

Predstavitev metapodatki

Open Data Support podpira Evropska komisija v okviru projekta SMART 2012/0107 'Lot 2: Zagotavljanje storitev objave, dostopa in vnovične uporabe javnih podatkov v Evropski uniji preko obstoječih portalov odprtih podatkov (Pogodba št. 30-CE-0530965/00-17).

© 2014 Evropska komisija



OPEN DATA SUPPORT

Training Module 1.4

Uvod v upravljanje z metapodatki

Učni cilji

Cilj tega modula je pojasniti:

- kaj so metapodatki,
- terminologijo in cilje upravljanja z metapodatki,
- različne razsežnosti kakovosti metapodatkov,
- uporabo kontroliranih slovarjev za metapodatke;
- izmenjavo in zbiranje metapodatkov,
- upravljanje z metapodatki v Open Data Support.

Vsebina

Ta modul vsebuje ...

razlago o metapodatkih,

oris življenjskega cikla metapodatkov,

- uvod v kakovost metapodatkov,
- pregled upravljanja z metapodatki in pristop izmenjave, ki se izvaja z uporabo Open Data Support preko platforme Open Data Interoperability Platform.

Kaj so metapodatki?

Opredefinitev, primeri in standardi vnovične uporabe.

Kaj so metapodatki?

“Metapodatki so strukturirane informacije, ki opisujejo, pojasnjujejo, opredeljujejo ali kako drugače omogočajo lažje pridobivanje, uporabo ali upravljanje z viri informacij. Metapodatkom pogosto pravimo podatki o podatkih ali informacije o informacijah.”

-- National Information Standards Organization

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Metapodatki zagotavljajo informacije, ki nam pomagajo osmisлити **podatke** (npr. dokumente, slike, nabore podatkov), **koncepte** (npr. klasifikacijske sheme) in **subjekte iz realnega sveta** (npr. ljudi, organizacije, kraje, slike, produkte).

Vrste metapodatkov

- **Opisni (deskriptivni) metapodatki** opisujejo vir z namenom odkrivanja in prepoznavanja.
- **Strukturni metapodatki**, npr. podatkovni modeli in referenčni podatki.
- **Administrativni metapodatki** so informacije, ki pomagajo upravljati z virom.

To predavanje v glavnem obravnava deskriptivne metapodatke za nabore podatkov.

Delno so opisani tudi administrativni metapodatki.

Primeri metapodatkov

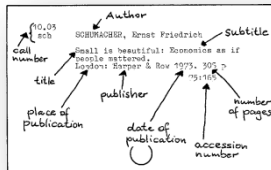
nalepka



Vsebuje metapodatke



kataloška kartica



knjigi



opis nabora podatkov (DCAT)

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;
dct:description "Data from seven weather stations
showing temperature, humidity,
wind direction and wind speed" ;
dct:modified "2013-07-01" ;
dct:publisher <http://myweather.com/id/myweather> ;
dcat:keyword "weather" ;
dcat:landingpage <http://myweather.com/stations1-7.html> ;
dcat:distribution :weatherdata1-xlsx
.

:weatherdata1-7-xlsx a dcat:Distribution ;
dct:format <http://publications.europa.eu/resource/authority/file-type/XLSX> ;
dct:licence <http://creativecommons.org/licenses/by/4.0/> ;
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```

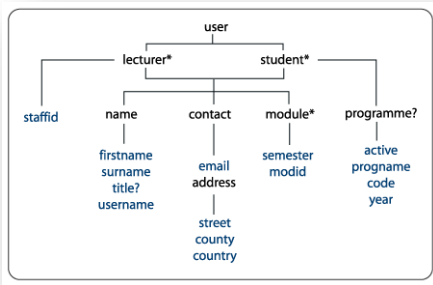


naboru podatkov

	Temp. °C	Humidity %	Wind direction	Wind speed km/h
Station 1	18.1	60	WSW	18
Station 2	17.5	59	WSW	20
Station 3	18.2	55	SW	22
Station 4	19.0	62	SW	18
Station 5	18.0	65	WSW	19
Station 6	18.2	63	SSW	21
Station 7	17.9	61	SW	22

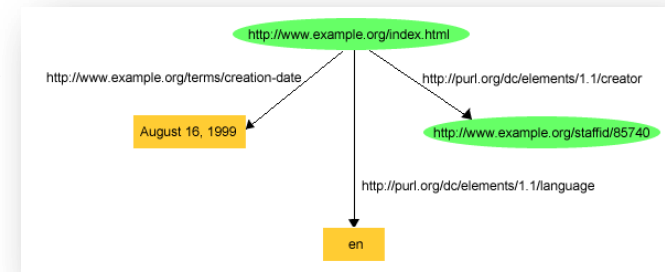
Dva pristopa zagotavljanja metapodatkov na spletu

