

Sémantické štandardy pre ISVS

<http://data.gov.sk/doc/semantics-standards/2016-05-21.pdf>

**- pracovná verzia poslaná na pripomienkovanie do pracovnej skupiny
PS1 na Ministerstve Financíí a do otvoreného odborného fóra
slovakia.digital -**

celý obsah metodiky je možné stiahnuť z
<http://www.semantickyweb.sk/standards/data.gov.sk-semanticweb/2016-05-21.zip>

Verzia	Popis zmeny	Autor
2015-11-02	Prvá verzia v roku 2015 poslaná na pripomienkovanie do pracovnej skupiny PS1.	Miroslav Líška Marek Šurek
2016-05-21	Druhá prepracovaná verzia poslaná na pripomienkovanie do PS1.	Miroslav Líška Marek Šurek

ČASŤ A. ÚVOD

- §A1 Úvodné definície (Sémantický web v kocke)
- §A2 Strategický zámer použitia sémantiky na verejné dáta

ČASŤ B. URI Identifikátor všeobecne

- §B1 Všeobecné požiadavky na data.gov.sk URI identifikátora
- §B2 Všeobecný návrhový vzor data.gov.sk URI identifikátora
- §B3 Spôsob verziovania URI
- §B4 Spôsob jazykovej lokalizácie hodnôt vlastností URI
- §B5 Šablóna pre data.gov.sk URI

ČASŤ C. Znalosti

- §C1 URI pre data.gov.sk ontológie
- §C2 URI pre data.gov.sk triedy
- §C3 URI pre data.gov.sk vlastnosti
- §C4 URI pre data.gov.sk definície
- §C5 Zoznam podporovaných štandardizovaných ontológií

ČASŤ D. Individuá

- §D1 Všeobecné URI pre individuá
- §D2 URI pre fyzickú osobu
- §D3 URI pre organizáciu

ČASŤ E. Dokumenty

- §E1 Všeobecné URI pre dokument
- §E2 Definícia URI pre elektronický formulár

ČASŤ F. Datasetsy

- §F1 URI pre číselník/hierarchiu/katalóg
- §F2 URI pre položku číselníka
- §F3 URI pre distribúciu číselníka/hierarchie/katalógu

ČASŤ G. URI referenčných údajov

- §G1 Základné pravidlá tvorby URI referenčných údajov

- + Dokumentácia
- + Príklady OWL/XML
- + Referencie

ČASŤ A. Úvod

§A1 Úvodné definície (Sémantický web v kocke)

URI (Uniform Resource Identifier) – Jednotný referencovateľný identifikátor je reťazec znakov používaných na identifikáciu entity v sémantickom webe v tvare `http://...` Na rozdiel od URL (Uniform Resource Locator) neslúži na navigáciu na stránku kvôli jej zobrazeniu, ale slúži na identifikáciu predmetnej entity.

RDF (Resource Description Framework) – Jazyk pre popis informácií o zdrojoch na sémantickom webe. Zdrojom môže byť entita reprezentujúca osobu, rieku, knihu, ... , ale aj vlastnosť, trieda, klasifikácia a podobne.

RDF Zdroj (RDF Resource) – je zdroj referencovateľný prostredníctvom URI

Nasledovný príklad prezentuje použitie URI na identifikáciu troch RDF zdrojov: konkrétnej osoby, vlastnosti priezvisko a identifikáciu triedy fyzická osoba.

```
<http://data.gov.sk/id/president/andrej-kiska>
<http://data.gov.sk/def/ontology/physical-person/familyName>
<http://data.gov.sk/def/ontology/physical-person/PhysicalPerson>
```

Indivídium – je objekt reálneho sveta, v sémantickom webe reprezentované ako URI. V predchádzajúcom príklade je len prvé URI indivídium.

Triplet – je reprezentácia znalosti vo forme trojice Subjekt —Predikát—> Objekt. Prvé dva členy sú vždy reprezentované ako URI, tretí byť URI nemusí. Ak je, predikát sa nazýva objektová vlastnosť (Object Property). Ak nie, objektom tripletu je konkrétna dátová hodnota, a vtedy sa predikát nazýva dátová vlastnosť (Data Type Property).

Nasledovné príklady prezentujú dva triplety. Prvý definuje zamestnanie Andreja Kisku a tu je použitá objektová vlastnosť. Druhý triplet definuje priezvisko nášho prezidenta a tu je použitá dátová vlastnosť.

```
<http://data.gov.sk/id/president/andrej-kiska>
<http://data.gov.sk/def/ontology/physical-person/occupation-type>
<http://data.gov.sk/def/occupation-type/1111001> .

<http://data.gov.sk/id/president/andrej-kiska>
<http://data.gov.sk/def/ontology/physical-person/familyName>
"Kiska"^^xsd:string .
```

Odvođený triplet – je triplet, ktorý vznikol strojovým odvođeným na základe iných tripletov.

Vyššie uvedený triplet definoval, že nejaký RDF zdroj má priezvisko Kiska. Keďže platí, že dátová vlastnosť priezvisko je definovaná medzi Osobou a reťazcom (tj. doména relácie je Osoba a rozsah je reťazec), teda

```
<http://data.gov.sk/def/ontology/physical-person/givenName>
rdfs:domain
<http://data.gov.sk/def/ontology/physical-person/PhysicalPerson> .
```

potom je možné odvodiť, že predmetný RDF zdroj reprezentuje fyzickú osobu, tj.

```
<http://data.gov.sk/id/president/andrej-kiska>
rdf:type
<http://data.gov.sk/def/ontology/physical-person/PhysicalPerson> .
```

Odvodzovanie – Strojové odvodzovanie (machine reasoning) je disciplína umelej inteligencie, ktorá sa zaoberá odvodzovaním nových znalostí (tripletovej) z existujúcich. Inými slovami, stroj sa snaží odvodiť, aké nové triplety vie odvodiť z existujúcich.

