



UvA-DARE (Digital Academic Repository)

Linked Open Data: Meer werk maken van (meta)data, ook in bibliotheken, archieven en musea

Koster, L.

Published in:
Informatie Professional

[Link to publication](#)

Citation for published version (APA):

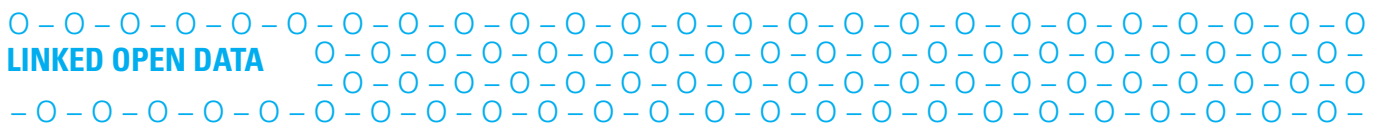
Koster, L. (2014). Linked Open Data: Meer werk maken van (meta)data, ook in bibliotheken, archieven en musea. *Informatie Professional*, 18(1), 30-31.

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.



Meer werk maken van (meta)data, ook in bibliotheken, archieven en musea

In een korte serie gaat Lukas Koster in op het fenomeen Linked Open Data. In de erfgoedsector lijkt de adoptie ervan langzamer te gaan dan daarbuiten, vermoedelijk doordat de informatiesystemen die er gebruikt worden hiervoor niet zijn ontworpen. Wat is Linked Open Data en wat gaat het betekenen voor bibliotheken, archieven en musea? In de komende afleveringen is er aandacht voor open data, ontologies en authority files, technische aspecten en beschikbare apps. Maar begonnen wordt met een 'Linked Open Data for dummies'.

Door: Lukas Koster

Linked Open Data lijkt hard op weg mainstream te worden. Grote organisaties doen eraan, zoals de BBC, Europeana, de Nationale Bibliotheek van Frankrijk, ja zelfs commerciële partijen als OCLC. Er wordt veel over geschreven en gepraat, in tijdschriften, op social media en op congressen. Meestal betreft het algemene theoretische beschouwingen of individuele projecten. Algemeen toepasbare oplossingen voor bibliotheken, archieven en musea ontbreken vooralsnog. Wat betekent dit nu voor informatieprofessionals en de organisaties waar zij voor werken? Met Linked Open Data kunnen de barrières tussen bestaande informatiesystemen en de daarin opgesloten informatie worden overwonnen, wat kan leiden tot vermindering van redundantie, grotere efficiency en nieuwe kennis.

Term

De term Linked Open Data is een samenvoeging van twee verschillende zaken. Linked Data is een technisch concept, gericht op het op gestructureerde en gestandaardiseerde wijze aan elkaar koppelen van digitale informatie uit verschil-

lende bronnen. Op die manier kan nieuwe of verrijkte informatie worden aangeboden.

Open Data is een juridisch concept. Het is erop gericht om met licenties toestemming te geven voor (her-)gebruik van digitale informatie door anderen dan de formeel rechthebbende. Linked Data zonder Open Data zou vooral een theoretisch idee blijven, vandaar dat het vaak in één adem wordt genoemd.

De basisbegrippen van Linked Data – RDF, URI's, links en het web of data – zijn in vier regels beschreven door Tim Berners-Lee, de bedenker van het World Wide Web en HTTP; het 'web of documents' (<http://www.w3.org/DesignIssues/LinkedData.html>).

Niet nieuw

Het uitgangspunt van Linked Data op zich is niet nieuw. Het gaat om relaties tussen eenheden van informatie. Bijvoorbeeld: een boek is geschreven door een persoon. De relatie 'geschreven door' geeft meer

informatie over het boek. Andersom gezien: een persoon is de schrijver van een boek. De rol van deze persoon ('schrijver') is de beschrijving van een relatie van die persoon met een bepaald boek.

In een database wordt deze relatie geïmplementeerd door de unieke sleutel van het Personen-tabelrecord op te nemen in het betreffende Boeken-tabelrecord. Dit is tenminste het geval in systemen met een werkelijke relationele database. In digitale catalogus-systemen worden helaas maar al te vaak de namen van de auteurs als tekststring in het boekenrecord opgenomen. Het voordeel van het gebruiken van sleutels in plaats van tekstwaarden is dat de betreffende entiteiten slechts eenmaal worden gedefinieerd en beschreven, wat grotere efficiency en minder fouten oplevert. In traditionele, op zichzelf staande computersystemen werken deze relaties alleen binnen de betreffende systemen.