XML (Tree/container approach)



```
<?xml version="1.0"?>
<!DOCTYPE user SYSTEM "users.dtd">
<user>
  <student>
    <name>
      <firstname>Joe</firstname>
      <surname>Smith</surname>
      <title>Mr.</title>
      <username>smithj</username>
    </name>
    <contact>
      <address>
        <street>54 Maple Rise, Santry</street>
        <country>Dublin</country>
        <country>Ireland</country>
      </address>
      <email>smithj@dcu.ie</email>
    </contact>
    <programme active="true">
      <programe>M.Eng in Electronic Systems</programe>
      <code>MEN</code>
      <year>1</year>
    </programme>
    <module semester="2">
      <modid>EE557</modid>
    </module>
    <module semester="1">
      <modid>EE553</modid>
    </module>
  </student>
```

RDF (Triple-based approach)



```
ex:index.html dc:creator exstaff:85740 .
ex:index.html exterm:creation-date "August 16, 1999" .
ex:index.html dc:language "en" .
```


Upravljanje z metapodatki vaših naborov podatkov

Upravljanje z metapodatki je pomembno

Metapodatke je treba upravljati zato, da se zagotovijo ...

- **razpoložljivost:** metapodatke je treba shraniti na mestu, kjer je do njih možno dostopati in jih indeksirati tako, da jih je mogoče najti;
- **kakovost:** metapodatki morajo biti konsistentne kakovosti, tako jim uporabniki lahko zaupajo;
- **trajnost:** metapodatke je treba hraniti nekaj časa;
- **odprte licence:** metapodatki morajo biti na razpolago pod licenco javne domene, da se omogoči njihove vnovična uporaba.

*Življenjski cikel metapodatkov je **obsežnejši** od življenjskega cikla podatkov:*

- Metapodatke je treba **ustvariti že preden se ustvarijo ali zajamejo podatki**, da se zagotovi informacija, ki bo na voljo v prihodnosti.
- Metapodatke je treba ohraniti tudi **po odstranitvi podatkov**, npr. da se zagotovi informacija o podatkih, ki so bili odstranjeni.

Metapodatkovna shema

“Poimenovalni, označevalni ali kodirni sistem, ki se uporablja za katalogizacijo informacij ali strukturiranje deskriptivnih evidenc. Metapodatkovna shema ustvarja ali opredeljuje podatkovne elemente in pravila, ki urejajo uporabo podatkovnih elementov za opis vira.”

XML
shema

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE RDF>
- <rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rd="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcam="http://purl.org/dc/dcam/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
  - <rdf:Description rdf:about="http://purl.org/dc/terms/">
    <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
    <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/title">
    <rdfs:label xml:lang="en">Title</rdfs:label>
    <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  - <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title-002"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
    <rdfs:label xml:lang="en">Creator</rdfs:label>
    <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
    <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator-002"/>
    <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
    <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
    <rdfs:label xml:lang="en">Subject</rdfs:label>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE RDF>
- <rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rd="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcam="http://purl.org/dc/dcam/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
  - <rdf:Description rdf:about="http://purl.org/dc/terms/">
    <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
    <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/title">
    <rdfs:label xml:lang="en">Title</rdfs:label>
    <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  - <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title-002"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
    <rdfs:label xml:lang="en">Creator</rdfs:label>
    <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
    <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
    <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/">
    <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-10-11</dcterms:modified>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
    <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator-002"/>
    <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
    <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
    <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
  - <rdf:Description>
  - <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
    <rdfs:label xml:lang="en">Subject</rdfs:label>
```

RDF
shema

Vnovična uporaba slovarjev za zagotavljanje metapodatkov do vaših virov

Splošni standardi in specifikacije:

- **Dublin Core** za objavljena gradiva (besedila, slike):
<http://dublincore.org/documents/dcmi-terms/>
- **FOAF** za osebe in organizacije: <http://xmlns.com/foaf/spec/>
- **SKOS** za konceptualne zbirke: <http://www.w3.org/TR/skos-reference>
- **ADMS** za interoperabilna sredstva: <http://www.w3.org/TR/vocab-adms/>

Specifični standardi za nabore podatkov:

- **Data Catalog Vocabulary DCAT**, <http://www.w3.org/TR/vocab-dcat/>

Specifična uporaba DCAT in drugih slovarjev za podporo interoperabilnosti podatkovnih portalov v Evropi:

- **DCAT profil aplikacij za podatkovne portale v Evropi:**
http://joinup.ec.europa.eu/asset/dcat_application_profile/description

Oblikovanje metapodatkovne sheme s shemo RDF (RDFS) – vnovična uporaba, če je le mogoče

Shema RDF je zlasti dobra za kombiniranje izrazov iz različnih standardov in specifikacij.

Kadar oblikujete shemo RDF, **ne izumljajte izrazov**, ki so že bili opredeljeni drugje – izraze uporabite **vnovič**, če je le mogoče.

- Npr. aplikacija profilov DCAT za podatkovne portale v Evropi (DCAT-AP) uporablja izraze iz DCAT, Dublin Core, FOAF, SKOS, ADMS in drugih standardov.

7.3. Dataset

7.3.1. Mandatory properties for Dataset

Property	URI	Range	Usage note	Card
description	dct:description	rdfs:Literal	This property contains a free-text account of the dataset. This property can be repeated for parallel language versions of the description.	1..n
publisher	dct:publisher	foaf:Organization	This property refers to an organisation responsible for making the dataset available.	1..1
title	dct:title	rdfs:Literal	This property contains a name given to the dataset. This property can be repeated for parallel language versions of the name.	1..n

7.3.2. Recommended properties for Dataset

Property	URI	Range	Usage note	Card
contact point	adms:contactPoint	v:VCard	This property contains contact information that can be used for flagging errors in the dataset or sending comments	0..n
dataset distribution	dcat:distribution	dcat:Distribution	This property links the dataset to an available distribution.	0..n
theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the dataset. A dataset can have multiple themes.	0..n

7.3.3. Optional properties for Dataset

Property	URI	Range	Usage note	Card
frequency	dct:accrualPeriodicity	dct:Frequency	This property refers to the frequency at which dataset is published.	0..1
identifier	dct:identifier	rdfs:Literal	This property contains the main identifier for the dataset, e.g. the URI or other unique identifier in the context of the Catalog	0..n
keyword/ tag	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the dataset.	0..n

Primer: opis odprtega podatkovnega niza z DCAT-AP

opis kataloga

