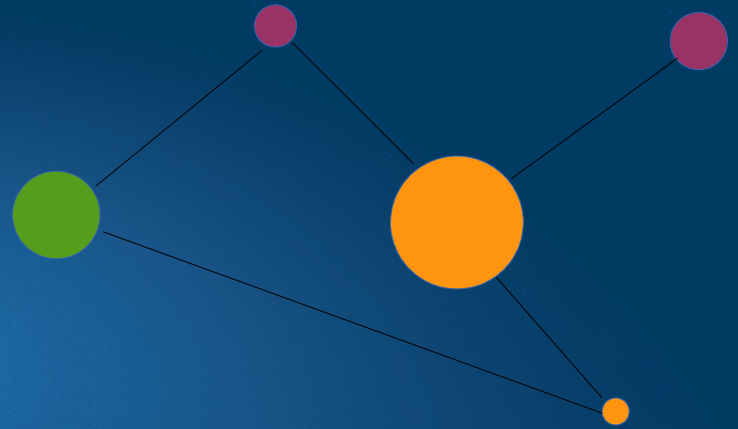
6. Prenos číselníkov do sémantiky

7. Eformulár + XMLDataContainer

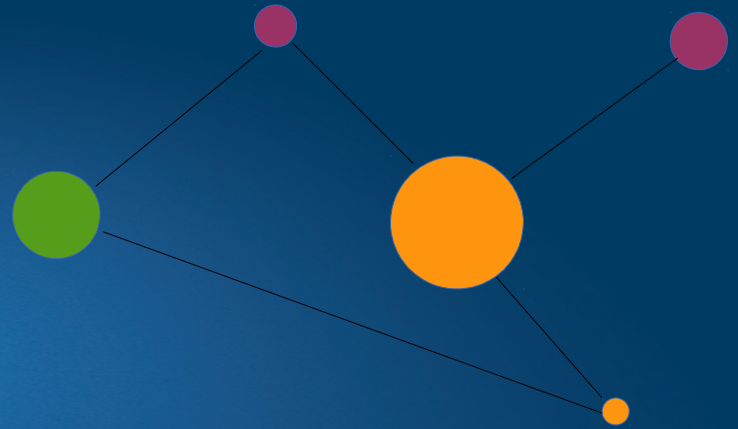
IV. Problémové časti

8. Všeobecný formát URI pre verejné dáta





I. Úvod



II. Bezproblémové časti

2) SK GOV URIs - Ontológie

- SEMIC (EÚ)



- ODP (SK)

Ontológia Dátových Prvkov

<http://data.gov.sk/def/ontology/odp/>

Trieda

[http://data.gov.sk/def/ontology/odp/\[Class\]](http://data.gov.sk/def/ontology/odp/[Class])

Objektová vlastnosť

[http://data.gov.sk/def/ontology/odp/\[objectProperty\]](http://data.gov.sk/def/ontology/odp/[objectProperty])

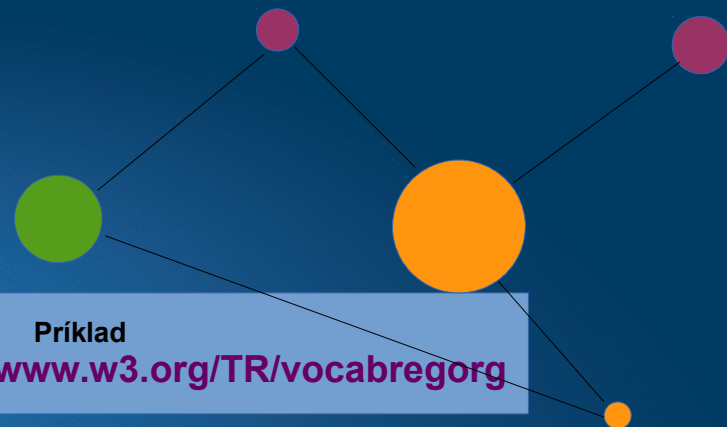
Dátová vlastnosť

[http://data.gov.sk/def/ontology/odp/\[dataTypeProperty\]](http://data.gov.sk/def/ontology/odp/[dataTypeProperty])