'Linked Data is een technisch concept, Open Data is een juridisch concept'

Wel nieuw

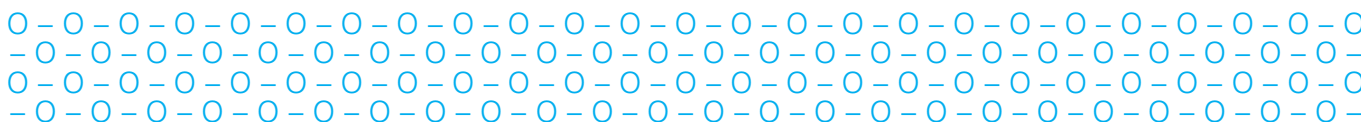
Wat wel nieuw is aan Linked Data, is dat hier het World Wide Web het uitgangspunt is. En waarbij het mogelijk is externe relaties te leggen tussen verschillende systemen op het web. Heeft de persoon die een bepaald boek heeft geschreven, bijvoorbeeld ook schilderijen gemaakt die geregistreerd zijn in een aparte museumdatabase, dan is dat in een introverte omgeving niet te zien in de boekencatalogus. Met Linked Data op het web zou dat wel mogelijk zijn.

Primaire sleutel

In stand-alone systemen worden individuele records onderscheiden door een unieke sleutel ('primary key' of 'unique identifier') binnen de betreffende tabel. Het 'adres' van een record bestaat dan uit de tabelnaam + de primaire sleutel, bijvoorbeeld 'Personen.1234'.

Op het web is dat duidelijk niet voldoende, omdat er talloze databasetabellen met identieke namen en sleutels bestaan. Adressen op het web worden gevormd door unieke URL's, die bestaan uit het protocol, het domain, een subdomain of map, en een binnen dat domain unieke tekststring om een bepaald document aan te duiden. Bijvoorbeeld: <http://nl.wikipedia.org/wiki/Armando>.

Bij Linked Data wordt de URI (Uniform Resource Identifier) gebruikt als primaire sleutel voor entiteiten op het web. Een URI heeft dezelfde structuur als een URL (voorbeeld: <http://viaf.org/viaf/9885610/>). Systemen die met Linked Data werken, dienen URI's te kunnen opslaan als verwijzing, en deze ook te kunnen verwerken om de voor



mensen begrijpelijke informatie die erachter zit te tonen.

Triple

Om URI's te kunnen verwerken, is het nodig dat informatie waarnaar een URI verwijst in een bepaald standaardformaat wordt aangeleverd. De basis hiervoor is de voornoemde relatie tussen twee entiteiten. 'Boek - geschreven door - persoon'. Dit is dezelfde constructie als die gebruikt wordt in zinnen in natuurlijke taal, zoals we allemaal op school hebben geleerd: onderwerp - gezegde - lijdend voorwerp. In Linked Data wordt hiervoor de Engelse terminologie gebruikt: subject - predicate - object. Zo'n relatie wordt 'triple' genoemd. Linked Data bestaat uit allemaal triples:

- > 'Uit Berlijn' is geschreven door Armando
- > De Openbare Bibliotheek Amsterdam heeft 'Uit Berlijn'
- > Armando is de schilder van 'Schuldig Landschap'.

Ontologies/vocabularies

Het metadatamodel om triples te beschrijven is RDF (Resource Description Framework). Maar RDF op zich is niet genoeg. Het zegt alleen dat er triples worden gebruikt, maar het zegt nog niets over wat voor soort informatie wordt beschreven. Daarvoor zijn inhoudelijke datamodellen nodig die een bepaald gebied beschrijven.

Deze modellen worden 'ontologies' of 'vocabularies' genoemd. Hierin worden ook triples gebruikt, bijvoorbeeld: de eerder genoemde relatie 'Boek - geschreven door - persoon'. Dit is vergelijkbaar met de tabelstructuur in databases. In RDF worden ook de relaties met een URI gedefinieerd.

Het is gebruikelijk ontologies/vocabularies op het web te publiceren om het hergebruik mogelijk te maken van zowel deze vocabularies als de data die met die vocabularies beschikbaar gemaakt zijn. Veelgebruikte ontologies zijn FOAF (persoonsbeschrijving), SKOS (trefwoorden en thesauri), BIBO (bibliografische informatie) en DC (erfgoed).

De triple 'boek geschreven door persoon' kan schematisch als volgt worden weergegeven:

bibo:Book-dc:Creator-foaf:Person
De triples die werkelijke objecten beschrijven, bevatten in plaats van de typebeschrijvingen de daadwerkelijke verschijningen. Ook hier worden voor 'subject - predicate - object' waar mogelijk URI's gebruikt.