```
:catalog
  a dcat:Catalog ;
  dct:title "Imaginary Catalog" ;
  rdfs:label "Imaginary Catalog" ;
  foaf:homepage <http://example.org/catalog> ;
  dct:publisher :transparency-office ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dcat:dataset :dataset-001 , :dataset-002 , :dataset-003 ;
  .
```

opis podatkov

```
:dataset-001
  a dcat:Dataset ;
  dct:title "Imaginary dataset" ;
  dcat:keyword "accountability","transparency" ,"payments" ;
  dct:issued "2011-12-05"^^xsd:date ;
  dct:modified "2011-12-05"^^xsd:date ;
  dct:publisher :finance-ministry ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dcat:distribution :dataset-001-csv ;
  .
```

opis distribucije

```
:dataset-001-csv
  a dcat:Distribution ;
  dcat:downloadURL <http://www.example.org/files/001.csv> ;
  dct:title "CSV distribution of imaginary dataset 001" ;
  dcat:mediaType "text/csv" ;
  dcat:byteSize "5120"^^xsd:decimal ;
  .
```

Kontrolirani slovarji

Uporaba tezavrov, taksonomij in standardiziranih seznamov izrazov za določanje vrednosti podatkovnih lastnosti.

Kaj so kontrolirani slovarji?

Kontrolirani slovar je vnaprej določen seznam vrednosti, ki se uporabijo kot vrednosti za specifično lastnost vaše metapodatkovne sheme.

- Poleg pazljivega oblikovanja shem so za izmenjavo informacij in interoperabilnost pomembni vrednostni prostori lastnosti metapodatkov.
- Zaradi skupnih kontroliranih slovarjev za vrednostne prostore so metapodatki razumljivi v vseh sistemih.

Kateri kontrolirani slovar naj se uporabi za posamezno vrsto lastnosti?

- Uporabljajte **sezname kod, ki so kontrolirani slovar za prosto besedilo ali lastnosti v nizu.**

primer lastnosti DCAT-AP:

keyword/ tag	dcat:keyword	rdfs:Literal	This property contains a keyword or tag describing the dataset.
--------------	--------------	--------------	---

- Uporabljajte **koncepte, ki jih URI prepoznava kot reference na "stvari"**

- Primer lastnosti DCAT-AP:

theme/ category	dcat:theme, subproperty of dct:subject	skos:Concept	This property refers to a category of the dataset.
--------------------	--	--------------	--

Vzorčni seznam kod -
ObjectInCrimeClass (ListPoint)

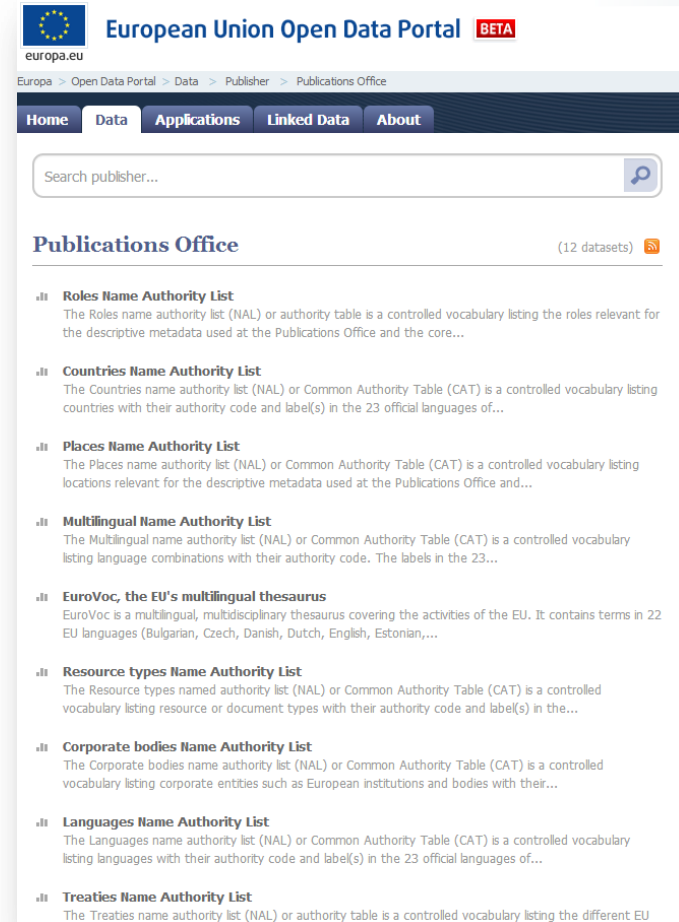
Code	Meaning
1	Used
2	Taken Without Consent
3	Eliminated
4	Not Involved
5	Involved
6	Crime Weapon
7	Handled / Received
11	Exhibit

- Vzorčna taksonomija z izrazi, ki imajo URI - EuroVoc

