



UvA-DARE (Digital Academic Repository)

Linked Open Data: Vocabulary als handleiding voor een RDF-dataset

Koster, L.

Published in:
Informatie Professional

[Link to publication](#)

Citation for published version (APA):

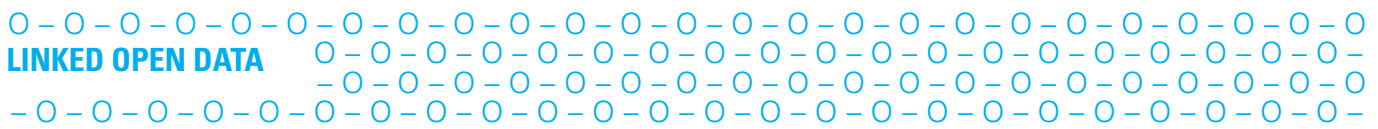
Koster, L. (2014). Linked Open Data: Vocabulary als handleiding voor een RDF-dataset. *Informatie Professional*, 18(4), 38-39.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Vocabulary als handleiding voor een RDF-dataset

Een RDF-dataset is zonder een gedetailleerde handleiding te interpreteren noch te gebruiken. Zo'n blauwdruk wordt met de term 'vocabulary' aangeduid. Een vocabulary verschilt aanzienlijk van de meer traditionele methoden voor het uitwisselen van informatie tussen systemen.

Door: **Lukas Koster**

In de vorige twee afleveringen is beschreven dat een linked open dataset voor iedereen toegankelijk op het web beschikbaar is en dat die dataset triples bevat die met behulp van de universele datanotatiewijze RDF beschreven zijn. Triples zijn relaties tussen telkens twee objecten van een bepaald type (zoals een Persoon en een Boek) met bijbehorende kenmerken.

Een RDF-dataset kan zonder een gedetailleerde handleiding niet geïnterpreteerd en gebruikt worden. Zo'n blauwdruk wordt met de term 'vocabulary' aangeduid. De eveneens in zwang zijnde, via het Semantic Web aan de filosofie ontleende, term 'ontology' wordt vaak gebruikt in een iets abstractere betekenis. Een vocabulary kun je zien als een beschrijving van de interne structuur van een dataset met als een van de belangrijkste toepassingen het uitwisselen van informatie tussen systemen.

Een vocabulary verschilt aanzienlijk van de meer traditionele methoden voor het uitwisselen van informatie tussen systemen. Bij traditionele informatiesystemen is het vaak niet eenvoudig data uit de onderliggende database te krijgen voor hergebruik in andere systemen. De interne structuur van de database is alleen duidelijk voor de systeemontwikkelaars, en de informatie in de database kan alleen worden benaderd via interne systeemfuncties. RDF-vocabulary's vormen een poging los te komen van de beper-

kingen van naar binnen gerichte datastructuren en te komen tot een open, flexibele en naar buiten gerichte datastructuur met gebruikmaking van de mogelijkheden van het web.

Informatie ophalen

De belangrijkste manier waarop informatie uit een traditioneel systeem kan worden gehaald is via de eindgebruikersinterface (wat er op het scherm wordt getoond). Deze manier is niet geschikt voor hergebruik van data door andere systemen. Hiervoor staan in de regel twee opties open: export/import en API's (application programming interface).

Bij export/import wordt de hele database of een deel ervan volgens bepaalde vastgestelde formaten in een extern bestand gezet. Vervolgens kan dat bestand in de database van het doelsysteem worden ingelezen. Bij een API-call wordt 'on the fly' de data over een bepaald object uit het bronsysteem in een vastgestelde structuur verkregen, waarna die informatie meteen verwerkt en gepresenteerd wordt. Bij beide manieren geschiedt de communicatie niet direct met de database, maar via apart ontwikkelde systeemfuncties en datastructuren.

Voor de buitenwereld is het niet te zien hoe de onderliggende database in elkaar zit, welke data er in zitten en of de geproduceerde data overeenstemmen met wat er gevraagd is.

Een RDF-dataset kent die beperkingen niet. De gehele onderliggende datastructuur van de dataset wordt met een voor iedereen op het web beschikbare vocabulary beschreven. Een vocabulary maakt zo alle data uit de dataset beschikbaar voor uitwisseling via universele standaarden, zonder dat aparte systeemfuncties en exportformaten nodig zijn. In tegenstelling tot traditionele systemen is het model voor datauitwisseling identiek aan de beschrijving van de interne datastructuur.

Een voorbeeld

Ter illustratie gebruiken we het Boek-Auteur-voorbeeld. In RDF worden typen en kenmerken aangeduid als classes (bijvoorbeeld 'Book', 'Person') en property's (bijvoorbeeld 'title', 'author', 'name'). Het moge duidelijk zijn dat de voertaal Engels is, omdat het gaat om universele uitwisseling op het web. Een daadwerkelijk voorkomen van een class kunnen we aanduiden met de term 'object' (bijvoorbeeld het object

'Uit Berlijn' is van het type (class) 'Book', 'Armando' is een Persoon). Dit is wat we in een RDF-dataset zien. De vocabulary fungeert als de sleutel tot het interpreteren van de dataset.

Alle classes en property's worden met URI's (unieke sleutels op het web) aangeduid, toegankelijk voor iedereen. Dat geldt eveneens voor de vocabulary zelf. Een relatie met een andere class wordt ook property genoemd. Er is geen onderscheid omdat elke property ofwel door een tekstwaarde ofwel door een relatie, in de vorm van een URI, kan worden weergegeven, bijvoorbeeld `author="Armando"` of `author="http://viaf.org/viaf/9885610/"`.

