

Sémantické dátové štandardy pre údaje verejnej správy Slovenskej Republiky

<http://data.gov.sk/doc/pdf/semantics-standards-isvs/2015-11-02>

**- 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-2015-11-02.zip>

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

ČASŤ A. ÚVOD

§A1 Úvodné definície (Sémantický web v kocke)

§A2 Strategický zámer použitia sématickej 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 lokalizácie hodnôt vlastností 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 Zoznam podporovaných štandardizovaných ontológií

ČASŤ D. Individuá

§D1 Všeobecné URI pre individuá

§D2 URI šablóna pre individuá

§D3 URI pre fyzickú osobu

§D4 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

+ 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/odp/familyName>  
<http://data.gov.sk/def/ontology/odp/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/odp/occupation>  
<http://data.gov.sk/def/occupation/1111001> .  
  
<http://data.gov.sk/id/president/andrej-kiska>  
<http://data.gov.sk/def/ontology/odp/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/odp/givenName>  
rdfs:domain  
<http://data.gov.sk/def/ontology/odp/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/odp/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.

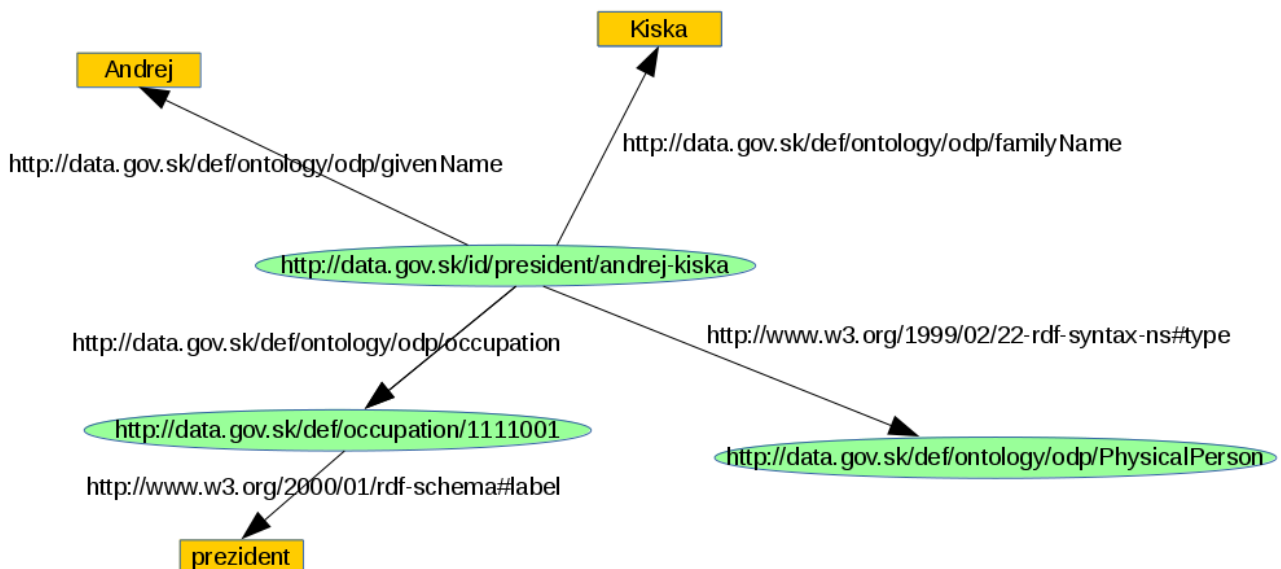
Príklad

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 vyhľadávaní, 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 vkladaní 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 aplikáciách 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 5 tripletov reprezentujúcich súčasného prezidenta SR Andreja Kisku.



Ontológia – je množina *tripletovej* ktorá opisuje vybranú doménu. V súčasnosti existuje obrovské množstvo ontológií. Napr. ontológia osoby, ontológia rastlín, ontológia odvetví sektora XY a podobne. Takmer každá dôležitá zahraničná klasifikácia niečoho existuje vo forme ontológie, ako napr. ontológia geografických prvkov, ontológia medzinárodnej klasifikácie chorôb, ontológia nomenklatury 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.

