

## A.1.1 Úvodné definície (Sémantický web v kocke)

**URI** (Uniform Resource Identifier) – je reťazec znakov používaných na identifikáciu entity v sémantickom webe v tvare `https://...` 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, riek, 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.

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

**Individuum** – je objekt reálneho sveta, v sémantickom webe reprezentované ako URI. V predchádzajúcom príklade je len prvé URI individuum.

**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.

```
<https://data.gov.sk/id/president/andrej-kiska> <https://data.gov.sk/def/ontology/physical-person/occupation-
type> <https://data.gov.sk/def/occupation-type/1111001> .
<https://data.gov.sk/id/president/andrej-kiska> <https://data.gov.sk/def/ontology/physical-person/familyName>
"Kiska"^^xsd:string .
```

**Odvođený triplet** – je triplet, ktorý vznikol strojovým odvođením z existujúcich tripletov v znalostnej báze na základe definovaného pravidla.

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

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

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

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

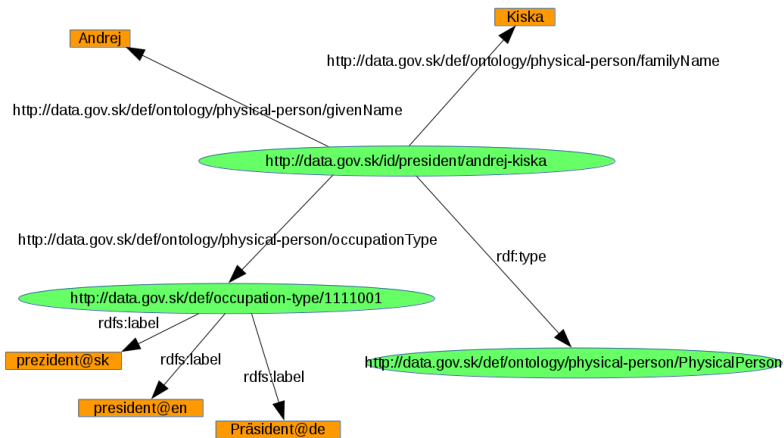
**Odvodzovanie** – Strojové odvodzovanie (machine reasoning) je disciplína umelej inteligencie, ktorá sa zaoberá odvodzovaním nových znalostí (tripletov) z existujúcich. Inými slovami, stroj sa snaží odvodiť, aké nové triplety vie odvodiť z existujúcich.

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 je 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 reprezentuje RDF graf obsahujúci 7 tripletov reprezentujúcich súčasného prezidenta SR Andreja Kisku.



**Ontológia** – je množina *tripletov* ktorá opisuje vybranú doménu. V súčasnosti samozrejme existuje obrovské množstvo ontológií, a 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. sémantickú ontológiu.

**Asset** - verzionovateľná entita

**Dataset** – verzionovateľná množina údajov definovaných v [A.4.5.3 Pravidlá pre publikáciu katalógu, datasetu a distribúcie](#).

**Distribúcia** – je reprezentácia datasetu.

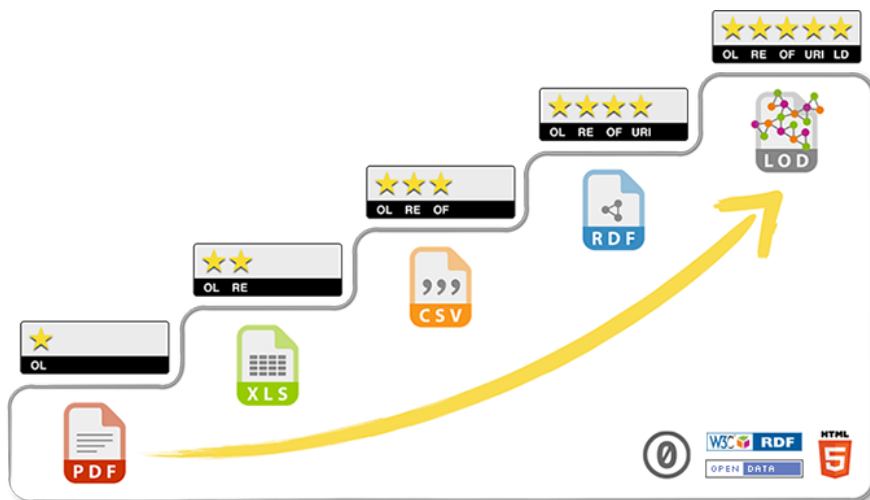
**Katalóg** - množina datasetov.

**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.

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



#### 4 vs 5 dáta

Rozdiel medzi 4 a 5 hviezdikovými dátami :

##### 4 hviezdíčky

- použitie referencovateľných identifikátorov na identifikáciu entít
- použitie vlastných ontológií na dátový model
- nutnosť mapovania ontológií v najväčšej možnej miere na 5 hviezdíček na zachovanie čo najväčšej interoperability
- dereferenciácia referencovateľných identifikátorov je voľiteľná
- zverejnenie vlastných ontológií je odporúčané
- zaregistrovanie šablón referencovateľných identifikátorov na MetaIS je voľiteľné
- pri registrácii je nutné vždy požiadať Centrálnu dátovú kanceláriu na UPPVII o zaradenie dát medzi 5 hviezdíkové. Len v prípade odmietnutia sa stávajú 4 hviezdíkové.
- negarantujú dátovú interoperabilitu medzi systémami nakoľko nie sú súčasťou centrálného modelu údajov
- referencovateľné identifikátory musia byť tvorené metodikou, ktorá je súčasťou [Sémantické dátové štandardy ISVS](#)

##### 5 hviezdíček

Všetko zo 4 hviezdíček, ale :

- reprezentujú centrálny model verejnej správy prostredníctvom ontológií
- sú používané jednotné referencovateľné identifikátory
- model údajov je pod prísnou správou pracovnej skupiny PS1 a zaručuje úplnú interoperabilitu údajov v prostredí verejnej správy
- doména všetkých URI je <https://data.gov.sk>
- dereferenciácia je zabezpečená integráciou MetaIS a portálu data.gov.sk

Dôležitým elementom je, že každá entita musí mať ambíciu stať sa 5 hviezdíkovou entitou. Pracovná skupina PS1 bude dozerať a schvalovať žiadosti o priradenie entity do centrálného modelu údajov.

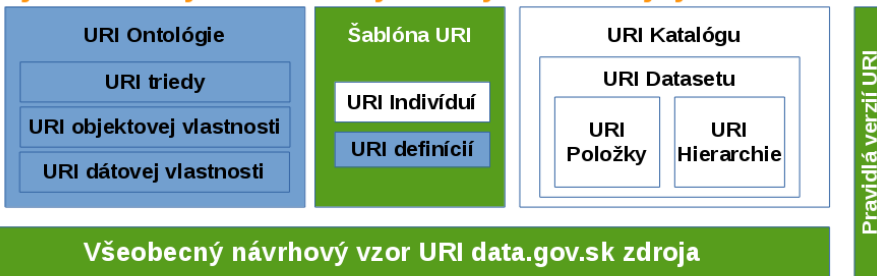
**Jednotný referencovateľný identifikátor** - vyjadruje URI identifikátor, ktorý unikátne identifikuje entitu v prostredí ISVS a ktorý je záväzný používať pre každú entitu, ktorá takýto identifikátor má pridelený.

**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.

Individuá  
Znalosti  
všeobecné

data.gov.sk-semanticweb

## jednoduchý /rozšíriteľný URI systém verejných dát SR



URI