Príklady

<http://data.gov.sk/def/ontology/odp>
<http://data.gov.sk/def/ontology/odp/CorporateBody>
<http://data.gov.sk/def/ontology/odp/address>
<http://data.gov.sk/def/ontology/odp/validFrom>

PREFIX odp:
odp:CorporateBody
odp:address
odp:validFrom



3) SK GOV URIs - Individuá

návrhový vzor: [http://data.gov.sk/id/\[TRIEDA/SUBTRIEDA\]/\[ID\]](http://data.gov.sk/id/[TRIEDA/SUBTRIEDA]/[ID])

klasifikácia identita

Príklady

<http://data.gov.sk/id/corporatebody/35810734>
<http://data.gov.sk/id/region/sk01>
<http://data.gov.sk/id/municipality/SK0102529320>

4) SK GOV URIs - Katalógy

Katalógy (množina datasetov)

[http://data.gov.sk/set/cat/\[katalogID\]](http://data.gov.sk/set/cat/[katalogID])

Príklad

<http://data.gov.sk/set/cat/registered-drugs>

Datasety (napr. aj číselníky)

[http://data.gov.sk/set/\[obsah\]/\[datasetID\]](http://data.gov.sk/set/[obsah]/[datasetID])

Príklad

<http://data.gov.sk/set/codelist/cl0024>

<http://data.gov.sk/set/drugs/categorizeddrugs>

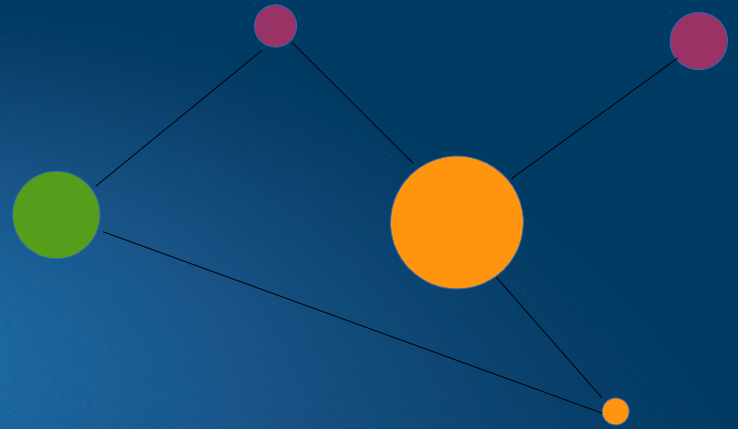
Distribúcia datasetu (publikovanie datasetov v čase)

[http://data.gov.sk/set//\[obsah\]/\[datasetID\]/\[dateTime\]](http://data.gov.sk/set//[obsah]/[datasetID]/[dateTime])

Príklad

<http://data.gov.sk/set/distribution/codelist/cl0024/1993-01-01>

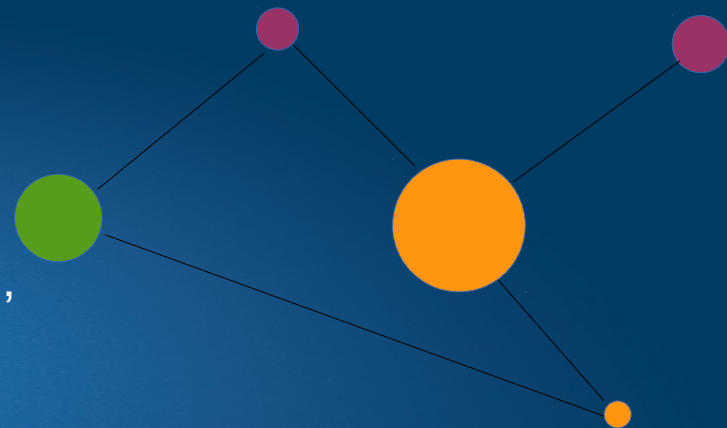




III. Nové časti na schválenie

5) Verziovane URI

Na koniec URI typu def, doc a set je možné pridať údaj o verzii. Neplatí to pre individuum, ktoré nemá verzie.