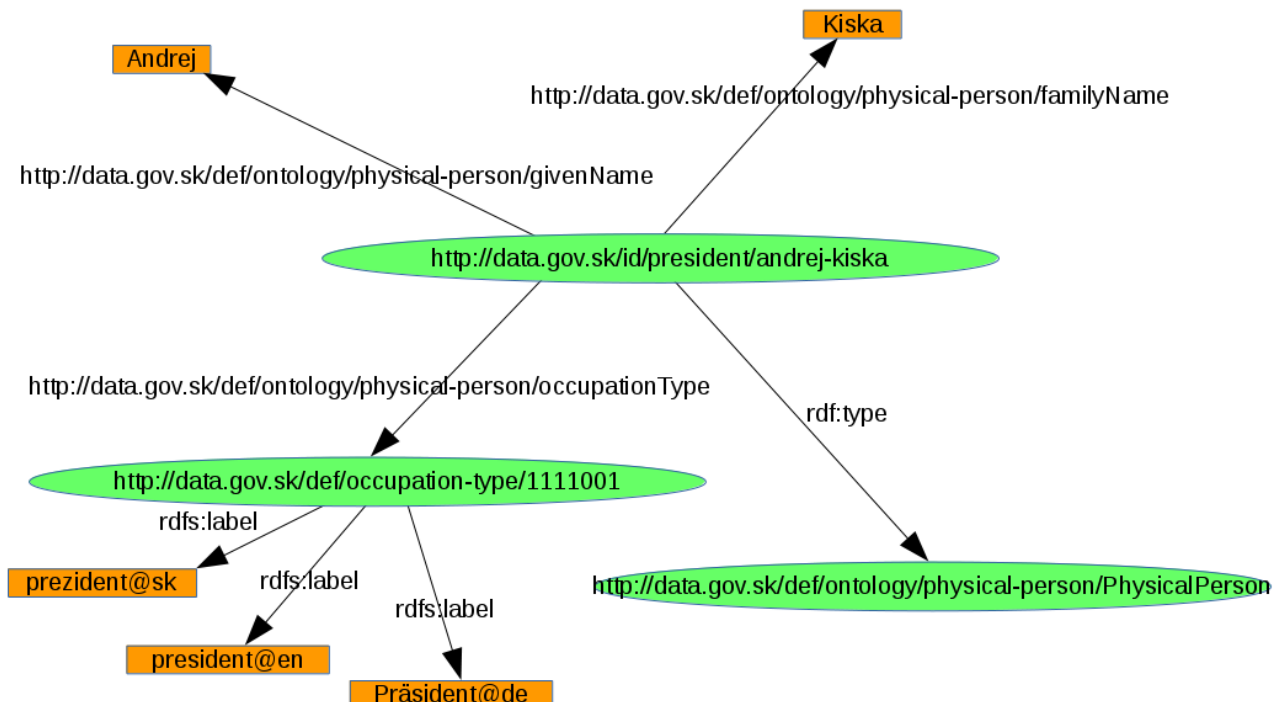
Príklad

Predstavme si situáciu, že v systéme je uložený triplet: Lukáško má brata Slávka. Klasický relačný systém vie odpovedať správne iba na otázku, kto je brat Lukáška? Slávko. Ale ak sa spýtam, kto je brat Slávka, tak mi správne odpovie len sémantický systém, pretože ten si z tripletu: Lukáško má brata Slávka odvodí inverzný triplet: Slávko má brata Lukáška, nakoľko vlastnosť „má brata“ je symetrická.

Význam odvodzovania nie len vo vyhľadávani, ale najmä pri celkovej integrácii rôznorodých informačných systémov, pri overovaní logických faktov a podobne. Reálne sa odvodzovanie vykonáva prostredníctvom odvodzovača (reasonera), čo býva štandardný komponent sémantickej databázy (triplestore). Existujú dva základné princípy odvodzovania – dopredné zreťazenie a spätné zreťazenie. Prvé sa deje hneď pri vkladani tripletov do databázy, čiže spomínaný inverzný triplet by sa vytvoril okamžite, kým druhý typ odvodzovania sa deje až pri dopyte, tj. až keď sa spýtam Kto je bratom Slávka. Až vtedy sa odvodí, resp. pokúsi odvodiť inverzný triplet . Pre kľúčovú požiadavku rýchlosti sa v komerčných aplikáciach preferuje prvý prípad. Pri vložení sa okamžite odvodí nové triplety aby odpovede na databázu boli už vypočítané.

RDF Graf – je orientovaný graf, tj. množina vrcholov a hrán, ktoré sú tvorené RDF tripletmi. Vrcholy sú buď URI zdroje alebo konkrétne dátové hodnoty. Hrany, tj. vlastnosti sú vždy URI zdroje.

Nasledovný príklad prezentuje RDF graf obsahujúci 7 tripletov reprezentujúcich súčasného prezidenta SR Andreja Kisku.



Ontológia – je množina *tripleto*v ktorá opisuje vybranú doménu. V súčasnosti samozrejme existuje obrovské množstvo ontológií, aj to aj v zmysle ich strojovo spracovateľnej formy (OWL, RDF). Napr. ontológia osoby, organizácie, ontológia rastlín, ontológia odvetví sektora XY a podobne. Takmer každá dôležitá klasifikácia niečoho existuje už vo forme strojovo-spracovateľnej ontológie, ako napr. ontológia geografických prvkov, ontológia medzinárodnej klasifikácie chorôb, ontológia nomenklatúry územných entít a podobne.

Vo všeobecnosti, ontológia v prvom rade definuje triedy a ich vzájomné vzťahy, pričom medzi základné vzťahy patria ich hierarchické relácie ako `rdfs:subClassOf` (podtrieda, resp. dedenie). V tomto kontexte je dôležité povedať, že ontológia je všeobecnejší pojem ako model, pretože model (napr. UML model) je vlastne slabá ontológia. Nie je možné použiť také silne odvodzovanie ako v prípade OWL, tj. sémantickej ontológie.