```
<rdf:Description rdf:about="http://eurovoc.europa.eu/300">
  <xl:altLabel rdf:resource="http://eurovoc.europa.eu/415040"/>
  <s04:prefLabel xml:lang="da">international kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="sv">internationell kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="en">international credit</s04:prefLabel>
  <s04:prefLabel xml:lang="de">internationaler Kredit</s04:prefLabel>
  <s04:prefLabel xml:lang="nl">internationaal krediet</s04:prefLabel>
</rdf:Description>
```

Primer – Seznami poimenovanih organov Urada za publikacije

- Ti sezname imajo vnovič uporabljive kontrolirane slovarje za:
 - države
 - družbe
 - vrste datotek
 - medinstitucionalne postopke
 - jezike
 - večjezičnost
 - vrste virov
 - vloge
 - mednarodne pogodbe



The screenshot shows the European Union Open Data Portal interface. At the top, there is the European Union flag and the text "European Union Open Data Portal BETA" with the URL "europa.eu". Below this is a navigation bar with "Home", "Data", "Applications", "Linked Data", and "About". A search bar is present with the placeholder text "Search publisher...". The main content area is titled "Publications Office" and indicates "(12 datasets)". A list of authority lists is displayed, each with a brief description:

- Roles Name Authority List**: The Roles name authority list (NAL) or authority table is a controlled vocabulary listing the roles relevant for the descriptive metadata used at the Publications Office and the core...
- Countries Name Authority List**: The Countries name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing countries with their authority code and label(s) in the 23 official languages of...
- Places Name Authority List**: The Places name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing locations relevant for the descriptive metadata used at the Publications Office and...
- Multilingual Name Authority List**: The Multilingual name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing language combinations with their authority code. The labels in the 23...
- EuroVoc, the EU's multilingual thesaurus**: EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 22 EU languages (Bulgarian, Czech, Danish, Dutch, English, Estonian,...
- Resource types Name Authority List**: The Resource types named authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing resource or document types with their authority code and label(s) in the...
- Corporate bodies Name Authority List**: The Corporate bodies name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing corporate entities such as European institutions and bodies with their...
- Languages Name Authority List**: The Languages name authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing languages with their authority code and label(s) in the 23 official languages of...
- Treaties Name Authority List**: The Treaties name authority list (NAL) or authority table is a controlled vocabulary listing the different EU

Življenjski cikel metapodatkov

Ustvarjanje, vzdrževanje, posodabljanje, shranjevanje, objava metapodatkov in izbris podatkov.

Ustvarjanje vaših metapodatkov

Ustvarjanje metapodatkov je lahko podprto s (pol)avtomatskimi procesi.
pr

- Lastnosti dokumentov, ki se ustvarijo z (pisarniški) orodji, npr. datum nastanka.
- Informacija o kraju in času, ki jo zabeleži kamera, senzor ...
- Informacija iz procesa objave, npr. lokacija datoteke ali URL.

Vendar pa druge lastnosti zahtevajo posredovanje človeka:

- O čem govori vir (npr. povezovanje do predmetnega slovarja)?
- Kako se vir lahko uporabi (npr. povezovanje z licenco)?