Príklad: Číselník právnych foriem má nasledovné URI

SK GOV URI

<http://data.gov.sk/set/codelist/cl0056/>

Pričom konkrétna verzia datasetu publikovaná v čase (distribúcia) má nasledovné URI

SK GOV URI

<http://data.gov.sk/set/codelist/cl0056/2005-10-15>

6) Prenos číselníkov do sémantiky

Prístup B

Položky číselníka budú indivíduá.

Napr. Bratislavský kraj bude

SK GOV URI

<http://data.gov.sk/id/region/1>

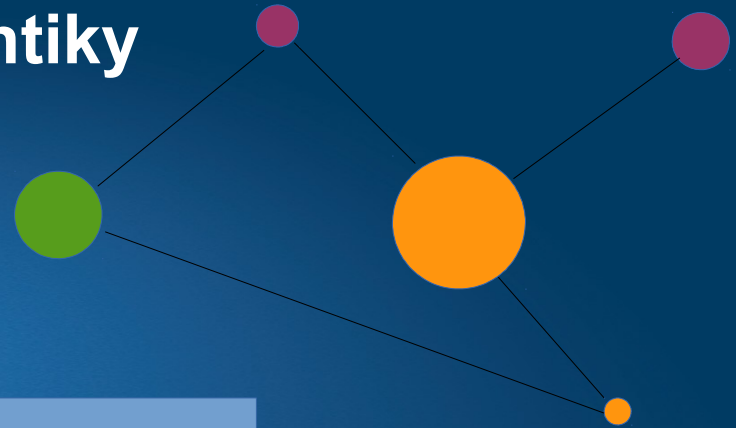
- Každá pôvodná položka číselníka bude mať svoju sémantickú triedu definovanú v ODP. Napr. uvedený BSK je typu odp:Region

```
<http://data.gov.sk/id/region/1 >
```

```
rdf:type
```