Dataset – sprístupnená publikovaná množina údajov, v sémantickom webe tripletov. Hlavné typy datasetu sú číselník alebo hierarchia. Kým číselník sa zameriava na výčet entít z ich základnými vlastnosťami, hierarchia obohacuje číselníky o hierarchické relácie, ako napr. `skos:broader` a `skos:narrower` (všeobecnejší a presnejší).

Distribúcia – je verzia (najčastejšie časová) katalógu / datasetu.

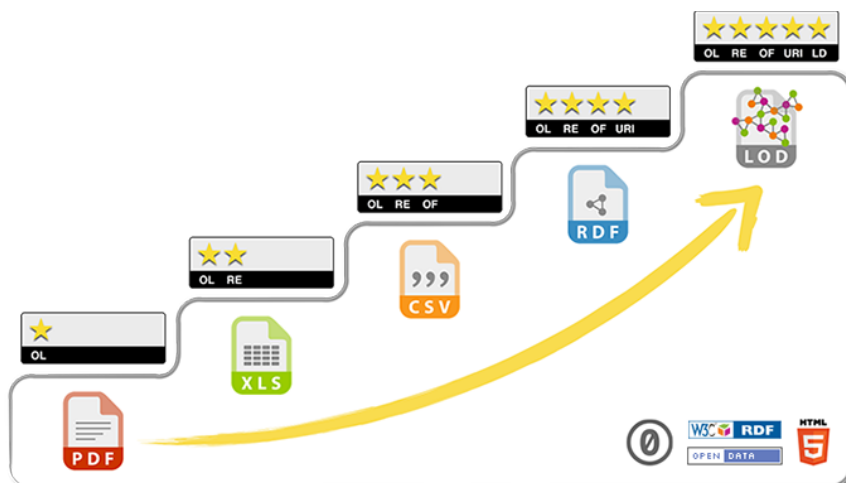
Katalóg - množina datasetov.

RDF Databáza / Sémantická databáza – je databáza určená na spracovanie RDF Grafu.

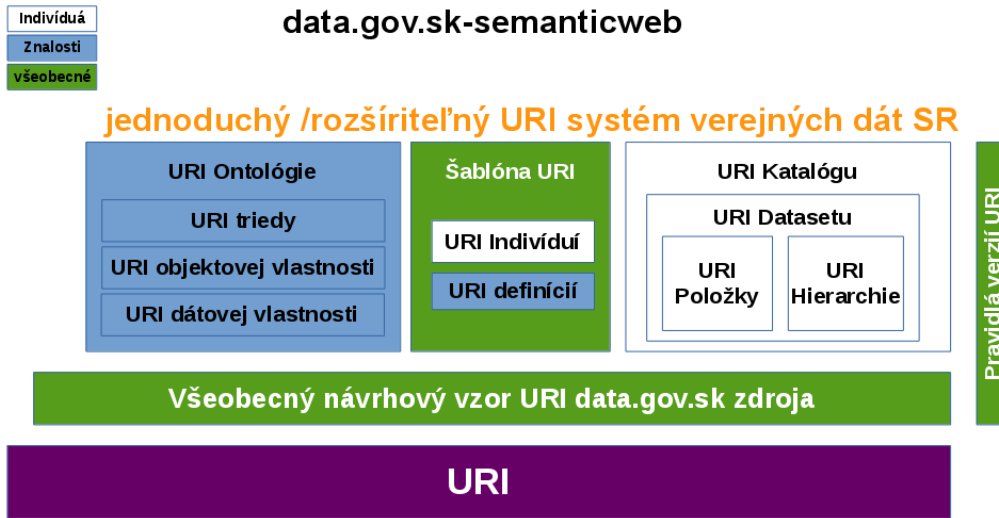
Sémantický web – je nová generácia Webu (3.0), ktorú definuje konzorcium W3C. Informácie na webe sú reprezentované prostredníctvom RDF/OWL, teda je možné ich spracovávať aj s ohľadom na ich význam. Tvorcom Sémantického webu je Tim Berners Lee, ktorý je súčasne aj autorom klasického webu, tj. WWW.

Linked Data – Prepojené dáta, je iný pojem na sémantické dáta, resp. triplet, resp. RDF/OWL. Týmto pomenovaním sa vyzdvihuje skutočnosť orientácie návrhu dát na ich maximalizáciu prepojitelnosti.

5★Open Data – Tim Berners-Lee, tvorca webu a iniciátor sémantického webu odporučil tzv. 5 hviezdikovú klasifikáciu otvorených dát. Platí, čím viac hviezdíček, tým sú otvorené dáta lepšie otvorené. Je zrejme, že kvoli výhodám sémantiky sú takto anotované dáta vyhodnotené ako najlepšie.



data.gov.sk-semanticweb – je metodika sémantických štandardov (5★ OpenData) pre údaje ISVS. Jej základom je rozšíriteľný URI systém, umožňujúci popísať (katalogizovať) údaje verejnej správy do komplexného RDF grafu.



§A2 Strategický zámer použitia sémantiky vo verejných dátach

- i Poskytovanie všetkých verejných dát vo forme 4★ a 5★ z pohľadu 5★open data pre nasledovné potreby
 - o Umožnenie efektívne prepojiť všetky dáta verejnej správy, tj, rôzne registre, datasety, číselníky, hierarchie prostredníctvom sémantiky zavedením jednotných URI identifikátorov a štandardizovaných ontológií pre jednotlivé entity.
 - o Zvýšenie transparentnosti a otvorenosti dát verejnej správy prostredníctvom prepojenia rôznych údajov z rôznych datasetov o zainteresovaných osobách, organizáciach a podobne
 - o Umožnenie silno-efektívneho použitia prepojených dát na tzv. sémantický mashup, keď je potrebné spracovať viac datasetov bez nutnosti nákladej integrácie

