



## UvA-DARE (Digital Academic Repository)

### Linked Open Data: Autoriteit en vertrouwen

Koster, L.

**Publication date**

2014

**Document Version**

Final published version

**Published in**

Informatie Professional

[Link to publication](#)

**Citation for published version (APA):**

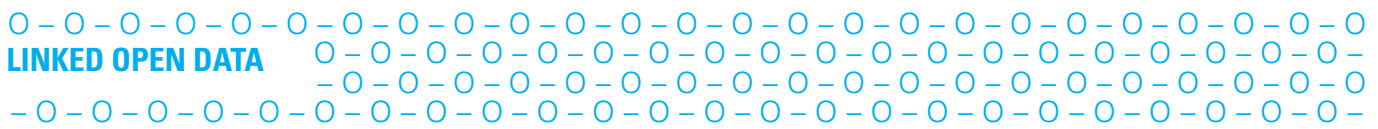
Koster, L. (2014). Linked Open Data: Autoriteit en vertrouwen. *Informatie Professional*, 18(5), 30-31.

**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: <https://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.



# Autoriteit en vertrouwen

Een belangrijk punt bij het publiceren en gebruiken van linked open data is vertrouwen. Wie garandeert de betrouwbaarheid van externe databronnen? Zit die vocabulary wel goed in elkaar? Klopt die door anderen gelegde relatie tussen twee objecten wel? Wie zegt dat jouw eigen data wel verantwoord hergebruikt worden?

Door: **Lukas Koster**

Om met die laatste vraag te beginnen: het antwoord is niemand! Er is geen garantie dat de door jou gepubliceerde data verantwoord worden hergebruikt. Als je je daar als instelling druk over maakt, dan kun je er beter niet aan beginnen. Goed, over tot belangrijker zaken.

## Vocabulary's

Als het gaat om de validiteit van een vocabulary, dan komt het goed uit dat linked open data draait om beschikbaarheid voor iedereen op het web. Iedereen kan vrij eenvoudig een vocabulary online bekijken en controleren. Bij verschillen van inzicht staan er twee opties open: het gebruik van delen van de vocabulary aangevuld met classes en property's uit andere vocabulary's, of het bijdragen aan de ontwikkeling van de betreffende vocabulary door verbeteringsuggesties in te dienen. Alleen maar heel transparant. W3C, het World Wide Web Consortium, stelt voor alle vocabulary's onder beheer van W3C te brengen (<http://www.w3.org/blog/data/2014/01/06/vocabularies-at-w3c/>). Tegenstanders vinden dat dit de open en gedistribueerde aard van linked open data zou aantasten.

## Centraal - gedistribueerd

De tegenstelling 'centrale autoriteit - gedistribueerde verantwoordelijkheid' heeft ook betrekking op de beide overige vragen uit de inleiding. Erfgoedinstellingen en informatieprofessionals beschouwen

zichzelf als autoriteiten op het gebied van informatie. Als ze al niet de oorspronkelijke bron van die informatie zijn, dan staan ze wel in voor de betrouwbaarheid van informatie van anderen.

In de erfgoedsector zijn er twee soorten data waarbij het gebruik van externe bronnen aan de orde kan zijn: de traditionele basismetadata (titel, auteurs, jaar, onderwerpen, herkomst, et cetera) en extra verrijkende informatie.

In het geval van basismetadata (verrijkende informatie komt aanbod bij 'Toezicht') komt het al vaak voor dat men data van externe partijen afneemt. Met name bibliotheken verkrijgen kant-en-klare metadatarecords van uitgeverij, andere erfgoedinstellingen of servicecentra. Daarnaast wordt voor auteursnamen en onderwerpen gebruikt gemaakt van door anderen beheerde thesauri en trefwoordenlijsten als de Nederlandse Basisclassificatie (NBC) en in toenemende mate de Library of Congress Subject Headings. In al deze gevallen gaat het om data afkomstig van vertrouwde organisaties.