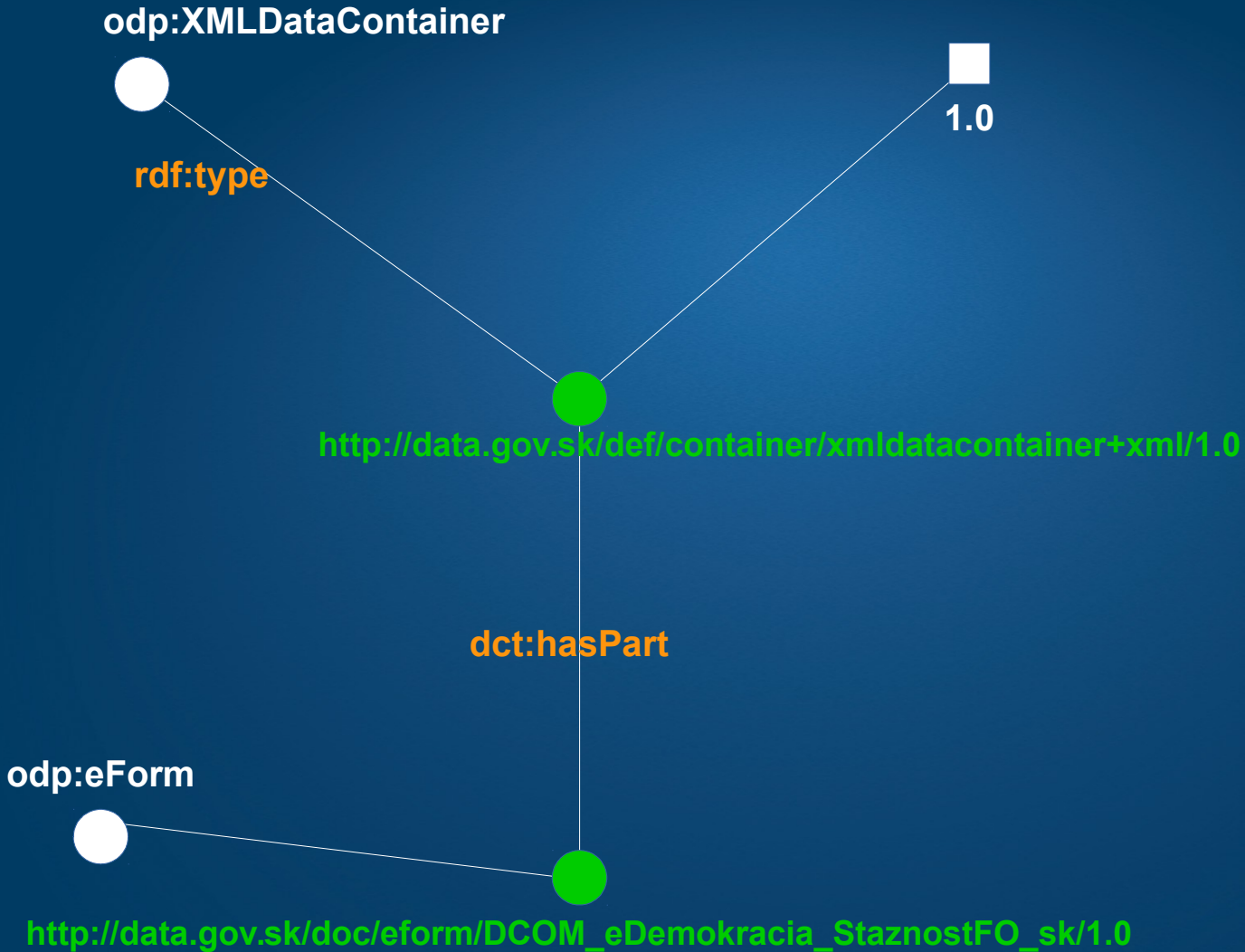
```
<http://data.gov.sk/def/ontology/odp
```

- Ak neexistuje pre položku číselníka prislúchajúca trieda, tak sa do ODP doplní.
- Pre podporu prechodu má každá trieda reprezentujúca číselník vlastnosť skos:historyNote nastavenú na kód číselníka.