ČASŤ B. URI Identifikátor všeobecne

§B1 Všeobecné požiadavky na data.gov.sk URI

- i URI musí byť čo najtrvácnejšie (perzistentné). Jeho reťazec nesmie obsahovať nejaký subreťazec ktorého platnosť je časovo veľmi obmedzená. Napr. nesmie obsahovať gestora, pretože gescia sa môže časom zmeniť.
- ii URI primárne neslúži na lokáciu k zdroju, ale na identifikáciu (referencovanie) nejakej entity. Zobrazovanie konkrétneho obsahu zabezpečuje tzv. dereferenciácia URI, tj. zobrazenie obsahu podľa typu konzumenta (používateľovi sa vrátia dáta v html obsahu, stroju zas strojovo čitateľný formát (RDF/XML))
- iii URI reprezentujúce konkrétnu časovú verziu súboru (napr. verzia ontológie, alebo distribúcia datasetu, verzia dokumentu, XSD súboru) nie je považované za trvácne, a môže referovať aj priamo na súbor.
- iv Keďže s URI nepracuje len stroj ale aj človek počas všetkých disciplín softvérového procesu (analýza, návrh, implementácia, testovanie), podstatná časť URI, najmä jej klasifikácia, musí byť čo najviac ľudsky čitateľná. Samotná identita v časti URI môže byť reprezentovaná štandardným IDčkom.
- v URI nesmie obsahovať ani programátorské znaky používané pre potrebu URL ako napr. page.asp?id=4557&gui=777.

§B2 Všeobecný návrhový vzor pre data.gov.sk URI

URI vzor

```
http://data.gov.sk/[type]/[classification]+/[identity]/{version}
```

Dovolená Syntax

[type] =

“def“ ak URI reprezentuje entitu ontológie

“id“ ak URI reprezentuje individuuum

“set“ ak URI reprezentuje dataset

“doc“ ak URI reprezentuje dokument

[classification]+ = trieda alebo koncept, pričom klasifikácia môže mať viac úrovní.

Účelom viacúrovňovej definície je umožnenie presnejšej definície a zabezpečenie jednoznačnosti URI.

[identity] = klasické pôvodné IDčko

{version} = Platí, že individuá, ktoré môžu existovať súčasne v rôznych verziách môžu mať aj verziu, ostatné nie.

Príklad

Prvé URI je príkladom osoby, samozrejme bez verzie (Andrej Kiska ako človek nemôže súčasne existovať v dvoch rôznych verziach). Druhé URI reprezentuje akciovú spoločnosť (URI je vytvorené na základe kódu položky 121 v číselníku právnych foriem organizácie, klasifikácia je odvodená od pomenovania triedy ktorej položky sú inštancie). Tretie URI reprezentuje samotný číselník právnych foriem organizácie, pričom štvrté URI je jeho konkrétnou stiahnuteľnou distribúciou vo formáte linked-data.

```
<http://data.gov.sk/id/president/andrej-kiska>
<http://data.gov.sk/def/legal-form/121>
<http://data.gov.sk/set/codelist/legal-form/>
<http://data.gov.sk/set/codelist/legal-form/2016-05-29.rdf>
```

§B3 Spôsob verziovania URI**Popis**

V zmysle použitia verzionovania entít data.gov.sk sa verzionovanie nepoužíva na fyzické osoby ani právnické osoby (tj. Agent) a súčasne ani geopriestorové entity. Na druhej strane, verzionovanie sa najčastejšie používa pre softvér, tj. dokumenty, programy, a podobne.

Príklad

Nasledovné triplety definujú aktuálnu verziu tohto dokumentu a verziu 1.0 Informačného systému pre verejné obstarávanie.

```
<http://data.gov.sk/doc/semantics-standards/2016-05-19-en.pdf>
rdf:type
<http://data.gov.sk/def/ontology/document/Document> .

<http://data.gov.sk/id/egov/isvs/5854/1.0 >
rdf:type
<http://data.gov.sk/def/ontology/egov/ISVS> .
```

§B4 Spôsob jazykovej lokalizácie hodnôt vlastností URI

Pri lokalizácii hodnôt dátových vlastností je nutné rozlišovať tzv. jazykovo lokalizovateľné a jazykovo nelokalizovateľné reťazce, ako napr. názvy osôb a podobne. V prvom je sa jazyková verzia udáva, v druhom sa definuje len typ hodnoty.

Príklad

V nasledovnom príklade, keď je možné reťazec „preložiť“ do iného jazyka, tak sa daná dátová vlastnosť použije viackrát s upresnením jednotlivých jazykových verzií (@sk, @en ...). Nasledovné triplety definujú triedu informačného systému pre verejnú správu SR v dvoch jazykových verziách.

```
<http://data.gov.sk/def/ontology/egov/ISVS>
rdf:type
  <http://www.w3.org/2002/07/owl#Class> ;
rdfs:label
  "informačný systém verejnej správy"@sk ;
rdfs:label
  "public administration information system"@en ;
```

Druhý príklad definuje priezvisko nášeho súčasného prezidenta ako príklad jazykovo nelokalizovateľného reťazca.

```
<http://data.gov.sk/id/president/andrej-kiska>
odp:familyName
  "Kiska"^^xsd:string .
```

§B5 URI šablóna

Popis

Keďže pre potreby registrácie URI cez MetaIS nie je efektívne registrovať každé individuum a definíciu samostatne (priveľký počet), tak individuá a definície je možné registrovať len ako celú množinu prostredníctvom tzv. šablóny URI.

