

A.4.3.1 Všeobecné pravidlá tvorby metadát verzionovateľných objektov

Kľúčovou ontológiou pri verzionovateľných entitách je ontológia adms

```
http://www.w3.org/ns/adms#
```

Každý prvok, ktorý je považovaný za verzionovateľný musí dediť od adms:Asset

```
x http://www.w3.org/2000/01/rdf-schema#subClassOf http://www.w3.org/ns/adms#Asset
```

Zdroj sémantických dát a práca s ním

V praxi potrebujú vývojári ako aj analytici často vedieť zdroj informácií. Ako príklad by sa dal uviesť číselník ulíc, ktorý potrebuje mať aplikácia nahrať vo svojom systéme. Nakoľko tento zdroj je verzionovateľný, tým pádom sa môže vyvíjať v čase a je nutné vedieť s ktorou verziou systém aktuálne pracuje, napríklad pre potreby integrácií medzi systémami.

V tomto ohľade je možné spracovávať verzionovateľné dáta týmito troma prístupmi :

1. Dáta v nesémantických databázach

- a. je nutné poznať verziu na dáta ako celok
Systém si v separetnej tabuľke-štruktúre zaznamená, že má v sebe nahrať dáta daného typu v konkrétnej verzii jednoznačne definovanej pomocou referencovateľného identifikátora. Napríklad má v sebe v tabuľke s názvom "kraje" dáta z <https://data.gov.sk/set/data/region/2017-01-01> a rovnakým princípom sú definované aj ostatné tabuľky, ktoré staticky používajú dáta vo svojom systéme.
- b. je nutné poznať verziu na každú položku z nahratých dát
V prípade, že človek potrebuje poznať každú jednu položku a jej príslušnosť k danému zdroju, na tento prípad je použitý vzťah rdfs:isDefinedBy. Príklad položky z <https://data.gov.sk/set/data/region/2017-01-01> , ktorá je reprezentovaná distribúciou <https://data.gov.sk/set/data/region/2017-01-01.rdf>

```
<loca:Region rdfs:about="https://data.gov.sk/id/region/2">
  <rdfs:label xml:lang="sk"><![CDATA[Bratislavský kraj]]></rdfs:label>
  <rdfs:label xml:lang="en"><![CDATA[Region of Bratislava]]></rdfs:label>
  <rdfs:label xml:lang="de"><![CDATA[Region of Bratislava]]></rdfs:label>
  <adms:identifier rdfs:resource="https://data.gov.sk/id/internal-is-id-type/2/2" />
  <rdfs:isDefinedBy rdfs:resource="https://data.gov.sk/set/data/region/2017-01-01" />
</loca:Region>
```

Ako je vidieť, uzol má v sebe reláciu príslušnosti k danej verzii dátového zdroja (datasetu). Takto je možné spätne dohľadávať ako informácie v systéme vznikali a v akých verziách ich systém používa až na úroveň položky.

2. Dáta v sémantických databázach

Sémantické databázy dokážu problém rozdielnych verzií riešiť aj na základe natívnych mechanizmov. Sémantická databáza je reprezentovaná grafom. Zároveň je možné v každej databáze špecifikovať aj podgraf (kontext) a doňho nahrávať dáta, ktoré môžu byť ohraničované práve definíciou do ktorého podgrafu patria. Tak je možné vložiť obsah súboru adms:AssetDistribution do podgrafu.

a) Príklad extrakcie dát konkrétnej verzie :

Dáta z nášho príkladu krajov vložíme do sémantickej databázy do podgrafu s URI identifikátorom rovnakým aký ma daný adms:Asset tj. <https://data.gov.sk/set/data/region/2017-01-01> .

Následne vieme robiť také dopyty nad databázou, ktoré obmedzia jeho výsledky na daný subgraf. V jazyku SPARQL by to vyzeralo nasledovne :

```
SELECT ?p ?o WHERE {
  GRAPH <https://data.gov.sk/set/data/region/2017-01-01> {
    {
      <https://data.gov.sk/id/region/2> ?p ?o
    }
  }
}
```

b) Príklad extrakcie dát v prípade existencie viacerých entít toho istého datasetu nahratých v databáze :

Povedzme, že vznikne nová verzia daného datasetu s URI identifikátorom <https://data.gov.sk/set/data/region/2017-01-02> v ktorom sa zmení názov kraja

```
<loca:Region rdf:about="https://data.gov.sk/id/region/2">
  <rdfs:label xml:lang="sk"><![CDATA[Kraj bratislavský]]></rdfs:label>
  <rdfs:label xml:lang="en"><![CDATA[Bratislava Region]]></rdfs:label>
  <rdfs:label xml:lang="de"><![CDATA[Bratislava Region]]></rdfs:label>
  <adms:identifier rdf:resource="https://data.gov.sk/id/internal-is-id-type/2/2"/>
  <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/set/data/region/2017-01-02"/>
</loca:Region>
```

Pre vytiahnutie aktuálnych dát z datasetu stačí zmeniť URI daného podgrafu :

```
SELECT ?p ?o WHERE {
  GRAPH <https://data.gov.sk/set/data/region/2017-01-02> {
    {
      <https://data.gov.sk/id/region/2> ?p ?o
    }
  }
}
```

Ak nezadáme podgraf, dostávame konsolidovaný pohľad na dáta.

```
SELECT ?p ?o ?graph WHERE {
  GRAPH ?graph {
    {
      <https://data.gov.sk/id/region/2> ?p ?o
    }
  }
}
```

Vráti zoznam všetkých vlastností v rôznych verziách:

?p		?
o		
?graph		
rdfs:label	"Kraj bratislavský"	
@sk		https://data.gov.sk/set/data/region/2017-01-02
rdfs:label	"Bratislava Region"	
@en		https://data.gov.sk/set/data/region/2017-01-02
rdfs:label	"Bratislava Region"	
@de		https://data.gov.sk/set/data/region/2017-01-02
rdfs:label	"Kraj bratislavský"	
@sk		https://data.gov.sk/set/data/region/2017-01-01
rdfs:label	"Kraj bratislavský"	
@en		https://data.gov.sk/set/data/region/2017-01-01
rdfs:label	"Kraj bratislavský"	
@de		https://data.gov.sk/set/data/region/2017-01-01
adms:identifier	https://data.gov.sk/id/internal-is-id-type/2/2	
https://data.gov.sk/set/data/region/2017-01-02		
adms:identifier	https://data.gov.sk/id/internal-is-id-type/2/2	
https://data.gov.sk/set/data/region/2017-01-01		
rdfs:isDefinedBy	https://data.gov.sk/set/data/region/2017-01-02.rdf	https://data.gov.sk
/set/data/region/2017-01-02		
rdfs:isDefinedBy	https://data.gov.sk/set/data/region/2017-01-01.rdf	https://data.gov.sk
/set/data/region/2017-01-01		

c) Príklad hľadania všetkých zdrojov v ktorých sa dané URI nachádza

Ak máme nahratých viacero datasetov v rôznych verziách, ktoré hovoria o tej istej entite, vieme si ich dohľadať pomocou SPARQL dopytu:

```
SELECT DISTINCT ?context WHERE {
  GRAPH ?context {
    {
      <https://data.gov.sk/id/region/2> ?p ?o
    }
    UNION
    {
      ?s ?p <https://data.gov.sk/id/region/2>
    }
  }
}
```

Výsledkom takéhoto dopytu v našom prípade by bol zoznam dvoch hodnôt :

```
https://data.gov.sk/set/data/region/2017-01-01
https://data.gov.sk/set/data/region/2017-01-02
```