



Linked Open Data

**Pilot
Linked Open Data
Nederland**

Deel 1 - Het Managementoverzicht

Pilot Linked Open Data Nederland

Deel 1 - Het Managementoverzicht



Linked Open Data

Colofon

Dit boek is tot stand gekomen vanuit de Pilot Linked Open Data. De partijen die hiernaast staan, hebben een bijdrage geleverd aan de totstandkoming van deze pilot.

Editors

Erwin Folmer
Marcel Reuvers
Wilko Quak

Review

Paul Brandt

Communicatie

Lian Pattje

Opmaak

remwerk, Amersfoort

Drukwerk

Klomp Reproka, Amersfoort

Oplage

400

Contact

e.folmer@geonovum.nl



Forum Standaardisatie



Ministerie van Infrastructuur en Milieu



Inhoud

Voorwoord	4
Introductie Pilot Linked Open Data	9
Data, Data, Data: Big, Linked & Open	15
Linked Open Data in Nederland: een snelgroeiend gewas in vruchtbare bodem	31
Artikelen Deel 2 De samenvattingen	38



Voorwoord

Dit boek is een resultaat van de Pilot Linked Open Data Nederland. Ik heb deze pilot mogen leiden en dat was een groot genot. Enerzijds omdat het onderwerp Linked Open Data uitermate boeiend is en een enorme potentie heeft. Maar bovenal door de grote groep enthousiastelingen die gewoon aan de slag gingen en de pilot tot een succes hebben gemaakt. Het grootste succes is misschien wel het opbouwen van het netwerk (de community), de nieuwe contacten die zijn opgedaan, en de enorme hoeveelheid kennisuitwisseling die heeft plaatsgevonden.

Dit boek (deel 1 en deel 2) is, ondanks de enorme rijkheid aan kennis, slechts een kleine neerslag van de kennisuitwisseling die heeft plaatsgevonden, en hopelijk blijft plaatsvinden.

Vooraf had ik wel wat bedenkingen, want ik kreeg een strak projectplan met vooraf gestelde resultaten en een strakke tijdsplanning. Alleen kreeg ik geen budget om een team van mensen in te schakelen om aan de resultaten te gaan werken. Dat zou de 'community' wel gaan doen. Jaja, natuurlijk. Hoe fout kon ik zijn. Zeker hebben we de nodige afslagen genomen ten opzichte van het projectplan, maar die kunnen we scharen onder 'hernieuwd inzicht'. En uiteindelijk hebben we mooie resultaten weten te bereiken.

Dan kom ik tot het moeilijkste aspect van dit voorwoord: het bedanken van zoveel mensen, dat ik niet weet waar ik moet beginnen en eindigen. Elke poging is gedoemd te mislukken en ik zal zeker mensen vergeten, die het wel verdienen om genoemd te worden. En toch ga ik het proberen.



Ik begin met Marcel Reuvers, de aanstichter; zonder hem was deze pilot er niet geweest! Uiteraard alle stuurgroepleden, en de 'founding' organisaties (zie de logo's in de colofon). Met hun financiële bijdrage was het mogelijk bijeenkomsten te organiseren, en onder meer dit boek uit te delen. Daarnaast alle trekkers die hard gewerkt hebben aan deelonderwerpen: Paul Geurts, Paul Francissen, Wilko Quak en Arjen Santema.

De organisaties die de ruimtes beschikbaar hebben gesteld (Belastingdienst, Brandweer Amsterdam-Amstelland, Gemeente Nijmegen, Geonovum, Kadaster, Rijksdienst voor het Cultureel Erfgoed, TNO), en alle presentatoren tijdens de bijeenkomsten, met daarbij een bijzonder woord van dank aan Frank van Harmelen voor het inspireren en ook beschikbaar stellen van VU-onderzoekers. Christophe Guéret dank voor het delen van je enorme hoeveelheid kennis en kunde, en het geven van cursussen, wat ook geldt voor Paul Hermans en Lieke Verhelst. En dan nog eenieder die

demonstrators gebouwd heeft, data heeft omgezet, servers in de lucht heeft gebracht, met daarbij een bijzonder compliment aan Marcel van Mackelenbergh voor zijn enthousiasme en inbreng. Voor de totstandkoming van dit boek wil ik alle auteurs en reviewers bedanken.