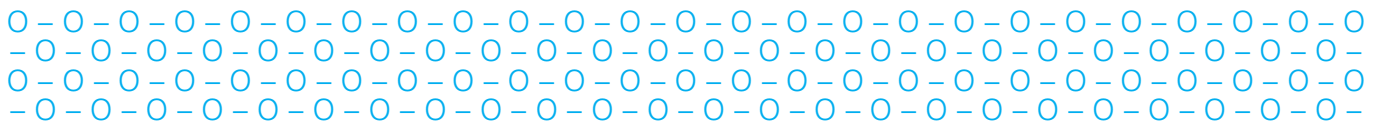
In RDF is het basisconcept de triple, een relatie tussen twee objecten. Een object is onderdeel van één of meer triples. (De term 'object' wordt hier anders gebruikt ('ding') dan in de nauwere Engelstalige beschrijving van de inhoud van een triple (subject - predicate - object: onderp - gezegde - voorwerp), waarbij 'object' (lijdend) 'voorwerp' betekent. Een 'object' kan zowel onderwerp als voorwerp in een triple zijn.) In tegenstelling tot traditionele databases kan elk object zoveel verschillende property's hebben als nodig is, binnen of buiten de oorspronkelijke dataset.

Veelgebruikte vocabulary's zijn FOAF (persoonsbeschrijving, <http://xmlns.com/foaf/0.1/>), SKOS (trefwoorden en thesauri, <http://www.w3.org/2004/02/skos/core#>), BIBO (bibliografische informatie, <http://purl.org/ontology/bibo/>) en DC-Terms (erfgoed, <http://purl.org/dc/terms/>). Zie voor een overzicht <http://lov.okfn.org>.

Datastructuur

Een vocabulary bestaat uit machi-

'Een vocabulary fungeert als model zowel voor datauitwisseling als voor interne datastructuur'



ne-leesbare tekst, bevattend notatiesymbolen en URI's van classes en property's, met verklarende tekst. De datastructuur Boek-Auteur uit ons voorbeeld kan als volgt in RDF worden beschreven:

```
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
bibo:Book ;
dcterms:title ;
dcterms:created ;
dcterms:creator ;
dcterms:publisher ;
dcterms:subject .
```

Hier wordt, in een klein deel van de BIBO-vocabulary, gebruik gemaakt van property's uit de DC-Terms-vocabulary. Er worden voor één class vijf mogelijke property's gedefinieerd. De property 'dcterms:title' verwijst naar <http://purl.org/dc/terms/title>, waar nadere informatie voor mens en machine te vinden is.

Het boek 'Uit Berlijn' kan worden weergegeven als:

```
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
<http://example.com/UitBerlijn>
rdf:type bibo:Book ;
dcterms:title "Uit Berlijn" ;
dcterms:created "1982" ;
dcterms:creator <http://viaf.org/viaf/9885610/> ;
dcterms:publisher <http://viaf.org/viaf/266801016> ;
dcterms:subject <http://dbpedia.org/resource/Berlin> .
```

Het boek heeft een eigen URI, hier een voorbeeld-URI. Er worden zes triples voor één object weergegeven, behalve de property's title, creator et cetera namelijk ook het type Book. Sommige zijn in de vorm van een URI, sommige in de vorm van tekst. Men zou `dcterms:creator` <<http://viaf.org/viaf/9885610/>>

zoals gezegd ook kunnen schrijven als `dcterms:creator "Armando"`, maar dan verdwijnt de connectie met de diverse naamvarianten en pseudoniemen die in de externe dataset VIAF beschikbaar is, en met andere informatiebronnen waarin ook deze URI wordt gebruikt. Bij meer auteurs of onderwerpen worden eenvoudig extra regels toegevoegd.

Het onderwerp is hier niet als Nederlandstalige tekst weergegeven, maar als URI, hetgeen in principe toegang geeft tot extra informatie in diverse talen over het onderwerp, en informatie uit andere bronnen die hiervan gebruik maken.

Een persoon kan ook als onderwerp worden opgenomen, met verwijzing naar dezelfde informatiebron als voor auteurs, bijvoorbeeld:

```
dcterms:subject <http://viaf.org/viaf/9885610/> .
```

Een autobiografie heeft dan:

'Een vocabulary maakt het publiceren van data voor hergebruik door anderen mogelijk'

```
dcterms:creator <http://viaf.org/viaf/9885610/> ;
dcterms:subject <http://viaf.org/viaf/9885610/> .
```

Tot slot

Samenvattend heeft een vocabulary de volgende voordelen:

- > een vocabulary kan door iedereen voor nieuwe toepassingen worden gebruikt;
- > een vocabulary is niet systeem- of domeingebonden;
- > een vocabulary kan eenvoudig uitgebreid worden;
- > een vocabulary fungeert als model zowel voor datauitwisseling als voor interne datastructuur;
- > een vocabulary biedt automatisch relaties met andere domeinen

(zoals Kunst en Geografie);
> een vocabulary maakt het publiceren van data voor hergebruik door anderen mogelijk, ook zonder bijbehorend systeem.

RDF-vocabulary's maken het in theorie mogelijk informatie uit alle mogelijke webbronnen geïntegreerd aan te bieden. Om dit te verwezenlijken zijn wel speciale applicaties nodig (daarover meer in een latere bijdrage). Hoe dan ook ligt een wereldwijd open datanetwerk nu binnen handbereik. <

Lukas Koster is Coördinator Bibliotheeksystemen bij de Bibliotheek van de Universiteit van Amsterdam.