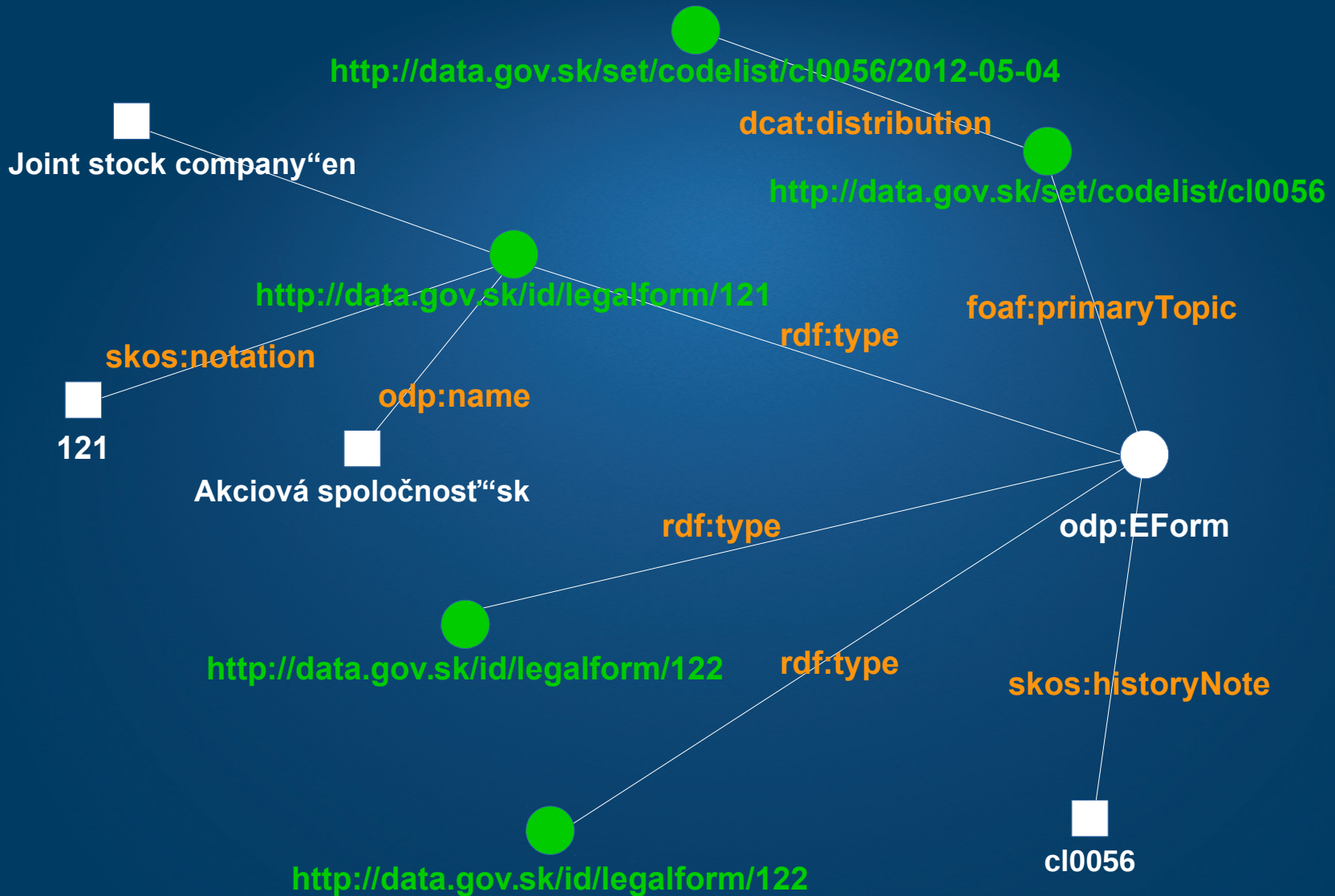


Ukážka na grafe >>

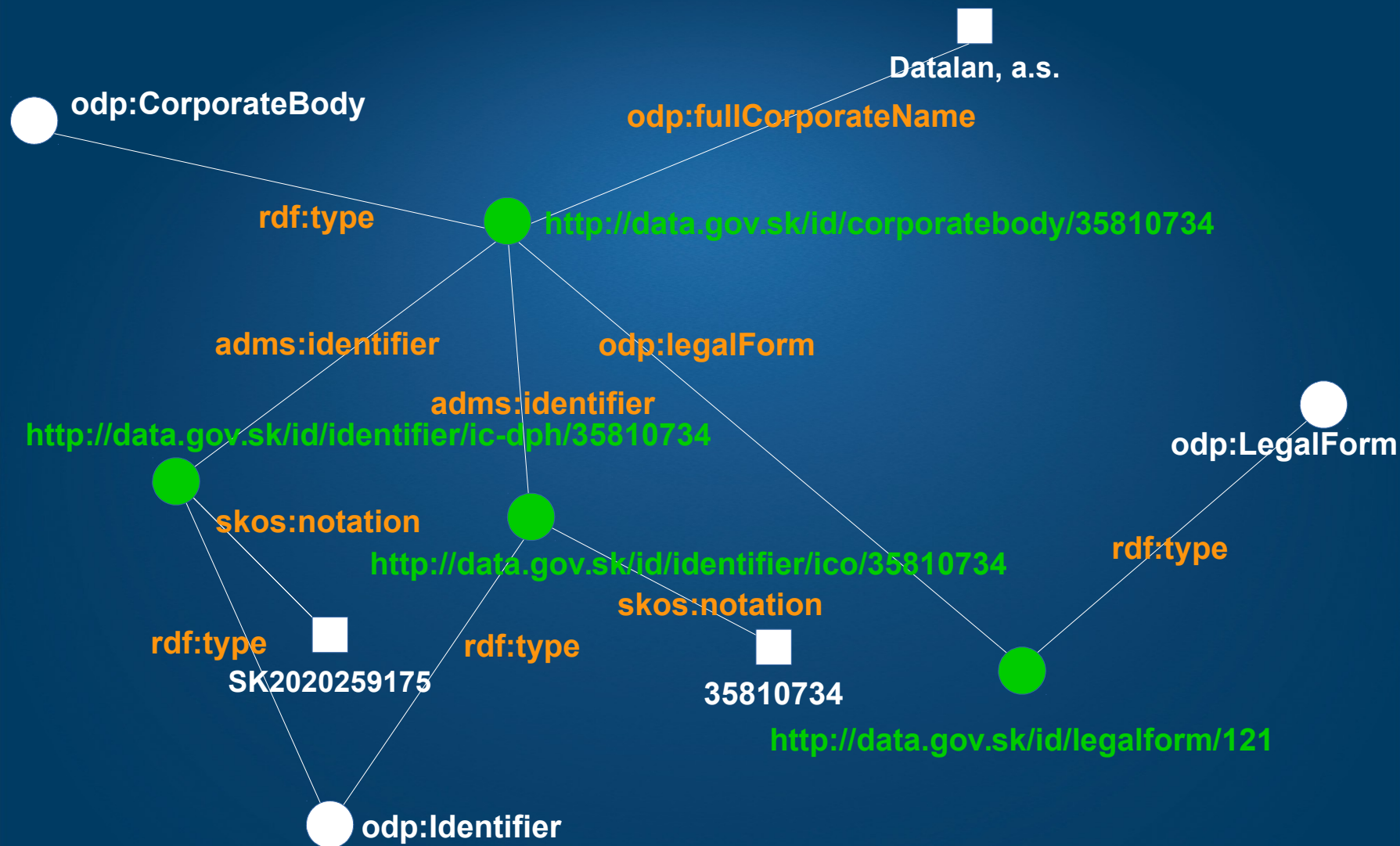
Príklad: Vzťah eForm + XMLDataContainer

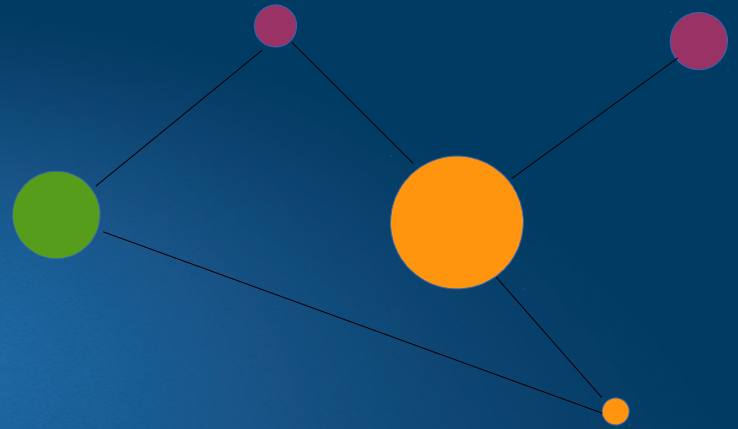


Príklad: Číselník právnych foriem a jeho distribúcia



Príklad: Právnická osoba + číselníky





IV. Problémové časti

8) Všeobecný formát SK GOV URI

Prístup B



SK GOV URI

`http://{data}{schemas}.gov.sk/[TYP]/{VLASTNIK}.[KLASIFIKACIA]/[IDENTITA]/{VERZIA}`