URI vzor

```
http://data.gov.sk/[type]/[concept]/[[IDENTITY]]
```

Príklad

Nasledovný triplet predstavuje definovanie jedinej organizácie samostatne.

```
<http://data.gov.sk/id/corporate-body/35810734>  
rdf:type  
<http://data.gov.sk/def/ontology/organization/CorporateBody> .
```

Ďalší triplet definuje šablónu pre individuá všetkých organizácií naraz (v MetaIS).

```
<http://data.gov.sk/id/corporate-body/[ICO]>  
<http://data.gov.sk/def/ontology/res/isURITemplateOf>  
<http://data.gov.sk/def/ontology/organization/CorporateBody>
```

Iný triplet definuje šablónu pre definície typov pohlaví (v MetaIS).

```
<http://data.gov.sk/def/sex/[ID]>  
<http://data.gov.sk/def/ontology/res/isURITemplateOf>  
<http://data.gov.sk/def/ontology/physical-person/Sex>
```

ČASŤ C. ZNALOSTI (Ontológie)

§C1 URI pre data.gov.sk ontológie

URI vzor

```
http://data.gov.sk/def/ontology/[ontologyname]/{version}
```

Dovolená Syntax

[ontologyname] = všetko malé písmená

Príklad

Nasledovné triplety definujú dve ontológie, Ontológiu fyzickej osoby a Ontológiu eGovernment entít.

```
<http://data.gov.sk/def/ontology/physical-person> rdf:type owl:Ontology .
<http://data.gov.sk/def/ontology/egov> rdf:type owl:Ontology .
```

§C2 URI pre data.gov.sk triedy

URI vzor

```
http://data.gov.sk/def/ontology/[ontologyname]/[Class]
```

Dovolená Syntax

[Class] = názov triedy sa začína vždy veľkým písmenom, ďalej nasleduje Camel formát, tj. striedanie veľkých a malých písmen, vždy na začiatku nového slova v reťazci.

Poznámka:

URI triedy sa neverzionuje. Ak vznikne zmena/doplnenie URI triedy, toto sa odráža v novej verzii ontológie, ktorej je daná trieda súčasťou.

Príklad

Nasledovné triplety definujú triedu Informačný systém verejnej správy ISVS v rámci ontológie EGOV a súčasne triedu Pohlavie v rámci ontológie fyzická osoba.

```
<http://data.gov.sk/def/ontology/egov/ISVS> rdf:type owl:Class .
<http://data.gov.sk/def/ontology/physical-person/Sex> rdf:type owl:Class .
```

§C3 URI pre data.gov.sk vlastnosti

URI vzor

```
http://data.gov.sk/def/ontology/[ontologyname]/[property]
```

Dovolená Syntax

[property] = názov vlastnosti sa začína vždy malým písmenom, ďalej nasleduje Camel formát, tj. striedanie veľkých a malých písmen, vždy na začiatku nového slova v reťazci.

Poznámka:

Podobne ako URI triedy, ani URI vlastnosti sa neverzionuje. Ak vznikne zmena/doplnenie URI vlastnosti, toto sa odráža v novej verzii ontológie, ktorej je daná vlastnosť súčasťou.

Príklad

Nasledovné triplety definujú objektovú vlastnosť sídlo firmy a dátovú vlastnosť priezvisko.

```
<http://data.gov.sk/def/ontology/organization/headquartes>
rdf:type
owl:ObjectProperty .

<http://data.gov.sk/def/ontology/physicalPerson/familyName>
rdf:type
owl:DatatypeProperty.
```

Pozn. objektová vlastnosť je relácia medzi dvoma URI entitami a dátová je medzi URI a dátovou hodnotu. Nasledovné triplety definujú URI adresy spoločnosti Datalan a priezvisko osoby Miroslav Líška.

```
<http://data.gov.sk/id/organization/35810734>
<http://data.gov.sk/def/ontology/organization/headquartes>
<http://data.gov.sk/id/organization/address/35810734> .

<http://data.gov.sk/id/creator/semantic-standards-miroslav-liska>
<http://data.gov.sk/def/ontology/physical-person/familyName>
"Líška"^^xsd:string .
```

§C4 URI pre data.gov.sk definície

Popis: Definície sú špeciálne dátové entity, ktoré vykazujú vlastnosti aj znalostí, aj indivíduí. Formálne sa jedná o definovanie triedy výčtom jej možných inštancií, čomu zodpovedá „sémantizácia“ položiek číselníkov s definíciami, ako napr. typ pohlavia, typ organizácie a podobne.

URI vzor

```
http://data.gov.sk/def/[concept]/[identity]
```

Príklad

Nasledovné triplety reprezentujú definovanie mužského a ženského pohlavia, čo odpovedá položkám číselníka Pohlavie. V príklade je ilustrované, že trieda Pohlavie má inštancie muž a žena.

```
<http://data.gov.sk/def/sex/1>
rdf:type
<http://data.gov.sk/def/ontology/physical-person/Sex> .

<http://data.gov.sk/def/sex/2>
rdf:type
<http://data.gov.sk/def/ontology/physical-person/Sex> .
```

§C5 Zoznam podporovaných štandardizovaných ontológií

Základné ontológie

RDF <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
RDFS <http://www.w3.org/2000/01/rdf-schema#>
OWL <http://www.w3.org/2002/07/owl#>
SKOS <http://www.w3.org/2004/02/skos/core#>
DCT <http://purl.org/dc/terms/>
WGS84_POS http://www.w3.org/2003/01/geo/wgs84_pos#
GEONAMES <http://www.geonames.org/ontology>
FOAF <http://xmlns.com/foaf/0.1/>

SEMIC Odporučené ontológie

ADMS <http://www.w3.org/ns/adms>
ADMS.SW <http://purl.org/adms/sw/>
CPSV <http://purl.org/vocab/cpsv>
DCAT <http://www.w3.org/ns/dcat>
LOCN <http://www.w3.org/ns/locn#>
PERSON <http://www.w3.org/ns/person#>
REGORG <http://www.w3.org/ns/regorg>
NUTS <http://rdfdata.eionet.europa.eu/ramon/ontology>

data.gov.sk ontológie

Zdroj <http://data.gov.sk/def/ontology/resource>
Fyzická osoba <http://data.gov.sk/def/ontology/physical-person>
Právnická osoba <http://data.gov.sk/def/ontology/organization>
Identifikátor <http://data.gov.sk/def/ontology/identifier>
EGOV <http://data.gov.sk/def/ontology/egov>
IKT <http://data.gov.sk/def/ontology/ikt>

Vzťah medzi data.gov.sk a ostatnými ontológiami je prepojený prostredníctvom mapovania v samotných ontológiách, z čoho vyplýva, že tieto sémantické dátové štandardy SR dosahujú úroveň 5 klasifikácie otvorených dát (5★OpenData).