En nu heb ik nog vele personen niet genoemd, maar ik heb een oplossing gevonden om te voorkomen dat ik iemand vergeet: ik wil een ieder bedanken die op welke manier dan ook heeft bijgedragen aan deze pilot. Alle namen staan achterin dit boek.

Ik hoop dat dit project, het prachtige onderwerp Linked Open Data, èn de energie in de community, jullie verder inspireert en energie geeft om verder door te pakken en te blijven innoveren!

Erwin Folmer (Geonovum)
Projectleider Pilot Linked Open Data
Nederland

Introductie Pilot Linked Open Data





Introductie Pilot Linked Open Data

Auteur

Erwin Folmer (Geonovum)

Dit hoofdstuk beschrijft de Pilot Linked Open Data, vanaf de oorsprong, naar de uitvoering, de resultaten en het leven na de pilot.

De oorsprong van de pilot

Een grote groep partijen onderschrijft in 2012 het belang om onderzoek te doen naar de mogelijkheden om aan elkaar gerelateerde informatie uit verschillende overheidsregistraties uniform te ontsluiten op het web als open, dus vrij beschikbare, data. Het oorspronkelijke doel van de Pilot Linked Open Data (LOD) was om met een integrale aanpak, startend met data preparatie, het linken van data, tot aan de toepassing ervan, aan het werk te gaan en om al doende op basis van praktijkervaring inzicht te krijgen in de mogelijkheden en wensen. De verkregen inzichten kunnen vervolgens leiden tot algemene afspraken en afgesproken standaarden voor ontsluiting van open data op het web.

Open Data kent vele facetten. De scope van de Pilot lag echter op het linken van Open Data dat veelal gezien wordt als de vervolgstap ná het beschikbaar stellen van Open Data. Buiten de scope van deze pilot vielen de facetten rondom de juridische kant van Open Data. Het plan voor de Pilot is geïnitieerd door de observatie dat op meerdere plaatsen Open Data gepubliceerd wordt maar dat er een gemis is aan een gemeenschappelijk referentiekader voor de ontsluiting ervan.

Dit gemis is voelbaar aan de technische kant van standaarden, met als direct gevolg dat vanuit de overheid verschillende formaten worden gebruikt. Dit limiteert het gemeenschappelijk gebruik van Open Data. De Open Data van de overheid worden als afzonderlijke silo's ontsloten en links tussen overheidsverzamelingen en verzamelingen op het web kunnen niet worden gelegd, terwijl deze wel aanwezig zijn. Daarnaast leidt het ontbreken van dit gemeenschappelijke referentiekader tot drempelvrees voor overheden om tot publicatie van Open Data te komen.

Het concept Linked Data biedt veel potentie en stond daarmee centraal in de pilot. Daarbij zijn drie gezichtspunten gekozen in de pilot. Allereerst de kant van toepassingen, want daar draait het natuurlijk allemaal om. Wat zijn potentiële toepassingen die het nut van Linked Data inzichtelijk kunnen maken? Ten tweede, om die toepassingen mogelijk te maken is er een 'aanbod' nodig aan Linked Open Data. Het derde gezichtspunt betreft 'techniek'. Om te komen tot toepassingen en aanbod van Linked Open Data is het noodzakelijk dat er 'technische' kennis wordt opgedaan en gedeeld; niet vanuit de premisse dat één gedeelde technische oplossing een noodzakelijke voorwaarde is, maar juist om de ontkoppeling van Linked Open Data met techniek te stimuleren.

