



Rozvoj a používanie DCAT-AP-SK pri publikácii a spracovaní Linked Open Data

NGHIA PHAM VAN

Problematika

- ▶ Nízka kvalita verejných dát
- ▶ Nekonzistencia
- ▶ Duplicitné identifikátory
- ▶ Chýba špecifikácia, štandard

DCAT – Data Catalog Vocabulary

- ▶ RDF slovník pre výmenu informácií medzi dátovými katalógmi
- ▶ Využíva sa na opis dátových katalógov a datasetov
- ▶ Definuje 3 hlavné triedy:
 - ▶ dcat:Catalog
 - ▶ dcat:Dataset
 - ▶ dcat:Distribution

DCAT Profil

- ▶ Špecifikácia pre dátové katalógy
- ▶ Dopĺňa dodatočné obmedzenia ako napríklad
 - ▶ Skupina minimálnych povinných metadát
 - ▶ Nové triedy a vlastnosti ktoré nie sú súčasťou DCAT

DCAT-AP

- ▶ Špecifikácia založená na DCAT
- ▶ Slúži na opis verejného sektoru v Európe
- ▶ Definuje triedy a vlastnosti na opis dátových katalógov
- ▶ Cieľom je zabezpečiť konzistenciu
- ▶ Je možné ho rozšíriť

DCAT-AP-SK

- ▶ Rozšírenie DCAT-AP pre Slovenské verejné dáta
- ▶ Opis dátových katalógov, datasetov a distribúcií
- ▶ Definuje povinné, odporúčané a nepovinné triedy a vlastnosti
- ▶ Reprezentuje špecifikáciu pri publikovaní dátových katalógov

DCAT-AP-SK Príklad

```
<!-- dcat:Catalog -->
<rdf:Description rdf:about="http://www.w3.org/ns/dcat#Catalog">
  <vann:usageNote xml:lang="sk">Povinná</vann:usageNote>
  <dct:identifier xml:lang="sk">dcat:Catalog</dct:identifier>
  <rdfs:label xml:lang="sk">Dátový katalóg</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="http://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#class-catalog"/>
  <rdfs:comment xml:lang="sk">Zbierka metadát o datasetoch.</rdfs:comment>
</rdf:Description>
```

```
<!-- Data Catalogue - dcat:dataset -->
<rdf:Description rdf:about="http://www.w3.org/ns/dcat#dataset">
  <vann:usageNote xml:lang="sk">Povinná vlastnosť. Táto vlastnosť spája Katalóg s Datasetom a je súčasťou datasetu.</vann:usageNote>
  <dct:identifier xml:lang="sk">dcat:dataset</dct:identifier>
  <rdfs:label xml:lang="sk">Dátový katalóg - dataset</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="https://www.w3.org/TR/2013/WD-vocab-dcat-20130312/#Class:_Dataset"/>
  <rdfs:comment xml:lang="sk">Dataset ktorý je súčasťou katalógu.</rdfs:comment>
</rdf:Description>

<!-- Data Catalogue - dct:description -->
<rdf:Description rdf:about="http://purl.org/dc/terms/description">
  <vann:usageNote xml:lang="sk">Povinná vlastnosť. Táto vlastnosť obsahuje textový opis Katalógu.
  Táto vlastnosť môže byť zopakovaná pre opisy v rôznych jazykoch.</vann:usageNote>
  <dct:identifier xml:lang="sk">dct:description</dct:identifier>
  <rdfs:label xml:lang="sk">Dátový katalóg - opis</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="http://dublincore.org/documents/dcmi-terms/#terms-description" />
  <rdfs:comment xml:lang="sk">Opisom môžu byť napríklad: abstrakt, obsah, grafická reprezentácia alebo textový opis zdroja.</rdfs:comment>
</rdf:Description>
```

DCAT-AP-SK moduly

- ▶ Pretože DCAT-AP-SK opisuje iba samotné dátové katalógy, datasety a distribúcie, navrhujeme vytvoriť doplnkové moduly pre domény/samotné dáta
- ▶ Definuje triedy a vlastnosti na opis domény
- ▶ Napríklad DCAT-AP-SK-Library pre knihy alebo DCAT-AP-SK-OR pre obchodný register