Príklad

Nasledovný triplet definuje prepojenie medzi data.gov.sk triedou právnická osoba patriacou do ontológie organizácia s triedou Registered Organization patriacou do ontológie REGORG.

```

<http://data.gov.sk/def/ontology/organization/CorporateBody>
rdfs:subClassOf
<http://www.w3.org/ns/regorg#RegisteredOrganization> .

```

ČASŤ D. INDIVÍDUÁ

§D1 Všeobecné URI pre indivídua

URI vzor

```
http://data.gov.sk/id/[classification]+/[identity]/{version}
```

Príklad

Nasledovné triplety definujú spoločnosť Datalan, Bratislavský samosprávny kraj, liek Concor, prezidenta Andreja Kisku, a verziu 1.0 Informačného systému pre verejné obstarávanie. Pozn. len posledný subjekt v triplete môže mať verziu.

```
<http://data.gov.sk/id/corporate-body/35810734> rdf:type org:Organization .
<http://data.gov.sk/id/nuts3/SK010> rdf:type loca:NUTS3 .
<http://data.gov.sk/id/drug/94164> rdf:type health:Drug .
<http://data.gov.sk/id/president/andrej-kiska> rdf:type per:PhysicalPerson .
<http://data.gov.sk/id/egov/isvs/5854/1.0> rdf:type egov:ISVS .
```

§D2 Definovanie URI pre fyzickú osobu

URI vzor

```
http://data.gov.sk/id/[person-classification]+/[person-id]
```

Dovolená Syntax

[person-id] = Cieľom je vytvoriť unikátne URI osoby. Do identity preto môžu vstúpiť aj niektoré technické identifikátory či hodnoty, ktoré zabezpečia unikátnosť.

Príklad

Nasledovné triplety definujú nášho aktuálneho prezidenta, autora a spoluautora tohto dokumentu.

```
<http://data.gov.sk/def/ontology/physical-person/PhysicalPerson>
  owl:oneOf
  <http://data.gov.sk/id/president/andrej-kiska>
  <http://data.gov.sk/id/creator/semanticke-standardy-2016-05-21-liska> ;
  <http://data.gov.sk/id/contributor/semanticke-standardy-2016-05-21-surek> .
```

§D3 Definovanie URI pre organizáciu

URI vzor

```
http://data.gov.sk/id/[organization-classification]+/[org-id]
```

Dovolená Syntax

[org-id] = jedinečný kód organizácie, najčastejšie IČO. V inom prípade nesmie byť org-id duplikátne k IČO.

Príklad

Nasledovné triplety definujú: Štatistický úrad, Ministerstvo Financí a Slovensko Digital.

```
<http://data.gov.sk/def/ontology/odp/Organization>
  owl:oneOf
  <http://data.gov.sk/id/corporate-body/00166197> ;
  <http://data.gov.sk/id/corporate-body/00151742> ;
  <http://data.gov.sk/id/corporate-body/00165221> ;
  <http://data.gov.sk/id/corporate-body/05621456> .
```

ČASŤ E. Dokumenty

§E1 URI pre dokumenty všeobecne

URI vzor

```
http://data.gov.sk/doc/[documenttype]/[documentname]/{version}
```

Dovolená Syntax

[documenttype] = účelom je upresnenie typu dokumentu, ktorý môže mať rôzne formáty. Presné pravidlá špecifikované nie sú, nakoľko presný typ súboru sa definuje prostredníctvom vlastnosti URI odp:mediaType.

Príklad

Nasledovné triplety definujú URI a typ tohto dokumentu, jeho názov, dátum vytvorenia, dátum platnosti a autora, jazyk a informáciu o verzii.

```
<http://data.gov.sk/doc/method/semantics-standards/2016-05-21.pdf>
  rdf:type
    <http://data.gov.sk/def/ontology/odp/Document> ;
  res:distributionOf
    <http://data.gov.sk/doc/method/semantics-standards> ;
  dct:title
    "Sémantické dátové štandardy ISVS"@sk ;
  dct:created
    "2015-11-01T13:49:55"^^xsd:date ;
  dct:modified
    "2016-05-21T20:46:55"^^xsd:date ;
  res:validFrom
    "2016-05-21T00:00:00"^^xsd:date ;
  dct:creator
    <http://data.gov.sk/id/person/semantic-standards-miroslav-liska> ;
  res:mediaType
    "application/pdf" ;
  res:language
    <http://data.gov.sk/id/language/sk> ;
  owl:versionInfo
    "2015-10-31"^^xsd:string .
```


§E2 URI pre elektronický formulár

URI vzor

```
http://data.gov.sk/doc/eform/[eformID]/{version}
```

Dovolená Syntax

[eformID] = poskytovateľ.názov_formulára.jazyk.doplňujúcainformácia , kde

- poskytovateľ sa uvádza ako IČO
- názov_formulára sa uvádza v slovenskom jazyku, čo predstavuje vecný názov formulára, pričom na oddelenie viacerých slov sa používa podtrhovník
- jazyk sa uvádza v zmysle kódu jazyka z číselníka jazykov
- doplňujúcainformácia-môže byť viacnásobná, oddelená bodkami alebo nemusí vôbec existovať

Príklad

Nasledovné triplety definujú elektronický formulár pre priznanie dani z nehnuteľnosti fyzickej osoby a jeho verziu 1.0.

```
<http://data.gov.sk/doc/eform/42499500.danove_priznanie_k_dani_z_prijmov_fyzick
ych_osob_2014.sk>
rdf:type
<http://data.gov.sk/def/ontology/egov/EForm> .

<http://data.gov.sk/doc/eform/42499500.danove_priznanie_k_dani_z_prijmov_fyzick
ych_osob_2014.sk/1.0>
res:isDistributionOf
<http://data.gov.sk/doc/eform/42499500.danove_priznanie_k_dani_z_prijmov_fyzick
ych_osob_2014.sk> .
```

ČASŤ F. DATASETY

§F1 URI pre dataset

URI vzor

```
http://data.gov.sk/set/[datasettype]/[classification]+/{distribution}
```