"Uit Berlijn" is geschreven door Armando:
<http://www.worldcat.org/oclc/10098995>
<http://purl.org/dc/terms/creator>
<http://viaf.org/viaf/9885610/>

'Armando is de schilder van "Schuldig Landschap":
<http://viaf.org/viaf/9885610/>
<http://purl.org/dc/terms/creator>
<http://hdl.handle.net/10934/RM0001.COLLECT.496040>

'Wat nieuw is aan Linked Data, is dat hier het World Wide Web het uitgangspunt is'

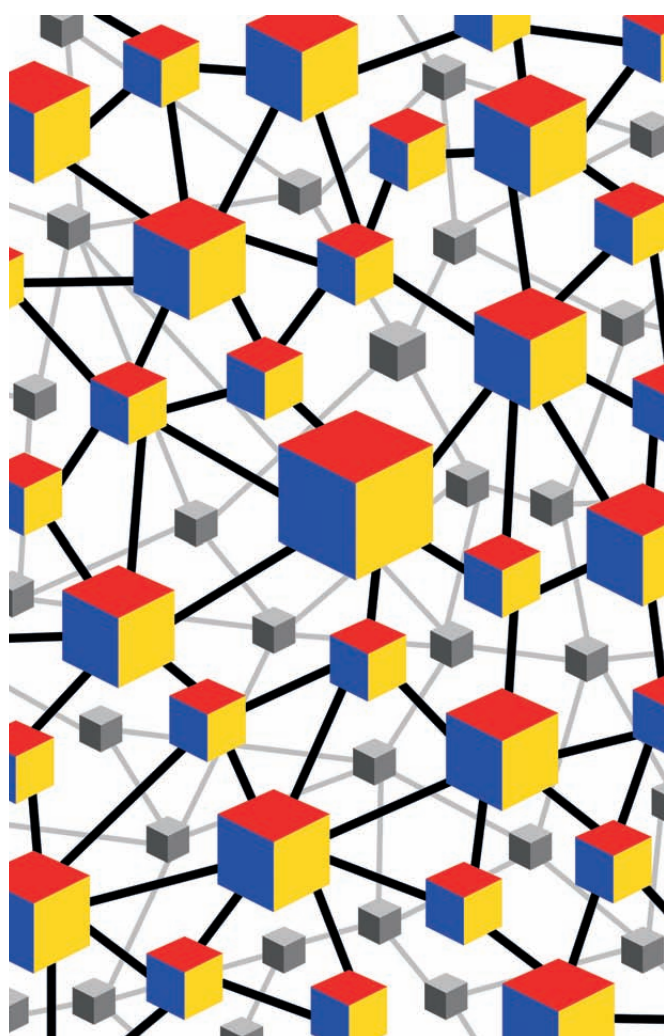
Het gebruik van URI's, oftewel concepten in plaats van tekstwaarden lost het probleem op van verschillende schrijfwijzen van auteurs,

locaties en ook trefwoorden in verschillende talen. Bij het presenteren van de concepten in leesbare vorm moeten de daarbij behorende, van toepassing zijnde termen worden opgehaald. Behalve het inzetten van Linked Data voor dit soort thesauri en voor het presenteren van bestaande informatie uit verschillende bronnen, zal het ook voorkomen dat door het combineren van informatie nieuwe kennis ontstaat.

Twee kanten

Het moge duidelijk zijn dat Linked Open Data twee kanten heeft: het publiceren en het gebruik ervan. Er worden al heel wat data gepubliceerd (VIAF, Worldcat, Library of Congress, et cetera). Het grote probleem is voornamelijk het gebruik van die gepubliceerde data. Bestaande systemen in de LAM-sector (libraries, archives, museums) zijn hier zoals gezegd niet voor ontworpen. Het zijn dus vooral losse projecten in instellingen met voldoende financiering waar met Linked Open Data wordt gewerkt. Gelukkig zijn er ook ontwikkelingen gaande bij leveranciers van veelgebruikte systemen, zowel commerciële als open source.

Als er in plaats van bibliotheek-, archief- en museumssystemen op Linked Open Data gebaseerde informatiesystemen zijn, zal het werk van informatieprofessionals ongetwijfeld veranderen. Behalve dat het traditionele catalogiseren veel meer gecentraliseerd zal gebeuren, zullen er ook verschuivingen optreden naar nieuwe, meer inhoudelijke taken. Dit zal andere kennis en ervaring vereisen en dus ook andere opleidingen.



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