## Authority files

Het gebruiken van extern beheerde 'authority files' hoeft in een linked

open data-omgeving niet anders te zijn. Het in vorige afleveringen van deze serie genoemde VIAF (Virtual International Authority File, <http://viaf.org>), beheerd door OCLC, is een goed voorbeeld. VIAF geeft toegang tot de 'name authority files' oftewel 'naamthesauri' van een groot aantal nationale bibliotheken en verwante organisaties. Het gaat hierbij om namen van personen, organisaties, geografische eenheden en dergelijke, met alle opgenomen varianten. Een vergelijkbaar persoonsnamenbestand is de ISNI-database (International Standard Name Identifier, <http://isni.org>) waarin de publieke identiteiten zijn opgenomen van personen die bijdragen aan creatieve werken (auteurs, componisten, regisseurs, acteurs, et cetera). De ISNI-database wordt onder auspiciën van een internationale raad (ISNI Registration Authority) opgebouwd uit een groot aantal bronnen, waaronder ook VIAF. De schrijver/schilder Armando heeft een VIAF-ID <http://viaf.org/viaf/9885610> en een ISNI <http://isni-url.oclc.nl/isni/0000000120171892>.

Een VIAF-ID en een ISNI worden gecreëerd op basis van bestaande data, maar er zijn ook vergelijkbare naamdatabases waar dat anders werkt. De steeds meer gebruikte

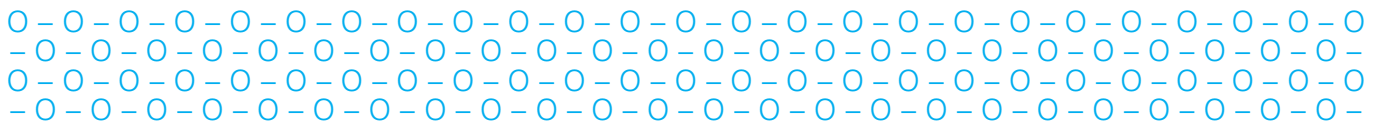
identificer voor wetenschappelijke onderzoekers ORCID (Open Researcher and Contributor Identifier, <http://orcid.org/>) kan door een onderzoeker zelf of door zijn of haar werkgever worden geregistreerd.

## Toezicht

ORCID kent net als VIAF en ISNI een formele toezichtstructuur (Board of Directors) met vertegenwoordigers van gezaghebbende instellingen uit de internationale wetenschappelijke onderzoekswereld. Dit ligt iets anders bij de breder gebruikte populaire online datasources Wikipedia (<http://wikipedia.org>) en zijn linked open data-versie DBpedia (<http://dbpedia.org>). Wikipedia is een online encyclopedie die wordt onderhouden door talloze vrijwilligers over de hele wereld. Er is een formele en informele beheer- en bewakingsstructuur onder auspiciën van de Wikimedia Foundation (<http://www.wikimedia.org>).

DBpedia kent een vergelijkbare structuur en publiceert informatie uit Wikipedia op gestructureerde wijze als linked open data. DBpedia URI's kunnen in principe worden gebruikt voor de traditionele basismetadata als auteurs en onderwerpen, maar ook om de traditionele data te verrijken. Armando kan worden gevonden in DBpedia: [http://dbpedia.org/resource/Armando\\_\(artist\)](http://dbpedia.org/resource/Armando_(artist)), afgeleid van de Wikipedia-pagina <http://nl.wikipedia.org/wiki/Armando> - [http://en.wikipedia.org/wiki/Armando\\_\(artist\)](http://en.wikipedia.org/wiki/Armando_(artist)).

Wat voor auteurs geldt, geldt ook voor onderwerpen. Men kan hiervoor een internationaal gezaghebbende database als de Library of Congress Authorities and Vocabularies (<http://id.loc.gov>) gebruiken. Het onderwerp van Armando's boek *Uit Berlijn* is de Duitse stad Berlijn. Bij de Library of Congress: ['Iedereen kan vrij eenvoudig een vocabulary online bekijken en controleren'](http://</a></p></div><div data-bbox=)



[id.loc.gov/authorities/classification/KKC5401-KKC5420.html](http://id.loc.gov/authorities/classification/KKC5401-KKC5420.html). Omdat Berlijn een geografische naam is, komt deze ook voor in VIAF: <http://viaf.org/viaf/122530980>. In DBpedia: <http://dbpedia.org/resource/Berlin>.