- Kje lahko najdem več informacij o tem viru (npr. povezovanje s stranjo na spletu ali dokumentacijo, ki opisuje vir)?

Vzdrževanje vaših metapodatkov

Pristopi k vzdrževanju metapodatkov morajo biti ustrezni glede na tip podatkov za objavo.

- Če se **podatek** ne **spreminja**, so **metapodatki** dokaj **stabilni**. Spremembe (masovne spremembe) se lahko po potrebi opravijo nemrežno.
- Če se **podatki spreminjajo pogosto** (npr. podatki, ki jih beleži senzor v realnem času), je treba **metapodatke** natančno uparjati s sprotnimi podatki o poteku dela, **spremembe** pa morajo prav tako biti **sprotne**.

Posodabljanje vaših metapodatkov – načrtovanje sprememb

Metapodatki delujejo v globalnem kontekstu, ki se spreminja!

- **Organizacija** – ustanovijo se oddelki, spojijo so z drugimi oddelki, odgovornost se prenese.
- **Uporaba podatkov** – ob podatkih vznikajo nove aplikacije.
- **Referenčni podatki** – kontrolirani podatki se razvijajo in se povezujejo.
- **Podatkovni standardi in tehnologije** – življenjski cikel tehnologije se ves čas krajša; kako bo splet videti jutri?
- **Orodja in sistemi** – razvoj shranjevanja, pasovne širine, mobilni ...

Metapodatke je treba osveževati v kar največji meri ob upoštevanju časa in denarja, ki ga imamo na voljo.

Hramba metapodatkov – katere možnosti so na voljo?

Metapodatke je mogoče vdelati skupaj s podatki ali pa jih shraniti ločeno od podatkov, odvisno od operativnih zahtev.

- Vdelovanje metapodatkov v podatke (npr. pisarniški dokumenti, MP3, JPG, podatki RDF); vdelani podatki omogočajo lažjo izmenjavo.
- Ločevanje metapodatkov od podatkov (npr. v podatkovni zbirki) s povezavami na ustrezne podatkovne datoteke omogoča lažje upravljanje.

Metapodatke je mogoče shraniti v “**klasično**” **relacijsko podatkovno** zbirko ali v **obliki RDF-trojke**, odvisno od tega, katera orodja so na razpolago in kakšne so zahteve glede zmogljivosti in kapacitet.

Postopki izbrisa podatkov

V mnogih primerih je treba metapodatke ohraniti nekaj časa po izbrisu podatkov, ki jih opisujejo.

Do **izločitve ali izbrisa podatkov** pride, kadar npr.:

- podatki niso več potrebni,
- podatki niso več veljavni,
- podatki so napačni,
- podatke umakne lastnik/tisti, ki jih je objavil.

V takem primeru naj metapodatki **vsebujejo informacijo**, da je bil podatek izbrisan, če pa je bil **arhiviran**, naj povedo, kje se lahko pridobi **arhivirana kopija**.

Objava metapodatkov – katere možnosti so na voljo?

- ‘Odprta’ objava: neposredni dostop do URI-jev.
 - To je možnost, ki je najbolj usklajena z vizijo povezanih odprtih podatkov (Linked Open Data) in omogoča delovanje po načelu ‘hodi za nosom’.
- Naj bodo vaši metapodatki na voljo preko **končne točke SPARQL**.
 - To zunanjim sistemom omogoča pošiljanje poizvedb trojčku RDF.
 - Nujno je znanje o shemi, ki se uporablja za trojček.
- Zakasnitev objave: dostop do izvožene datoteke v RDF
 - Nastane s konvertiranjem podatkov, ki niso v RDF-obliki, v podatke, ki so v RDF-obliki.
 - Omogoča nemrežno masovno žetev in predpomnenje podatkovnih zbirk.

See also:

<http://www.slideshare.net/OpenDataSupport/licence-your-data-metadata>

Kakovost metapodatkov

Kakovost in celovitost deskriptivnih metapodatkov vaših podatkovnih naborov neposredno vpliva na njihovo zmogljivost iskanja in vnovično uporabo.

O kakovosti metapodatkov govorimo, ko so zagotavljeni ... (1/3)

- **Točnost** metapodatkov - ali se lastnosti vira odražajo ustrezno?
 - *Npr. kažejo pravi naslov, pravo licenco, pravega objavitelja in omogoča uporabnikom, da odkrivajo vire, ki jih potrebujejo.*