De uitvoering van de pilot

De uitvoering van de pilot lag nadrukkelijk bij de deelnemende partijen uit zowel de wetenschappelijke, publieke als private sector. Het was een pilot die voor iedereen was opengesteld voor deelname. Een bijzondere vermelding verdienen de partijen: Geonovum, Kadaster, Ministerie van Infrastructuur en Milieu, Forum Standaardisatie, KING, GeoBusiness Nederland, Interprovinciaal Overleg, Programma Stelsel van Basisregistraties, Gemeente

Amersfoort en de Gemeente Nijmegen. Deze partijen vormden de stuurgroep en zorgden voor de financiën, beslissingen, en data en waren daarmee de katalysator van het proces. De andere deelnemers van de Pilot Linked Open Data deden kosteloos mee en stelden uren van zichzelf of van de organisatie ter beschikking. Gedurende de pilot konden deelnemers aanhaken. Elke deelname was welkom, waarbij sommige deelnemers betrokkenheid toonden door aan een activiteit deel te nemen en te produceren, en anderen zich vooral tot doel stelden om inzicht te verkrijgen in Linked Open Data, dan wel om juist als kennispartner te fungeren.

In september 2012 was het project uit de startblokken gevlogen met een inspirerende bijeenkomst te Amersfoort. Meer dan 100 aanwezigen hebben zich in verschillende groepjes verdeeld om verder te werken aan de drie genoemde gezichtspunten. Dat betekende het startschot van een levendige community voor dit onderwerp. Deze community is vaak bij elkaar gekomen, zoals tijdens de zestal plenaire bijeenkomsten (ook gehouden bij de deelnemers: Belastingdienst, TNO, Gemeente Nijmegen, Kadaster, Brandweer Amsterdam-Amstelland), tijdens de vele onderwerp-specifieke bijeenkomsten, maar tevens bij de

Conceptual Fridays © Geonovum. Daarnaast waren de leden ook actief op LinkedIn (LOD Nederland). We mochten daarbij ook rekenen op de steun van Vrije Universiteit die op basis van vouchers drie onderzoekers beschikbaar stelde. De resultaten van de pilot zijn ook internationaal uitgedragen: onder andere bij Google London tijdens de Open Data on the Web workshop van het W3C, en op het European Data Forum te Dublin.

Wat heeft de pilot opgeleverd?

Vanuit het gezichtspunt van de toepassingen is snel een keuze gemaakt voor twee specifieke toepassingen: 'Monumenten' en 'Huiskluis'. De groep rond Monumenten, onder aanvoering van de Gemeente Nijmegen, hield zich niet alleen bezig met de ontwikkeling van een demo app op basis van data omtrent monumenten, maar ook met het delen van ervaringen rond de kloof tussen het aanbod van Linked Open Data en de mogelijkheid tot het gebruik ervan in nuttige apps. De tweede toepassing betrof de 'Huiskluis'; een toepassing die eenvoudig uit te leggen is als een centrale plek om informatie over een huis te presenteren. Zou het niet prettig zijn als je in één oogopslag alle gegevens van je huis kan inzien? Op basis van Linked Data kan je eindeloos informatie blijven toevoegen aan de huiskluis. Vanuit het gezichtspunt van het

aanbod is door een groep gewerkt om datasets als Linked Open Data beschikbaar te stellen voor de demo toepassingen. Daarnaast is er gewerkt aan een architectuur voor het aanbieden van Linked Open Data. Tot slot de technische groep: deze heeft zich beziggehouden met een verscheidenheid aan technische uitdagingen hoe datasets ontsloten en toegepast kunnen worden.

Veel van deze resultaten zijn ook terecht gekomen in het boek over deze pilot. In hoofdstuk 5 van dit eerste deel kunt u de samenvattingen hiervan lezen, terwijl in deel 2 (of via Internet) de volledige tekstbijdrages beschikbaar zijn. En dat sluit dan ook aan bij misschien wel het belangrijkste resultaat van de Pilot Linked Open Data: de kennisuitwisseling. De energie die vloeide tijdens, en gedurende deze productie voortvloeit uit, deze kennisuitwisseling was en is een fantastisch resultaat. Dat begon al met de inspirerende sprekers tijdens de bijeenkomsten. Gedenkwaardig waren de cursussen die gegeven werden door de experts tijdens de bijeenkomst bij de Belastingdienst.