DCAT-AP-SK-OR Príklad

```
<!-- Legal Subject -->
<rdf:Description rdf:about="https://data.gov.sk/def/ontology/legal-subject/LegalSubject">
  <vann:usageNote xml:lang="en">Mandatory</vann:usageNote>
  <vann:usageNote xml:lang="sk">Povinná</vann:usageNote>
  <dct:identifier xml:lang="en">lsub:LegalSubject</dct:identifier>
  <dct:identifier xml:lang="sk">lsub:LegalSubject</dct:identifier>
  <rdfs:label xml:lang="en">Legal Subject</rdfs:label>
  <rdfs:label xml:lang="sk">Právny Subjekt</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/legal-subject/2017-01-01.owl"/>
  <rdfs:comment xml:lang="en">A complex datatype describing a legal subject or an entity,
    which is neither a legal subject nor a physical person according to the law of Slovakia.</rdfs:comment>
  <rdfs:comment xml:lang="sk">Zložený dátový prvok pre popis právnickej osoby alebo subjektu,
    ktorý nie je právnickou osobou ani fyzickou osobou podľa právneho poriadku Slovenskej republiky.</rdfs:comment>
</rdf:Description>
```

DCAT-AP-SK-OR Príklad

10








```
<!-- Legal Subject - Name -->
<rdf:Description rdf:about="https://data.gov.sk/def/ontology/legal-subject/name">
  <vann:usageNote xml:lang="en">Mandatory</vann:usageNote>
  <vann:usageNote xml:lang="sk">Povinná</vann:usageNote>
  <dct:identifier xml:lang="en">lsub:name</dct:identifier>
  <dct:identifier xml:lang="sk">lsub:name</dct:identifier>
  <rdfs:label xml:lang="en">Legal subject full name</rdfs:label>
  <rdfs:label xml:lang="sk">Názov právneho subjektu</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/legal-subject/2017-01-01.owl"/>
  <rdfs:comment xml:lang="en">The full name of the legal subject.</rdfs:comment>
  <rdfs:comment xml:lang="sk">Uvádza sa plný názov právnickej osoby.</rdfs:comment>
</rdf:Description>
```

```
<!-- Legal Subject - Established -->
<rdf:Description rdf:about="https://data.gov.sk/def/ontology/legal-subject/established">
  <vann:usageNote xml:lang="en">Mandatory</vann:usageNote>
  <vann:usageNote xml:lang="sk">Povinná</vann:usageNote>
  <dct:identifier xml:lang="en">lsub:established</dct:identifier>
  <dct:identifier xml:lang="sk">lsub:established</dct:identifier>
  <rdfs:label xml:lang="en">Established</rdfs:label>
  <rdfs:label xml:lang="sk">Vznik</rdfs:label>
  <rdfs:isDefinedBy rdf:resource="https://data.gov.sk/def/ontology/legal-subject/2017-01-01.owl"/>
  <rdfs:comment xml:lang="en">The date on which the legal subject was established.</rdfs:comment>
  <rdfs:comment xml:lang="sk">Uvádza sa dátum zriadenia právnickej osoby.</rdfs:comment>
</rdf:Description>
```

```
<!-- Legal Subject - Legal Form -->
```

Validácia

11

Class Name	Rule ID	Rule Severity	Rule Description	Message
<input type="text" value="Search Class Name"/>	<input type="text" value="Search Rule ID"/>	<input type="text" value="Search Rule Severity"/>	<input type="text" value="Search Rule Description"/>	<input type="text" value="Search Message"/>
Catalog	112	 error	dcat:Catalog does not exist.	The mandatory class dcat:Catalog does not exist.
Agent	113	 error	foaf:Agent does not exist.	The mandatory class foaf:Agent does not exist.
Dataset	114	 error	dcat:Dataset does not exist.	The mandatory class dcat:Dataset does not exist.
Distribution	163	 warning	dcat:Distribution does not exist.	The recommended class dcat:Distribution does not exist.
LicenseDocument	166	 warning	dct:LicenseDocument does not exist.	The recommended class dct:LicenseDocument does not exist.
Category	167	 warning	skos:Concept does not exist.	The recommended class skos:Concept does not exist.
Category scheme	168	 warning	skos:ConceptScheme does not exist.	The recommended class skos:ConceptScheme does not exist.

Show 10 entries

Budúca práca

- ▶ Používanie Reasonera na validáciu datasetov
- ▶ Spojenie s pravidlami ontológií
- ▶ Rozšírenie DCAT-AP-SK
- ▶ Vytvorenie modulov pre iné domény