- **Razpoložljivost** vaših metapodatkov – ali se do metapodatkov lahko dostopa sedaj ali še čez nekaj časa v prihodnosti?
 - *Npr. razpoložljivost za indeksiranje in prenos in vključenost v redni postopek varnostnega kopiranja.*
- **Popolnost** vaših metapodatkov – ali so zajete vse pomembne lastnosti vira (v smislu tega, kaj praktično in ekonomsko izvedljivo in nujno za uporabo)?
 - *Npr. navedba licence, ki ureja vnovično uporabo, ali formata distribucije, ki omogoča filtriranje teh vidikov.*

O kakovosti metapodatkov govorimo, ko so zagotavljeni ... (2/3)

- **Skladnost** vaših metapodatkov s sprejetimi standardi – ali se metapodatki skladajo s specifičnimi metapodatkovnimi standardi ali profilom aplikacije?
 - *Npr. opis nabora podatkov je skladen z DCAT-AP.*
- **Konsistentnost** vaših metapodatkov – ali podatki morda ne vsebujejo nasprotij?
 - *Npr. ob večkratnih in nasprotujočih si licenčnih izjav za isti podatek.*
- **Verodostojnost** in **izvor** vaših metapodatkov – ali metapodatki temeljijo na zaupanja vrednih virih?
 - *Npr. povezave z referenčnimi podatki, ki jih objavlja in upravlja neka trdna organizacija (npr. Urad za publikacije EU).*

O kakovosti metapodatkov govorimo, ko so zagotavljeni ... (3/3)

- **Zmožnost obdelave** metapodatkov – ali so metapodatki ustrezno strojno berljivi?
 - *Npr. metapodatki določenega nabora podatkov so na voljo v oblik RDF in/ali XML, ne pa kot prosto besedilo.*
- **Relevantnost** metapodatkov – ali metapodatki vsebujejo pravo količino informacij za določeno konkretno nalogo?
 - *Npr. omejiti informacijo, da se zagotovi storitev, ki optimalno odgovarja uporabnikovim potrebam.*
- **Pravočasnost** vaših metapodatkov – ali metapodatki ustrezajo dejanskim (trenutnim) lastnostim vira in ali so objavljeni dovolj zgodaj?
 - *Npr. navedba zadnjih sprememb datuma vira, s čimer se zagotovi svežina metapodatkov, tako da uporabnik vedno dobi najnovejšo informacijo.*

Izmenjava metapodatkov o naborih podatkov

Preslikava vaših metapodatkov za skupni metapodatkovni slovar, kakor je npr. DCAT-AP in izmenjava metapodatkov med platformami.

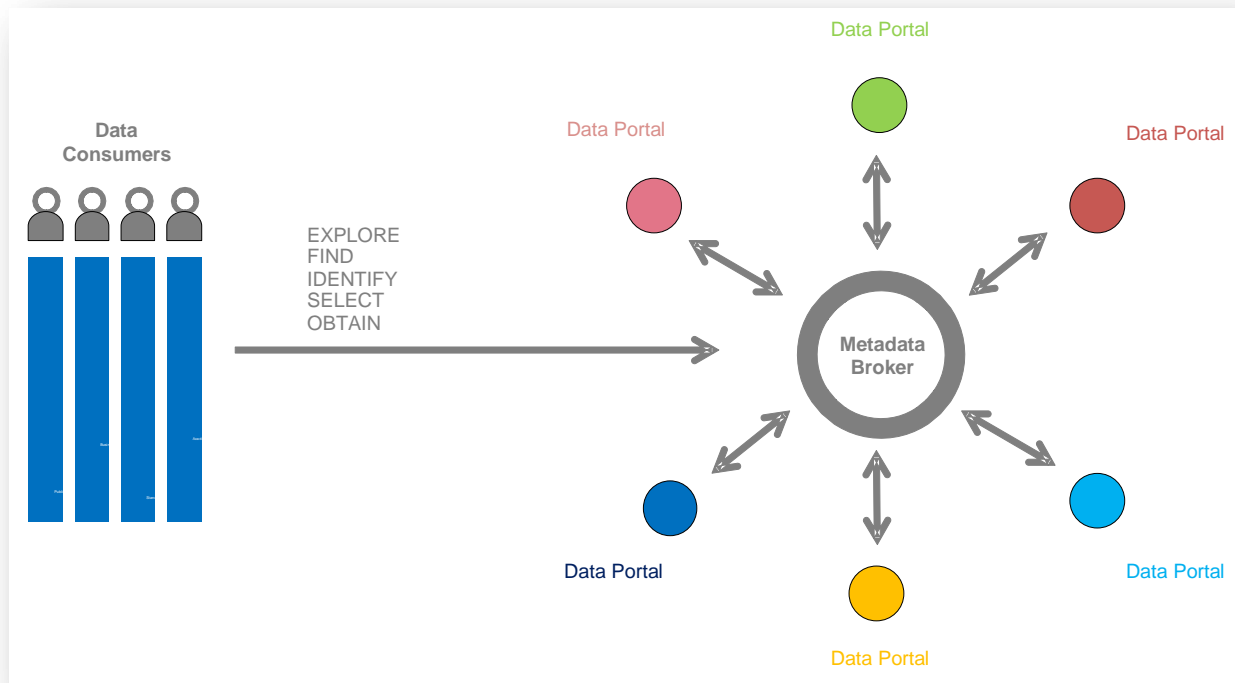
Homogeniziranje metapodatkov

Ob izmenjavi med sistemi naj se metapodatki preslikajo na skupni model, tako da imata pošiljatelj in prejemnik enako razumevanje pomena metapodatkov.

- Na **shematični ravni** lahko metapodatki, ki prihajajo iz različnih virov, temeljijo na **različnih metapodatkovnih shemah**, npr. DCAT, schema.org, CERIF, na lastnem modelu ...
- Na **ravni podatka (vrednosti)** naj se lastnostim metapodatkov pripišejo vrednosti iz **različnih kontroliranih slovarjev** ali **sintaks**, npr.:
 - jezik: angleščina se lahko izrazi kot <http://publications.europa.eu/resource/authority/language/ENG> or as <http://id.loc.gov/vocabulary/iso639-1/en>
 - datumi: ISO8601 (“20130101”) proti W3C DTF (“2013-01-01”)

Primer: Homogeniziranje metapodatkov o naborih podatkov (Profil aplikacije DCAT za podatkovne portale v Evropi)

DCAT-AP se lahko uporabi kot skupni model za izmenjavo metapodatkov s platformami odprtih podatkov po Evropi in/ali s posrednikom podatkov (npr. Open Data Interoperability Platform - ODIP).



See also:

http://joinup.ec.europa.eu/asset/dcat_application_profile/home

Primer preslikave - data.gov.uk

Scottish Road Accident Statistics **dct:title (Dataset)**

Data about injury road accidents, accident costs, vehicles involved, drivers and riders, drink-drive accidents, drivers breath tested, casualties and international comparisons.



Source agency: Scottish Government

Designation: National Statistics