Vanuit de pilot zijn interviews afgenomen, en deze zijn gepubliceerd in nieuwsbrieven over de pilot. Ook deze zijn opgenomen in deel 2:

Interview 1

Joop

Vanderheiden en
Bart Broex van de
Rijksdienst voor het
Cultureel Erfgoed



Interview 2

Frank van

Harmelen van de
Vrije Universiteit



Interview 3

Hayo Schreijer
van het Kennis- en
exploitatiecentrum
voor Officiële
Overheids
Publicaties (KOOP)



Interview 4

Willem Melder van
het Nederlands
Instituut voor Beeld
en Geluid



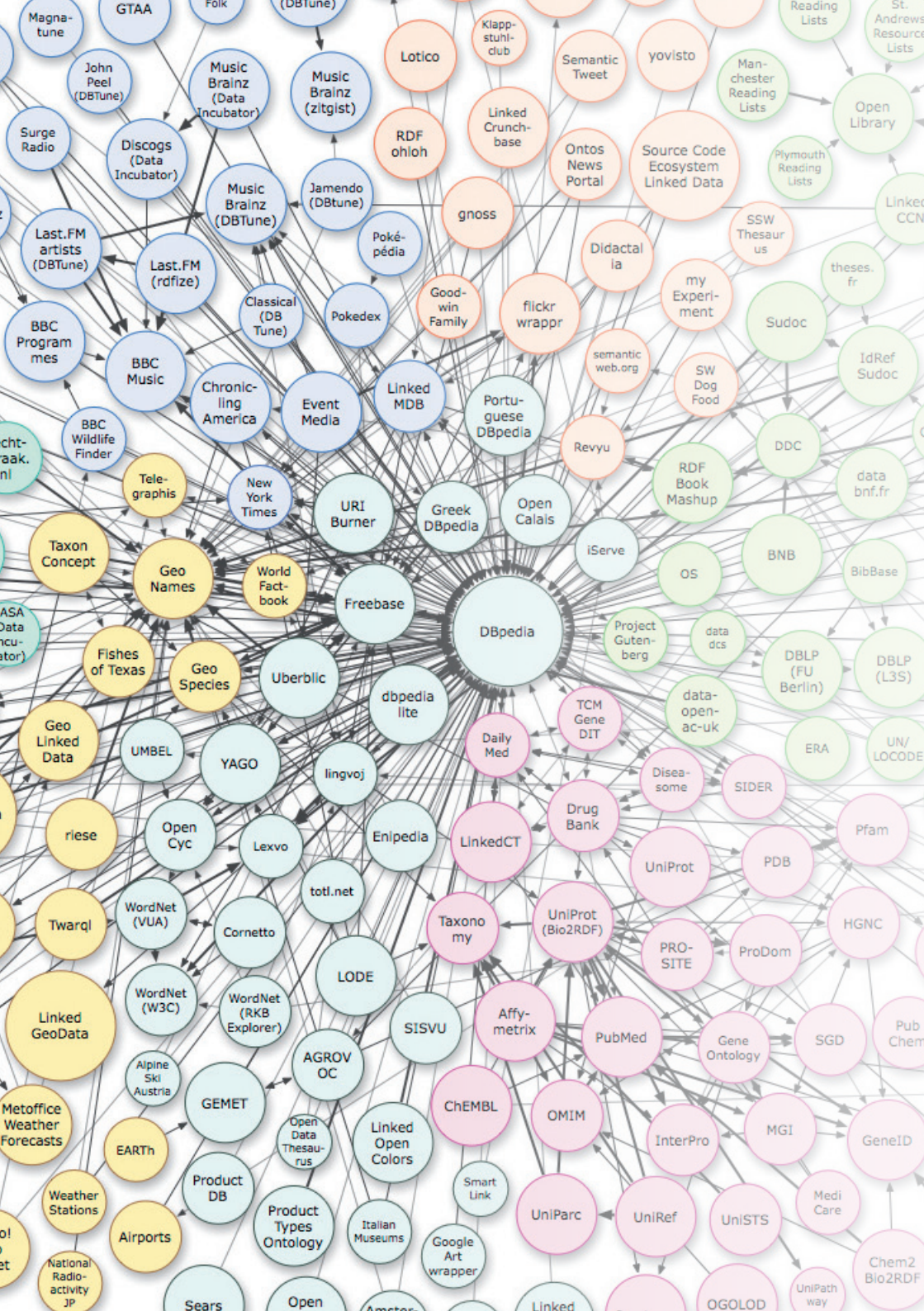
De resultaten uit de pilot zijn aangevuld met bijdrages van buiten de pilot. Samen vormen ze een inhoudelijk rijk en inspirerend boekwerk op het gebied van Linked Open Data: het eerste in zijn soort in Nederland. We zien het dan ook als een mooi sluitstuk om dit werk uit te delen op de laatste kennisdelingsactiviteit van deze pilot: het slotevenement op 3 juli bij de Rijksdienst voor het Cultureel Erfgoed in Amersfoort.