Nasledovné triplety sú výňatkami z Ontológie dátových prvkov, ktoré definujú že osoba môže byť fyzická alebo právnická.

```
<http://data.gov.sk/def/ontology/odp/PhysicalPerson>  
  rdfs:subClassOf  
<http://data.gov.sk/def/ontology/odp/Person> .  
  
<http://data.gov.sk/def/ontology/odp/Organization>  
  rdfs:subClassOf  
<http://data.gov.sk/def/ontology/odp/Person> .
```

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. triplety, 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.



§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

- ii URI musí byť čo najtrvácnejšie. Jeho reťazec nesmie obsahovať nejaký subreťazec ktorého platnosť je časovo veľmi obmedzená.
- iii URI nesmie obsahovať gestora. Napríklad URI pre typ stromu by nemal obsahovať kód ministerstva životného prostredia, pretože toto ministerstvo sa môže zlúčiť s iným, resp. gescia správy stromov prejde z jedného ministerstva na iné.
- 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 nie je ekvivalent URL. Ak je nutné identifikovať súbor, tak je ho potrebné definovať prostredníctvom URI, ktorý má vlastnosť fyzické uloženie súboru. Až hodnota tejto vlastnosti môže ukazovať na konkrétny súbor.
- vi 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 klasifikovanej jednou úrovňou, druhé a tretie dvoma. Druhé URI je príkladom verzie tohto dokumentu, pričom je upresnené že sa jedná o jeho PDF verziu. Tretie URI je príkladom URI pre fázu projektu. Keďže existujú rôzne typy projektov naprieč celým spektrom verejných dát, môžu existovať aj rôzne typy projektových fáz. Preto pre upresnenie že sa jedná o fázy EGOV projektu je použitá dvojité úroveň klasifikácie.

```
<http://data.gov.sk/id/president/andrej-kiska>
```

```
<http://data.gov.sk/doc/pdf/semantics-standards/2015-10-31>
```

```
<http://data.gov.sk/id/egov/project-phase/2>
```


V prípade použitia verzie v rámci URI, je ho potrebné samostatne definovať prostredníctvom vlastnosti verzia, aby bolo jasné, že URI obsahuje verziu. Kód reprezentujúci jazykovú, časovú alebo inú verziu je súčasťou reťazca verzia v URI. Nasledovné triplety definujú anglickú verziu tohto dokumentu.

```
<http://data.gov.sk/doc/pdf/method/semantics-standards/2015-11-02-en>
  odp:versionOf
    <http://data.gov.sk/doc/pdf/method/semantics-standards/2015-11-02>
  odp:versionInfo
    "2015-11-02-en"^^xsd:string ;
  odp:language
    <http://data.gov.sk/id/language/en> ;
  dct:created
    "2015-11-02"^^xsd:date .
```

§B3 Spôsob lokalizácie hodnôt vlastností URI

Pri lokalizácii hodnôt dátových vlastností je nutné rozlišovať tzv. jazykovo - nelokalizovateľné reťazce, ako napr. názvy osôb a podobné. V tomto prípade sa jazyková verzia reťazca neudáva, definuje sa len typ hodnoty.

Príklad

Nasledovný triplet definuje priezvisko prezidenta ako jazykovo nelokalizovateľný reťazec.

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

V inom prípade, keď je možné reťazec „preložiť“ do iného jazyka, tak sa daná dátová vlastnosť použije viackrát s upresnením jednotlivých jazykových verzií. 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 ;
```

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 dátových prvkov (ODP) a Ontológiu eGovernment entít (EGOV).

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

Príklad

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

```
<http://data.gov.sk/def/ontology/egov/ISVS> rdf:type owl:Class .  
<http://data.gov.sk/def/ontology/odp/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.

Príklad

Nasledovné triplety definujú objektovú vlastnosť adresa a dátovú vlastnosť priezvisko.

```
<http://data.gov.sk/def/ontology/odp/address> rdf:type owl:ObjectProperty .  
<http://data.gov.sk/def/ontology/odp/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/odp/address>  
<http://data.gov.sk/id/organization/address/35810734> .  
  
<http://data.gov.sk/id/person/semantic-standards-miroslav-liska>  
<http://data.gov.sk/def/ontology/odp/familyName>  
"Líška"^^xsd:string .
```

§C4 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 Odporúčané 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

ODP <http://data.gov.sk/def/ontology/odp>
EGOV <http://data.gov.sk/def/ontology/egov>
IKT <http://data.gov.sk/def/ontology/ikt>
HEALTH <http://data.gov.sk/def/ontology/health>

Vzťah medzi data.gov.sk a ostatnými ontológiami je prepojený prostredníctvom mapovania v samotných ontológiach, 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 triedov Organization patriacou do ontológie ODP s triedou Registered Organization patriacou do ontológie REGORG.