Language: English **dct:language**

Alternative title: Scottish Road Accident Statistics

dct:description

Licence **dct:licens**

UK Open Government Licence (OGL)



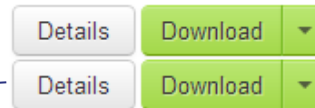
Dcat:accessURL

Data Resources **2**

Key statistics for 2007

2007 Volume

dct:title (Distribution)



dcat:downloadURL, dct:issued, dct:format, dct: description

Additional Information

Openness score	★★★★★
Geographic coverage	Scotland dct:spatia
National statistic	yes I
ONS Category	Travel and Transport dct:theme
Temporal coverage	No value dct:temporal
Date added computed	No value
Date updated computed	No value

dct:publisher

Publisher

Scottish Government

Enquiries:

No details supplied

FOI Contact:

adms:contactPoint

• Web:

<http://www.whatdotheyknow.com...>

Tags

- accident
- health-well-being-and-care
- road
- road-accidents
- road-safety
- roads
- safety
- transport
- transport-accidents-and-casualties
- travel-and-transport

dcat:keyword

About this dataset

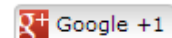
- Added to data.gov.uk: 10/12/2011
- Modified on data.gov.uk: 10/06/2013
- History of changes
- JSON, API and URI for developers

dct:issued

dct:modified

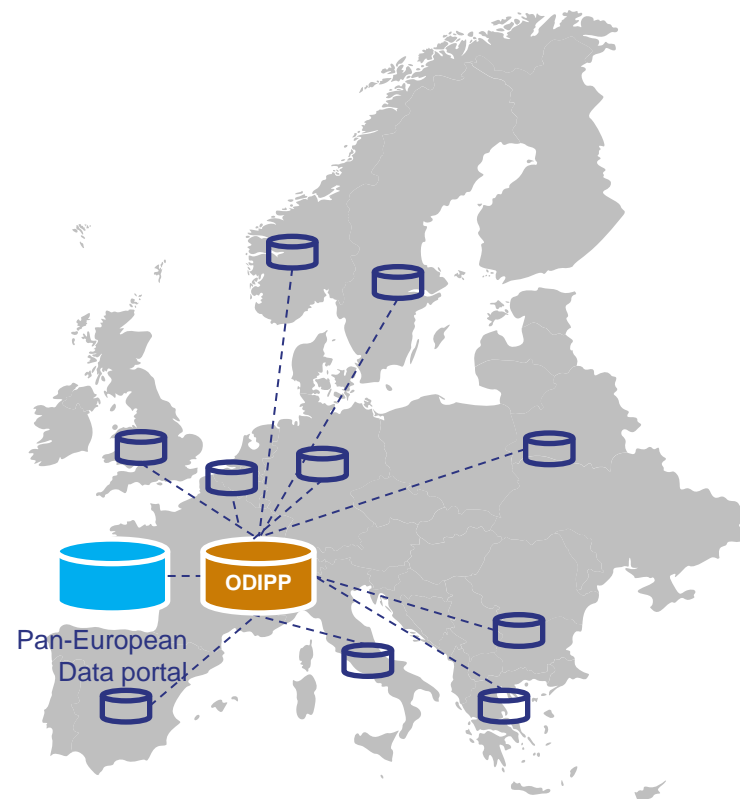
Do more with this data

- Share your app
- Share an idea
- Request new data



Kakšne so zmogljivosti interoperabilne platforme odprtih podatkov (Open Data Interoperability Platform)?

- **Žetev** metapodatkov s portala odprtih podatkov.
- **Preoblikovanje** metapodatkov v RDF.
- **Harmonizacija** metapodatkov v RDF, ki se ustvarijo v predhodnih korakih z DCAT-AP.
- **Validacija** harmoniziranih metapodatkov proti DCAT-AP.
- **Objava** deskriptivnih metapodatkov kot povezanih odprtih podatkov (Linked Open Data).



See also:

<http://www.slideshare.net/OpenDataSupport/promoting-the-re-use-of-open-data-through-odip>

Sklepi

- Metapodatki zagotavljajo informacije o vaših podatkih in virih. Kakovost metapodatkov neposredno vpliva na zmožnost odkrivanja virov in vnovično uporabo virov.
- Za upravljanje z metapodatki je treba slediti strukturiranemu pristopu.
- Življenjski krog metapodatkov podaljšuje življenjski krog naborov podatkov (metapodatki pred objavo in po izbrisu).
- Homogenizirani metapodatki omogočajo delovanje metapodatkovnih posrednikov, s tem pa se lahko zmanjšujejo ovire do vaših virov in izboljša prepoznavnost in možnosti, da te odkrijejo, kar pomeni, da se s tem krepi njihov potencial vnovične uporabe. .

Delo v skupini in vprašanja



<http://www.visualpharm.com>

Delo v dveh skupinah: izberite en nabor podatkov iz vaše države in ga opišite z DCAT Application Profile.



<http://www.visualpharm.com>

Ali ima vaša organizacija minimalni nabor metapodatkov, ki se zagotavljajo skupaj z odprtimi podatki?



<http://www.visualpharm.com>

Katere so po vašem mnenju glavne ovire za vnovično uporabo standardnih kontroliranih slovarjev za vaše metapodatke?



<http://www.visualpharm.com>

Ali imate in uporabljate podatkovno in/ali metapodatkovno metodologijo na ravni vaše organizacije?

Rešite spletni test tukaj!

Hvala!
... vprašanja?

To predstavitev je pripravila organizacija Open Data Support

Zavrnitev odgovornosti

Pogledi, izraženi v tej predstavitvi, so izključno pogledi avtorjev in se v nobenih okoliščinah ne smejo tolmačiti kot uradno mnenje Evropske komisije.

Evropska komisija ne odgovarja za točnost podatkov v tej predstavitvi niti ne prevzema nobene odgovornosti za njihovo uporabo .

Sklicevanje na katere koli posebne produkte, specifikacije, postopke ali storitve blagovnih znamk, proizvajalcev ali drugih oseb ne pomenijo nujno potrditev, priporočilo ali prednostno izbiro Evropske komisije.