Dovolená Syntax =

[datasettype] =

“catalog“ ak URI reprezentuje katalóg

“codelist“ ak URI reprezentuje tzv. plochý číselník, tj. výčet prvkov z ich základnými vlastnosťami

“hierarchy“ ak URI reprezentuje hierarchiu (SKOS)

“data“ ak URI reprezentuje komplexné dáta (napr. datová tabuľka)

[classification]+ = trieda alebo koncept

[distribution] = popis verzie / distribúcie

Príklad

Nasledovné triplety definujú URI pre Číselník krajov SR, Hierarchiu Nomenklatury územných jednotiek SR pre štatistické účely, Katalóg základných číselníkov pre informačné systémy verejnej správy a dataset organizácií v systéme Regis.

```
<http://data.gov.sk/set/codelist/nuts3> rdf:type dcat:Dataset .
<http://data.gov.sk/set/hierarchy/nuts> rdf:type dcat:Dataset .
<http://data.gov.sk/set/catalog/stat-codelist> rdf:type dcat:Catalog .
<http://data.gov.sk/set/data/organization-regis> rdf:type dcat:Dataset .
```

Rozdiel medzi číselníkom a hierarchiou je v tom, že kým úlohou číselníka je urobiť jednoduchý výčet prvkov, hierarchia organizuje jednotlivé položky do systému vzájomných vzťahov. Nasledovný triplet definuje vzťah medzi Bratislavským krajom (BSK) a Okresom BA I (Staré mesto). Pozn: keďže platí, že Okres BA I nie je to isté čo BSK, je použitá netranzitívna vlastnosť skos:narrower.

```
<http://data.gov.sk/id/nuts3/SK010>
skos:narrower
<http://data.gov.sk/id/lau1/SK0101> .
```

Niekedy je hierarchia založená na dedení medzi jej jednotlivými úrovňami. Nasledovný triplet definuje túto situáciu, tj. nižšia úroveň je súčasne i vyššou. Triplet definuje že jednobytová budova (objekt tripletu) je súčasne i bytová budova (subjekt tripletu).

```
<http://data.gov.sk/def/construction-type/11>
skos:narrowerTransitive
<http://data.gov.sk/def/construction-type/111> .
```

Keďže v hierarchii existuje i nasledovný triplet, tj. jednobytová budova je budova a používame tranzitívny vzťah medzi položkami hierarchie

```
<http://data.gov.sk/def/construction-type/1>
skos:narrowerTransitive
<http://data.gov.sk/def/construction-type/111> .
```

je možné odvodiť, že jednobytová budova je budova.

```
<http://data.gov.sk/def/construction-type/1>
skos:narrowerTransitive
<http://data.gov.sk/def/construction-type/111> .
```

§F2 URI pre položku číselníka

V prípade, že je položkou číselníka konkrétne individuum (napr. Bratislavský kraj), tak je URI pre položku tvorené prostredníctvom pravidla §D1 URI pre individua. V prípade, keď sú položkami číselníka definície typov (napr. pohlavie muž), tak je URI takejto položky tvorené pravidlom §C4 URI pre definície.

§F3 URI pre distribúciu datasetu/hierarchie/katalóg

Distribúcia datasetu je jeho konkrétna verzia (najčastejšie časová). Platí, že akákoľvek zmena v distribúcii (pridanie/odstránenie atribútu alebo prvku, oprava niektorých položiek, resp. pridanie jazykovej lokalizácie...) má vždy za následok vytvorenie novej distribúcie. Inými slovami, vytvorená distribúcia musí byť nemenná tj. akákoľvek zmena musí viesť k vytvoreniu novej distribúcie. Samotná distribúcia datasetu nie je považovaná za trvácú entitu, tj. URI môže obsahovať aj referenciu na konkrétny súbor.

URI vzor

```
http://data.gov.sk/set/[datasettype]/[classification]+/{distribution}
```

Príklad

Prvý triplet predstavuje definíciu distribúcie číselníka Okresov SR platnú od 29.10.2014, druhý definuje nahradenie tejto verzie novšou z 1. januára 2015.

```
<http://data.gov.sk/set/codelist/nuts/2014-10-29.owl>
rdf:type
dcat:Distribution .

<http://data.gov.sk/set/codelist/nuts/2015-01-01.owl>
dct:replaces
<http://data.gov.sk/set/codelist/nuts/2014-10-29.owl>
```