Is er leven na de pilot?

Jazeker. Deze pilot was een startschot en dat betekent dat er nog vele stappen gezet moeten worden voordat Linked Open Data breed toegepast gaat worden. Maar het was wel een noodzakelijk startschot. Nu is het tijd dat er doorgepakt wordt, want de potentie van Linked Open Data is glashelder. De community is deze uitdaging aangegaan, en pakt het nu verder op. Dus u hoort nog van ons!

Data, Data, Data: Big, Linked & Open





Data, Data, Data: Big, Linked & Open

Auteurs

Erwin Folmer (Geonovum)

Dennis Krukkert (TNO)

Silja Eckartz (TNO)

De gehele business en IT-wereld praat op dit moment over Big Data, een trend die medio 2013 Cloud Computing is gepasseerd (op basis van Google Trends). Ook beleidsmakers houden zich actief bezig met Big Data.

Neelie Kroes, vice-president van de Europese Commissie, spreekt over de 'Big Data Revolution' en de verandering naar een 'Data-driven Economy'; waarbij data als de nieuwe olie worden beschouwd die tientallen miljarden euro's waard zijn. [\[1\]](#)

Deze Big Data Revolutie wordt ook uitstekend beschreven in het boek 'De Big Data Revolutie', waarin big data wordt beschreven als bron van economische waarde en innovatie: 'Gegevens kunnen op een slimme manier worden hergebruikt en zo uitgroeien tot een rijke bron van innovaties en nieuwe diensten'. Een aspect daarbij is de gigantische toename van data en

de noodzakelijke machines, maar de echte revolutie zit in de gegevens zelf en de gebruikswijze. [\[2\]](#)

De bedrijvigheid en innovatieve aspecten worden breed gedragen, en in een overheidscontext vaak ook gerelateerd aan Open Data in het publieke domein. Zoals Kroes het heeft over 'Unlocking this gold mine' en 'Opening up public data means opening up business opportunities'. Maar ondanks de stevige uitspraken van Kroes is Europa geen koploper.

In mei 2013 heeft Obama een nieuwe wet voor Open Data ondertekend: 'And today I'm announcing that we're making even more government data available, and we're making it easier for people to find and to use. And that's going to help launch more start-ups. It's going to help launch more businesses... It's going to help more entrepreneurs come up with products and services that we haven't even imagined yet. This kind of innovation and ingenuity has the potential to transform the way we do almost everything.' Daarbij gaat Obama verder dan alleen het open publiceren: een open licentie, machine readable, en in een open formaat. Daarnaast zal de data gecombineerd

en gerelateerd moeten worden voor toepassingen. [3]

Ook het OECD heeft een rapport [4] gepubliceerd over Big Data, of eigenlijk de economische mogelijkheden van de nieuwe data society. Daarin wordt een vijftal toepassingen of trends beschreven:

- Data-driven R&D: verbetering in onderzoek door data toepassing
- Data(-intensive) products: het ontwikkelen van nieuwe producten/diensten waarin data het product is (of component)
- Data-driven processes: optimaliseren van productieprocessen
- Data-driven marketing: meer gerichte advertenties en gepersonaliseerde aanbevelingen
- Data-driven organisation: ontwikkelen van nieuwe benaderingen voor het management in organisaties

