



UvA-DARE (Digital Academic Repository)

Linked data

Koster, L.

Published in:

IK: tijdschrift over intellectueel kapitaal

[Link to publication](#)

Citation for published version (APA):

Koster, L. (2011). Linked data. *IK: tijdschrift over intellectueel kapitaal*, 10(2), 11-13.

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.

Lukas Koster

Linked data

Semantisch web, linked data, web 3.0... Het zijn allemaal termen die vaak door elkaar worden gebruikt. Onterecht vindt Lukas Koster van de Universiteit van Amsterdam. Wat houden die begrippen nu eigenlijk in?

De term *semantic web* is enigszins misleidend. De omschrijving “a group of methods and technologies to allow machines to understand the meaning - or ‘semantics’ - of information on the World Wide Web” (zie http://en.wikipedia.org/wiki/Semantic_Web) is te ambitieus. Het heeft niets te maken met kunstmatige intelligentie. Machines begrijpen niet.

Een betere omschrijving is “a web of data that can be processed directly and indirectly by machines” (Tim Berners-Lee). Onder ‘machines’ moeten we dan niet verstaan ‘computers’, maar digitale informatiesystemen die daarop draaien. *Web of data* staat hier tegenover het *web of documents*, het huidige web bestaande uit documenten, pagina’s, bestanden en andere meer of minder gestructureerde objecten van allerlei aard.

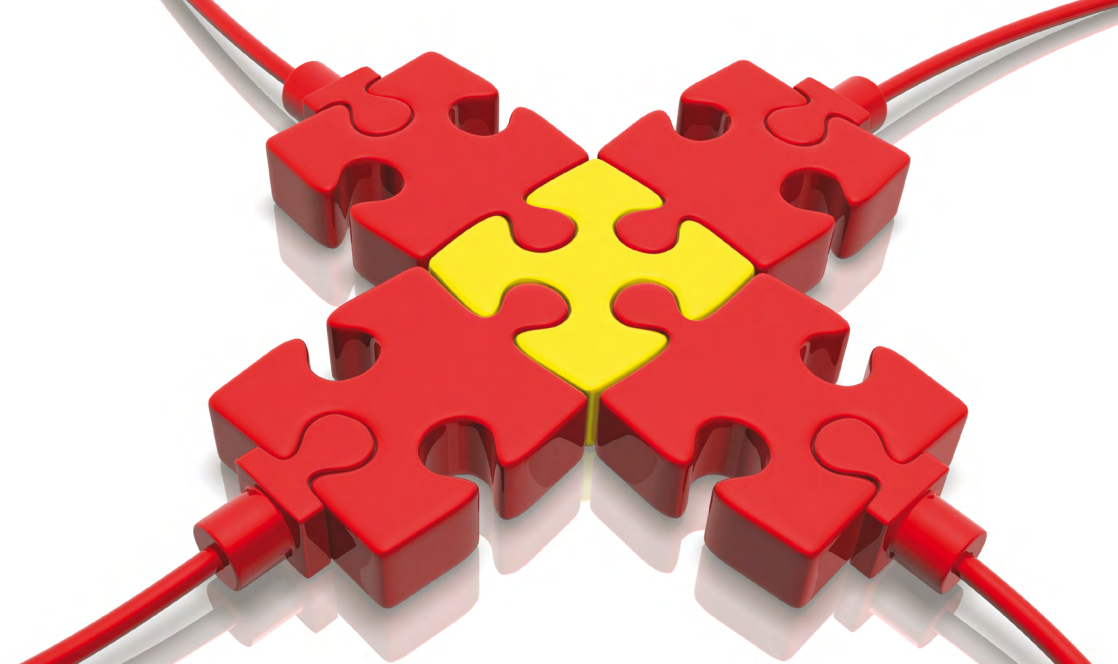
Beter nog zouden we moeten spreken van *linked data*, ofwel het gebruik van het web om ge-

kwalificeerde, omschreven links (*typed links*) te maken tussen data afkomstig uit verschillende bronnen. Tim Berners-Lee heeft hiervoor vier regels geformuleerd (zie www.w3.org/DesignIssues/LinkedData.html) die als volgt geïnterpreteerd kunnen worden:

Het linken van data is een methode voor het via het web beschikbaar stellen van relaties tussen dingen (data, concepten, documenten) die ergens op het web aanwezig zijn, met gebruikmaking van URI’s voor het identificeren, RDF voor het beschrijven en HTTP voor het publiceren van die dingen en relaties, zodanig dat zij zowel door mensen als door software begrepen en gebruikt kunnen worden.