Problém1 – Trieda URI sa vyjadruje nesprávne

- To že je niečo schéma sa dá vyjadriť cez dátový prvok, napr.
`http://data.gov.sk/def/ontology/odp/Scheme`

čiže napr. schéma pre daňové priznanie sa dá vyjadriť tripletom

```
<http://data.gov.sk/doc/eform/xsd/danove_priznanie_k_dani_z_prijmov_fyzickych_osob_2014-sk-1.0>  
rdf:type  
<http://data.gov.sk/def/ontology/odp/XMLScheme>
```

- Stačí teda len `http://data.gov.sk...`

8) Všeobecný formát SK GOV URI

Prístup B



SK GOV URI

`http://{data}{schemas}.gov.sk/[TYP]/{VLASTNIK}.[KLASIFIKACIA]/[IDENTITA]/{VERZIA}/{LOKACIA+SUBORY}`

Problém2 URI nie je URL

- URI neslúži na to, aby sa veci dali zobrazovať v prehliadači, ale aby boli referencovateľné cez http protokol. To, kde sa nachádza samotný súvisiaci súbor sa rieši prostredníctvom ďalších vlastností URI, napr. predikátom `odp:downloadURL` alebo inými.

Čiže toto je nesprávne:

`<http://data.gov.sk/doc/eform/DCOM_eDemokracia_StaznostFO_sk/1.0/schema.xsd>`

Ale toto je už správne:

`<http://data.gov.sk/doc/xsdschema/DCOM_eDemokracia_StaznostFO_sk/1.0>`

A teda platí že

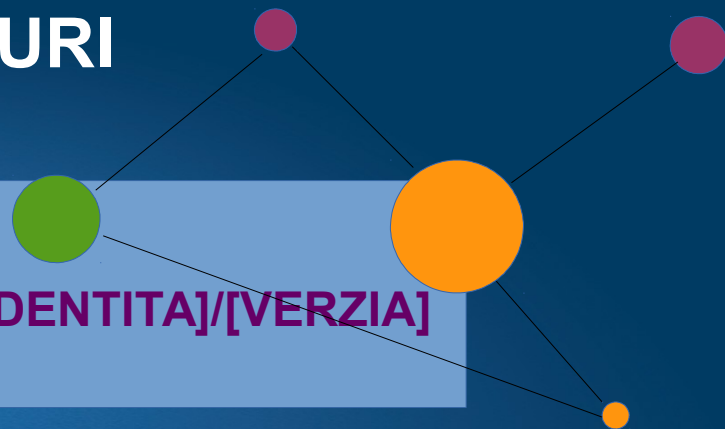
`<http://data.gov.sk/doc/xsdschema/DCOM_eDemokracia_StaznostFO_sk/1.0>`

`odp:downloadURL`

`<http://HOCIKDE/DOWLOAD/DCOM/eDemokracia_StaznostFO_sk/1.0/schema.xsd>`

8) Všeobecný formát SK GOV URI

Prístup A



SK GOV URI

[http://data.gov.sk/\[TYP\]/\[KLASIFIKACIA\]/\[IDENTITA\]/\[VERZIA\]](http://data.gov.sk/[TYP]/[KLASIFIKACIA]/[IDENTITA]/[VERZIA])

TYP

- id = konkrétne indivídium sveta (osoba, organizácia, obec ...)
- def = definovaná abstraktná entita (dátový prvok, dátový kontajner ...)
- doc = dokument, elektronický formulár...
- set = množina údajov, číselník