ČASŤ H. URI referenčných údajov

§H1 Základné pravidlá použitia URI referenčných údajov

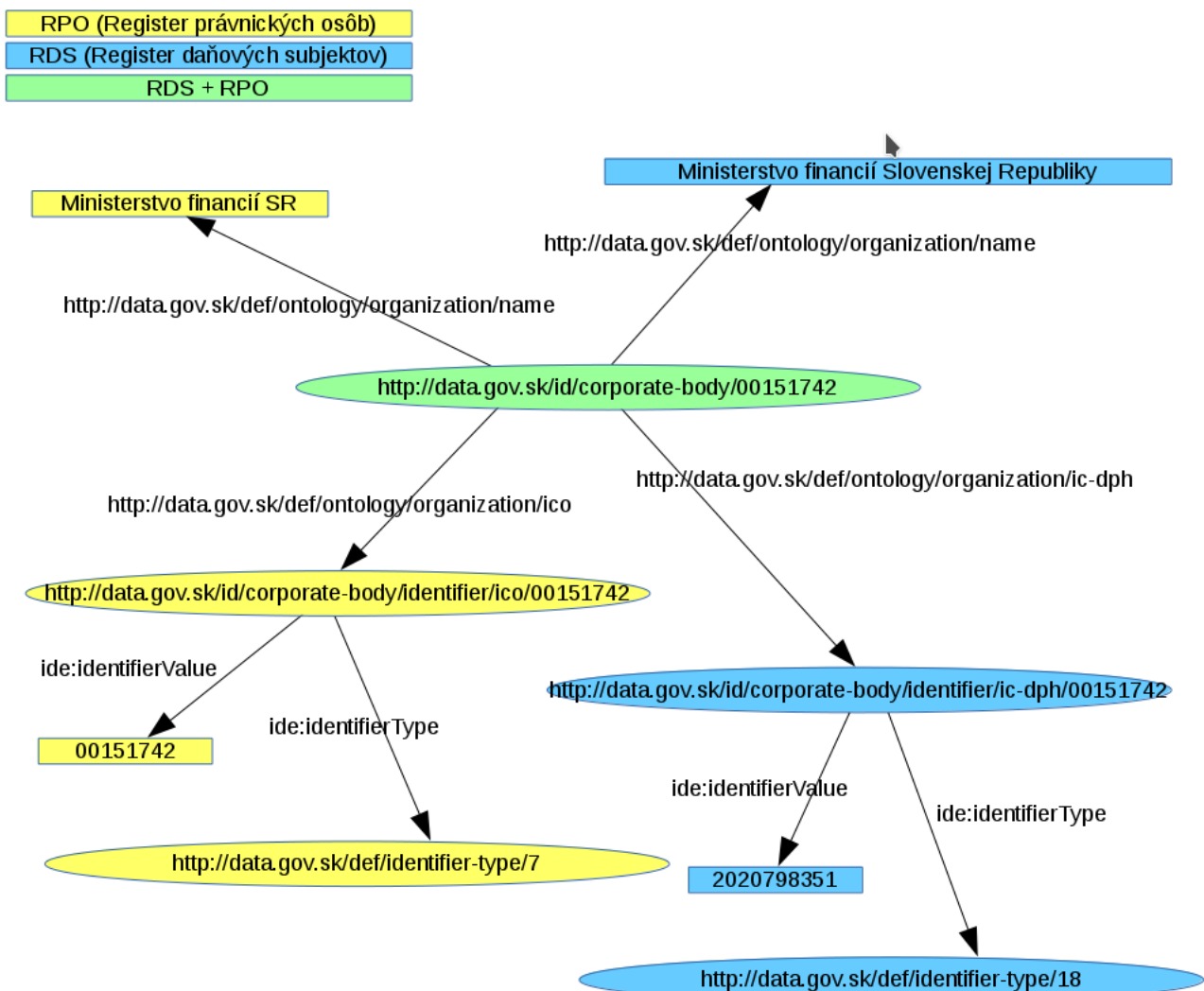
Popis

V súčasnosti sa uchováva množstvo informácií v rôznych ISVS duplicitne. Napr. obchodný register pre ministerstvo financií obsahuje názov „Ministerstvo financií SR“, kdežto Register daňových subjektov obsahuje názov „Ministerstvo financií Slovenskej republiky“. Čiže keby sa všetky informácie sústredili do jedinej databázy, nebolo by v podstate jasné, ktoré dáta sú smerodajné. A to je presne úloha referenčných údajov. Identifikovať, že globálne – z rámca nejakej databázy je daná entita referenčná, tj. smerodajná.

Formálne je referenčný údaj tvorený podľa pravidiel §C2 a §C3, tj. referenčným údajom môže byť trieda, objektová alebo dátová vlastnosť, pričom patrí do tzv. referenčného kontextu v rámci databázy, ktorá keďže obsahuje referenčné údaje, je referenčná databáza. Množina referenčných kontextov potom tvorí celkový graf platných prepojených údajov ISVS.

Príklad

Uvedený problém je ilustrovaný na nasledujúcom obrázku. Po zhraní údajov z RPO a RDS dostaneme duplicitnú vlastnosť názov organizácie, a je teda potrebné rozhodnúť, ktorá je smerodajná (referenčná) a ktorá nie.



Riešenie spočíva v definícii dátového prvku k príslušnému referenčnému registru. Nasledovný prvý triplet definuje, že názov organizácie je referenčným údajom v registry právnických osôb, a na druhej strane, identifikátor IČ-DPH je referenčným údajom v Registry daňových subjektov.

```
<http://data.gov.sk/def/ontology/organization/name>  
<http://data.gov.sk/def/ontology/resource/referenceRegister>  
<http://data.gov.sk/id/egov/isvs/420> .
```

```
<http://data.gov.sk/def/ontology/organization/ic-dph>  
<http://data.gov.sk/def/ontology/resource/referenceRegister>  
<http://data.gov.sk/id/egov/isvs/6113> .
```

Referenčné údaje majú svoju definíciu v príslušných ontológiach. Podľa týchto definícií je možné vyhodnotiť, či sú dané dáta o entite publikované správne.

Poznámka

Pojmy „referenčné údaje“ a „referencovateľné údaje“ majú odlišný význam a použitie, a je ich potrebné rozlišovať. Vo všeobecnosti platí, že všetky údaje sú referencovateľné - prostredníctvom URI (Jednotný referencovateľný identifikátor). Avšak, ako už bolo spomenuté, rôzne ISVS uchovávajú množstvo údajov duplicitne (napr. názov firmy je „udržiavaný“ vo viacerých registroch), a teda je nutné definovať, ktoré sú tie správne, tj. smerodajné. A tieto údaje sa nazývajú referenčné.

Referencie

Líška, M., Šurek, M.: *Toward Government Linked Data: A Slovak Case* . Semantics 2015. Wien, AT. [pdf available online at: <http://www.slideshare.net/semwebcompany/miroslav-liska-marek-urek-toward-government-linked-data-a-slovak-case>]

Líška, M.: *Na strastiplnej ceste k otvoreným päťhviezdičkovým údajom v Slovenskej republike*. In Petr Šaloun and Dušan Chlapek Proceedings of the DATAKON 2014: pages 40-51. Jasná, SR. [pdf available online at: <http://www.datakon.cz/?wpdmdl=458>]

Líška, M.: *Päťhviezdičkové otvorené údaje Slovenskej republiky*. Open Data Forum 2013. Bratislava, SR. [video (time=135:34) available online at: <http://www.metoo.sk/tbb?relacia=open-data>]