De vraag is of de door centrale gezaghebbende organisaties gecreëerde data als VIAF en ISNI betrouwbaarder zijn dan informatie die door iedereen kan worden gecreëerd, zoals ORCID en Wikipedia/DBpedia. Het is aannemelijk dat informatieprofessionals en hun instellingen voor wat betreft basismetadata meer vertrouwen hebben in databases die onder toezicht staan van erkende gezaghebbende instituten, maar dat ze geneigd zijn voor verkrijgende informatie het gebruik van 'crowd sourced' bronnen niet af te keuren. Dus auteur en onderwerp via VIAF of ISNI, extra toelichting en relaties uit DBpedia.

### SameAs-relaties

De verschillende bronnen voor het definiëren van basismetadata als personen en onderwerpen stellen informatiespecialisten en hun organisaties voor de keuze: aan welke moet de voorkeur worden gegeven? Het goede nieuws van linked open data is dat er geen keuze gemaakt hoeft te worden, omdat in principe alle relaties mogelijk zijn. Het is gebruikelijk dit te doen door middel van sameAs-relaties (<http://www.w3.org/TR/owl-ref/#sameAs-def>). Hiermee kan worden vastgelegd dat een object in een bepaalde database hetzelfde is als een object in een andere database, bijvoorbeeld:

<http://viaf.org/viaf/9885610>  
sameAs <http://isni-url.oclc.nl/isni/0000000120171892>

<http://viaf.org/viaf/9885610>  
sameAs [http://dbpedia.org/resource/Armando\\_\(artist\)](http://dbpedia.org/resource/Armando_(artist))

SameAs-relaties kunnen tegenstrijdige informatie over entiteiten in verschillende databronnen aan het licht brengen, hetgeen idealiter tot

een hogere kwaliteit zou moeten leiden. Informatieprofessionals kunnen hierin een rol spelen.

In VIAF worden links naar Wikipedia, ISNI en WorldCat Identities (<http://www.worldcat.org/identities>) gegeven. In Wikipedia komen ook VIAF en ISNI ID's voor. Uiteraard kunnen sameAs-relaties door iedereen in allerlei databases worden aangelegd.

### Provenance

Dit brengt ons bij de laatste te behandelen vraag uit de inleiding: klopt die door anderen gelegde relatie tussen twee objecten wel? Hiermee zijn we aangeland bij een belangrijk probleem: relaties in RDF hebben geen property's. In RDF kan geen informatie worden opgeslagen over herkomst ('provenance', door wie en wanneer?) van relaties. De betrouwbaarheid is dus niet zonder meer vast te stellen. Ook andersoor-

## 'Basismetadata worden vaak afgenomen van uitgevers en andere externe partijen'

tige elementaire informatie over relaties, zoals begin- en einddatum, kan niet worden vastgelegd. Dit is niet zo'n probleem bij auteurs van boeken, maar bijvoorbeeld wel bij tijdsgebonden relaties als koningschappen.

Er zijn verschillende oplossingen voor dit probleem bedacht, elk met zijn voor- en tegenstanders. Het gaat hierbij om het omzetten van relaties in aparte objecten met hun eigen property's, het opnemen van één of meer triples in een verzameling met eigen property's, of het gebruiken van de speciale PROV-ontology voor herkomstinformatie van individuele objecten en relaties.

Welke methode de voorkeur verdient is eigenlijk niet zo interessant voor informatieprofessionals. Dit is afhankelijk van de eisen die worden gesteld aan specifieke implementaties van linked open datasets en applicaties; daarover een volgende keer meer. Waar informatieprofessionals wel aandacht aan zouden moeten besteden, is hun rol bij het omgaan met nieuwe vormen van databeheer op het web, zoals Wikipedia. <

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