Avtor je z vso skrbnostjo poskrbel, da je po potrebi za uporabo pridobil vsa dovoljenja za dele besedil, vključujoč ilustracije, zemljevide in grafične prikaze, na katerih že obstajajo intelektualne pravice nosilcev teh pravic ali njihovih zakonitih zastopnikov.

Predstavitev je skrbno pripravila skupina PwC, ki pa ne odgovarja in ne jamči (niti izrecno niti implicitno) za popolnost ali točnost vsebovanih informacij. PwC ne odgovarja za informacije v tej predstavitvi ali odločitve ali posledice, ki bi nastale na njihovi podlagi. Prav tako ni odgovoren za škodo, ki bi nastala na podlagi informacij iz te predstavitve. Informacije v tej predstavitvi so splošne narave in namenjene zgolj kot napotek v zadevah splošnega interesa. Predstavitev ni nadomestek za strokovni nasvet v kakršni koli zadevi. Bralec naj pred ukrepanjem na podlagi informacij v tej publikaciji poišče strokovni nasvet.

Avtorji:

Max Dekkers, Michiel De Keyzer, Nikolaos Loutas and Stijn Goedertier

Reference

Slide 6, 7:

NISO. Understanding Metadata.

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

Slide 9:

- Dublin City University. Chapter 3: Introduction to XML.

<http://wiki.eeng.dcu.ie/ee557/q2/326-EE.html>

W3C. RDF Primer. <http://www.w3.org/TR/rdf-primer/>

Slide 12:

http://gondolin.rutgers.edu/MIC/text/how/catalog_glossary.htm

- Dublin Core. Example XML Schema. <http://dublincore.org/schemas/xmls/qdc/dc.xsd>
- Dublin Core, Example RDF Schema. <http://dublincore.org/2012/06/14/dcterms.rdf>

Slide 14, 33:

- The ISA Programme. DCAT Application Profile for Data Portals in Europe - Final Draft.

https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/dcat-application-profile-data-portals-europe-final-draf

Slide 18:

- ListPoint. ObjectInCrimeClass. <http://www.listpoint.co.uk/CodeList/details/ObjectInCrimeClass/1.2/1>

Slide 19:

- Publications Office. Countries Name Authority List. <http://open-data.europa.eu/en/data/dataset/2nM4aG8LdHG6RBMumfkNzQ>

Dodatno branje



Understanding Metadata, NISO.

<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>



Ben Jareo and Malcolm Saldanha. The value proposition of a metadata driven data governance program. Best Practices Metadata. May 2012.

https://community.informatica.com/mpresources/Communities/IW2012/Docs/bos_30.pdf



John R. Friedrich, II. Metadata Management Best Practices and Lessons Learned. The 10th Annual Wilshire Meta-Data Conference and the 18th Annual DAMA International Symposium. April 2006.

<http://www.metaintegration.net/Publications/2006-Wilshire-DAMA-MetaIntegrationBestPractices.pdf>

Povezane pobude



Metadata Management. Trainer screencasts, <http://managemetadata.com/screencasts/msa/>



MIT Libraries. Data Management and Publishing. Reasons to Manage and Publish Your Data, <http://libraries.mit.edu/guides/subjects/data-management/why.html>



ISA Programme. DCAT Application Profile for European Data Portals, https://joinup.ec.europa.eu/asset/dcat_application_profile/description



Generating ADMS-based descriptions of assets using Open Refine RDF, <https://joinup.ec.europa.eu/asset/adms/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf>



The Dublin Core Metadata Initiative, <http://dublincore.org/>

Pridružite se nam ...

Poiščite nas na



Open Data Support

<http://www.slideshare.net/OpenDataSupport>



Open Data Support

<http://goo.gl/y9ZZI>

Sledite ščebetu



@OpenDataSupport

Pridružite se nam



joinup

<http://www.opendatasupport.eu>

Navežite stik

contact@opendatasupport.eu