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

ČASŤ D. INDIVÍDUÁ

§D1 Všeobecné URI pre individua

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/organization/35810734> rdf:type odp:Organization .
<http://data.gov.sk/id/nuts3/SK010> rdf:type odp:NUTS3 .
<http://data.gov.sk/id/drug/94164> rdf:type health:Drug .
<http://data.gov.sk/id/president/andrej_kiska> rdf:type odp:PhysicalPerson .
<http://data.gov.sk/id/egov/isvs/5854/1.0> rdf:type egov:ISVS .
```

§D2 URI šablóna pre individua

URI vzor

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

Dovolená Syntax

[[ID-DESCRIPTION]] = kódový popis definície identity

Príklad

Keďže pre potreby registrácie URI cez MetaIS nie je efektívne registrovať každé individuum samostatne (prívelký počet), tak individua je možné registrovať len ko celú množinu individuí naraz prostredníctvom tzv. šablóny individuí. Nasledovné triplety predstavujú najskôr prvý prípad, tj. definovanie jedinej organizácie samostatne. Druhý triplet definuje šablónu pre individua všetkých organizácií naraz.

```
<http://data.gov.sk/id/organization/35810734>
rdf:type
<http://data.gov.sk/def/ontology/odp/Organization> .

<http://data.gov.sk/id/organization/[ICO]>
<http://data.gov.sk/def/ontology/odp/isIndividualURITemplateOf>
<http://data.gov.sk/def/ontology/odp/Organization>
```

§D3 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 vjsť aj niektoré technické identifikátory či hodnoty, ktoré zabezpečia unikátnosť.

Príklad

Nasledovné triplety definujú aktuálneho prezidenta SR Ing. Andreja Kisku.

```
<http://data.gov.sk/id/president/andrej_kiska>
  odp:givenName
    "Andrej"^^xsd:string ;
  odp:familyName
    "Kiska"^^xsd:string ;
  odp:occupation
    <http://data.gov.sk/def/occupation/1111001> ;
  odp:namePrefix
    <http://data.gov.sk/def/name-prefix/1> ;
  foaf:homePage
    <https://sk-sk.facebook.com/AndrejKiska/> ;
  foaf:mbox
    "kontakt@andrejkiska.sk"^^xsd:string .
```

§D4 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, Štátny ústav pre kontrolu liečiv, Datalan, Anext.

```
<http://data.gov.sk/def/ontology/odp/Organization>
  owl:oneOf
    <http://data.gov.sk/id/organization/00166197> ;
    <http://data.gov.sk/id/organization/00151742> ;
    <http://data.gov.sk/id/organization/00165221> ;
    <http://data.gov.sk/id/organization/35810734> ;
    <http://data.gov.sk/id/organization/24261122> .
```

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/pdf/method/semantics-standards/2015-10-31>
  rdf:type
    <http://data.gov.sk/def/ontology/odp/Document> ;
  odp:name
    "Sémantické dátové štandardy pre údaje verejnej správy Slovenskej
    Republiky"@sk ;
  dct:created
    "2015-11-01T13:49:55"^^xsd:dateTime ;
  dct:modified
    "2015-11-01T20:46:55"^^xsd:dateTime ;
  odp:validFrom
    "2016-01-01T00:00:00"^^xsd:dateTime ;
  dct:creator
    <http://data.gov.sk/id/person/semantic-standards-miroslav-liska> ;
  odp:mediaType
    <http://data.gov.sk/def/mediatype/application/pdf> ;
  odp:language
    <http://data.gov.sk/id/language/sk> ;
  odp:versionInfo
    "2015-10-31"^^xsd:string .
```