Belangrijkste begrippen

Een URI is een **Uniform Resource Identifier**, gebruikt om dingen op het web uniek te identificeren. Via het HTTP-protocol zijn deze URI’s via het web te benaderen. Een URI is vergelijkbaar met een unieke sleutel in een databasesysteem,



waarbij het grote verschil is dat deze laatste alleen binnen dat systeem geldig is, terwijl een URI ook buiten het systeem op het hele web geldt.

Het gebruik van URI's betekent niet dat elk ding slechts eenmaal op het web als linked data-object voorkomt. De aard van het web is nu eenmaal zo dat er altijd meer bronnen over één onderwerp beschikbaar zullen zijn. In linked data kan een *same as*-relatie gebruikt worden om aan te geven dat het om hetzelfde gaat. Hierbij speelt ook de betrouwbaarheid van informatie een rol.

Data = relaties. **Relaties** tussen dingen geven meer informatie. Bijvoorbeeld: er is een boek en er is een persoon. De relatie 'geschreven door' geeft meer informatie: dit boek is geschreven door die persoon. Andersom gezien: deze persoon is de schrijver van dat boek. Dit geeft aan dat een rol van een persoon (in dit geval

schrijver) de beschrijving is van een relatie (een typed link) van die persoon met een bepaald boek. In linked data wordt deze drie-eenheid van subject - predicate - object (onderwerp - gezegde - voorwerp) een *triple* genoemd.

Relaties worden ook gebruikt binnen traditionele databasesystemen (RDBMS - *relational database management systems*). Maar met linked data kunnen relaties ook gelegd worden tussen objecten in verschillende systemen op het web. Het boek kan dan bijvoorbeeld in een boeken-database staan en linken naar een persoon in een personendatabase.

RDF (Resource Description Framework) is als metadatumodel de algemene methode om triples herbruikbaar te beschrijven. Subject, object (dingen) en predicate (relatie) kunnen in de vorm van URI's of een *literal string* (tekenreeks) gedefiniëerd worden. Een voorbeeld:

- Een boek (URI) is geschreven door (URI) een persoon (URI).
- Een boek (URI) heeft de afmeting (URI) 120 bladzijden (literal)
- Een boek (URI) heeft de soort (URI) roman (URI)

RDF kan in verschillende vormen worden geïmplementeerd. De meest gebruikte is **RDF/XML**, maar er zijn ook andere niet XML-vormen, zoals **N5** en **Turtle**.

Om de volgens RDF gepubliceerde data herbruikbaar te maken is het gebruikelijk ook het domeinspecifieke datamodel in RDF op het web te publiceren. Deze RDF-datamodellen worden ontologies, of ook wel *vocabularies* genoemd. Deze termen worden door elkaar gebruikt, maar in feite is een ontology een meer abstract model, terwijl vocabularies vaak ook toegestane waarden bevatten. Een ontology omschrijft de entiteiten in een bepaald domein met daarbij de mogelijke attributen en relaties.

Het is goed bestaande vocabularies te gebruiken, en eventueel geheel of gedeeltelijk met andere vocabularies samen te voegen tot nieuwe modellen. Uiteraard kunnen op deze manier ook data uit verschillende domeinen aan elkaar gekoppeld worden.

Twee veelgebruikte algemene vocabularies:

- **SKOS for subjects**
- **FOAF for persons**

Een URI voor een type, attribuut of een relatie bevat dan zowel een verwijzing naar de gebruik-

te vocabulary, als de verwijzing naar het desbetreffende soort type, attribuut of relatie.

Tot slot: het linked data-concept kan pas echt optimaal gebruikt worden als de betreffende informatie ook open is, dat wil zeggen dat op de data een licentie van toepassing is die hergebruik toestaat. In dit verband wordt ook wel gesproken over *linked open data* (LOD).

Content negotiation is de term die wordt gebruikt voor het mechanisme om onderscheid te maken tussen informatieverzoeken door mensen en software. Een URI wordt door de betreffende server *geredirect* naar een voor mensen leesbaar document (bijvoorbeeld een HTML-pagina) als het verzoek vanuit een browser komt, of naar een RDF-document als het vanuit een linked data-applicatie komt. **►►**

Lukas Koster (<http://twitter.com/lukask>)

werkt bij de bibliotheek van de Universiteit van Amsterdam en is een van de initiatiefnemers van het **Dutch Culture Linked Open Data-symposium** dat op 13 mei plaatsvindt in de Openbare Bibliotheek van Haarlem. Diverse bekende sprekers in de erfgoedsector van Nederland en Vlaanderen zullen ingaan op de huidige succesvolle toepassingen met het linken van data.

De Creative Commons Attribution-ShareAlike 3.0 Unported

Licentie is van toepassing op dit werk. Ga naar <http://creativecommons.org/licenses/by-sa/3.0/> of stuur een brief naar

Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, VS om deze licentie te bekijken.