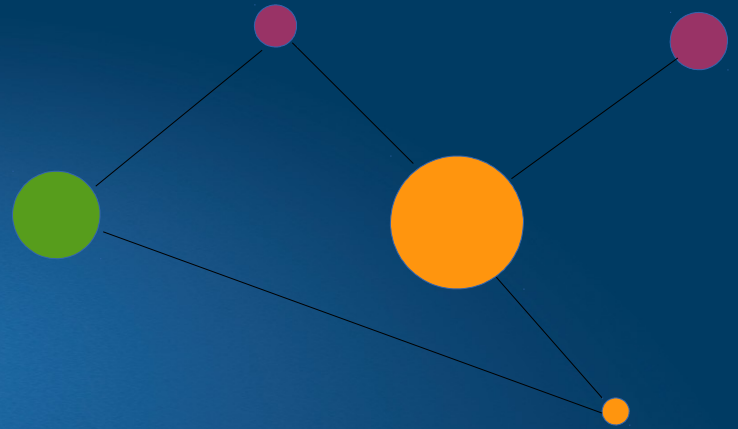
KLASIFIKACIA závisí od typu (vid' konkrétne príklady)

IDENTITA

- reťazec slúžiaci na jednoznačnú identifikáciu entity. Pozor, URI neslúži na parsovanie jeho vlastností (názov, vlastník, jazyk, verzia ...) alebo definovanie lokácie zdroja

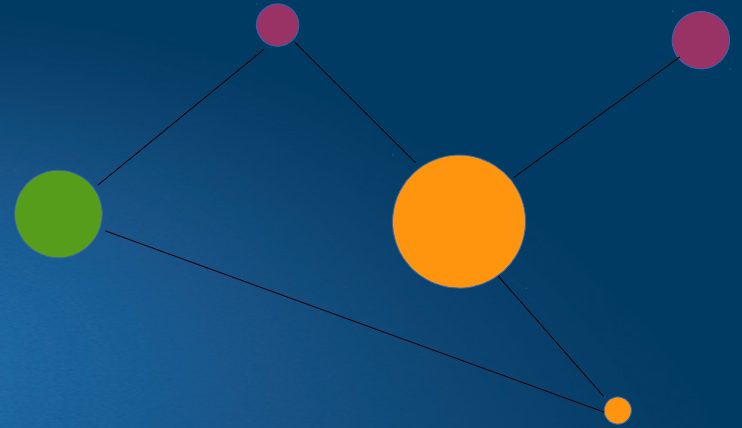
VERZIA

- je možná pre použitie s URI typu def, doc, set. Používa sa i na reprezentáciu distribúcie.



III. Vybrané príklady

Referencie



<http://sk.linkedin.com/in/miroslavliska/>

<http://www.datalan.sk>

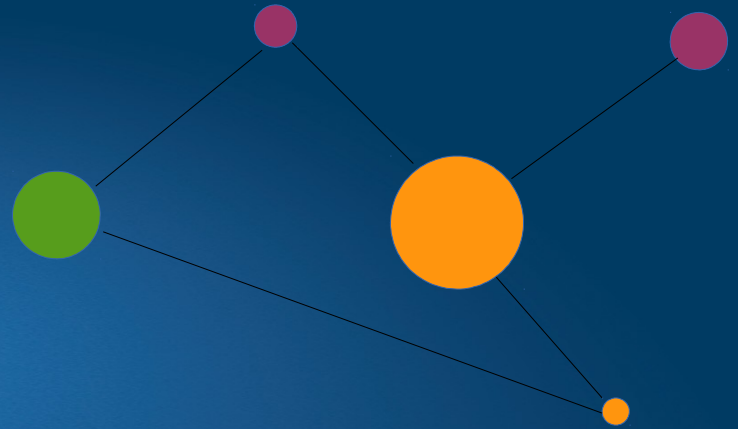
<http://www.w3.org/standards/semanticweb/>

<https://joinup.ec.europa.eu/community/semic/description>

<http://www.openrdf.org/>

<http://www.w3.org/OWL/>

<http://www.w3.org/RDF/>



Ďakujem za pozornosť