§E2 URI pre elektronický formulár

URI vzor

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

Príklad

Nasledovné triplety definujú elektronický formulár pre priznanie dani z nehnuteľnosti fyzickej osoby s jeho menným priestorom.

```
<http://data.gov.sk/doc/eform/00308307.A0000224.000000027.PriznanieDzN_FO/1.0>
  rdf:type
    <http://data.gov.sk/def/ontology/egov/EForm> ;
  odp:namespace
    <http://data.gov.sk/doc/namespace/00308307.A0000224.000000027.PriznanieDzN_FO/1.0> .
```

ČASŤ F. DATASETY

§F1 URI pre číselník/hierarchiu/katalóg

URI vzor

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

Dovolená Syntax =

[datasettype] =

“cat“ ak URI reprezentuje katalóg

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

“hierarchy“ ak URI reprezentuje hierarchiu (SKOS)

[classification]+ = trieda alebo koncept

[distribution] = popis verzie / distribúcie

Príklad

Nasledovné triplety definujú URI pre Číselník krajov SR, Hierarchiu Nomenklatúry územných jednotiek SR pre štatistické účely a Katalóg základných číselníkov pre informačné systémy verejnej správy.

```
<http://data.gov.sk/set/codelist/nuts3> rdf:type dcat:Dataset .  
<http://data.gov.sk/set/hierarchy/nuts> rdf:type skos:ConceptScheme .  
<http://data.gov.sk/set/cat/base-codelist-isvs> rdf:type dcat:Catalog .
```

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/11> .
```

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, tak je URI pre položku tvorené prostredníctvom pravidla §D1 URI pre individua. V prípade, keď sú položkami číselníka definície typov, tak je URI takejto položky tvorené nasledovným pravidlom:

URI vzor

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

Príklad

Nasledovný triplet definuje položku číselníka Pohlavie - Muž. V tomto prípade sa použije prístup OWL Full, tj. individuum Muž je typu Pohlavie a súčasne je i koncept.

```
<http://data.gov.sk/def/sex/1> rdf:type odp:Sex .  
<http://data.gov.sk/def/sex/1> rdf:type skos:Concept .
```

§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 distribúcii.

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


Dokumentácia

[20151102-data.gov.sk-semanticke-standardy.odt](#)

[20151102-data.gov.sk-semanticke-standardy.pdf](#)

[20151102-data.gov.sk-uri-identifikatory.ods](#)

- zoznamy navrhnutých Jednotných referencovateľných identifikátorov (URI)
 - číselníky a hierarchie publikované na statistics.sk
 - mapovanie Katalógu dátových prvkov (KDP) na Ontológiu dátových prvkov (ODP)
 - URI pre MetaIS entity
 - vybrané medicínske číselníky
 - vybrané eformuláre

Príklady OWL/XML

Ontológie

[examples/def/ontology/odp.owl](#)

[examples/def/ontology/egov.owl](#)

Datasety

Katalógy

[examples/set/cat/base-codelist-isvs.owl](#)

[examples/set/cat/codelist .owl](#)

[examples/set/cat/egov.owl](#)

[examples/set/cat/hierarchy.owl](#)

Číselníky (distribúcie)

[examples/set/codelist/base-ontology/2015-10-26.owl](#)

[examples/set/codelist/existential-status/2008-10-13.owl](#)

[examples/set/codelist/filiation-type/2008-10-13.owl](#)

[examples/set/codelist/identifier-type/2014-03-03.owl](#)

[examples/set/codelist/isvs/2015-08-05.owl](#)

[examples/set/codelist/lau1/2014-03-11.owl](#)

[examples/set/codelist/lau2/2014-09-22.owl](#)

[examples/set/codelist/marrital-status/2012-10-05.owl](#)

[examples/set/codelist/medical-service/2014-03-03.owl](#)

[examples/set/codelist/name-prefix/2013-06-28.owl](#)

[examples/set/codelist/name-suffix/2012-05-14.owl](#)

[examples/set/codelist/nuts2/2014-09-22.owl](#)

[examples/set/codelist/nuts3/2014-09-22.owl](#)

[examples/set/codelist/organization-legal-type/2015-05-10.owl](#)

[examples/set/codelist/phone-number-type/2008-10-13.owl](#)

[examples/set/codelist/population-density-type/2009-01-01.owl](#)

[examples/set/codelist/sex/2009-09-13.owl](#)

[examples/set/codelist/un-country/2012-12-05.owl](#)

Hierarchie (distribúcie)

[examples/set/hierarchy/nuts/2007-10-25.owl](#)

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>]