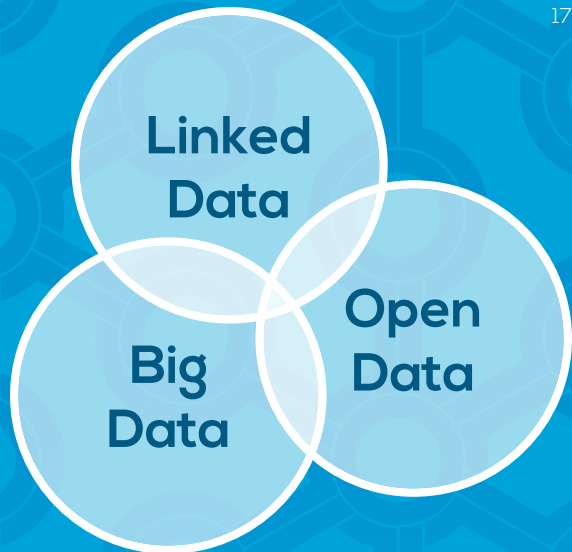
Maar is dit nu Big Data? De terminologie in ICT is al jaren een hot issue; termen en trends met spannende afkortingen zoals een SOA, volgen elkaar in rap tempo op. Zoals gezegd in 2013 is Big Data booming. Linked Open Data is een term die al een lange geschiedenis heeft en al jaren relatief stabiel is, maar het is sterk gerelateerd aan de trend Big Data, en uiteraard ook aan Open Data.

In het kort samengevat gaat Big Data over de slimme toepassing van data in alle soorten en maten die tot op heden nog niet mogelijk waren. Eigenlijk is de term 'big' dus misleidend. We hadden het beter gewoon 'Data, Data, Data' kunnen noemen zoals in de Roadmap ICT; dat dekt beter de lading. [5]

Regelmatig worden ook andere termen geïntroduceerd om de lading beter te dekken. Zoals Small Data [6], Long Data [7] of Smart Data [8], en voor specifieke toepassings-domeinen zoals Linked Enterprise Data. Maar het gaat om data, data, data.

De viewpoints op 'Data, Data, Data'

Er zijn vele 'viewpoints' mogelijk op 'data', maar de drie belangrijkste zijn toch wel Big, Open en Linked; deze overlappen elkaar ook. Op de volgende pagina's worden de viewpoints nader beschreven. De drie viewpoints hebben overlap, en waar ze elkaar overlappen wordt het vaak interessant. Vaak gaan initiatieven van Linked Data over vrij beschikbare (open) data, en spelen issues gerelateerd aan het Big viewpoint ook een rol.



Big Data

De benaming 'Big' Data view mag enigszins verwarrend overkomen, want in de vorige paragraaf lieten we zien dat de hype Big Data, niet heel specifiek over 'Big' gaat. Echter binnen de ruime hype term (die wij liever data, data, data noemen) is er zeker wel een onderwerp dat specifiek over 'Big' gaat. Meestal worden dan de drie V's (volume, velocity, en variety) gebruikt om aan te tonen dat het 'Big' is, ook al zijn er geen eisen aan bijvoorbeeld het Volume gedefinieerd voordat we over 'Big' zouden moeten praten. Maar als je aan een informaticus vraagt wat Big Data is dan komt hij met een krappere, technischer antwoord: Big data bestaat uit datasets die te groot zijn om met reguliere databasemanagementsystemen te onderhouden. Denk hier bijvoorbeeld aan sensor data of data streams.

Hier wordt ook de term NoSQL voor gebruikt om uit te drukken dat het om ongestructureerde data in schema-loze databases gaat. Voor een informaticus spelen dan ook data mining en data analytics een heel belangrijke rol voor big data. Verschillende open source softwaremodellen ondersteunen data-intensieve gedistribueerde applicaties. Een voorbeeld is MapReduce, een raamwerk voor het verwerken van problemen over enorme datasets met behulp van een groot aantal computers (nodes), ook cluster of grid genoemd. Een populaire vrije implementatie is Apache Hadoop. Hierbij gaat het ook over process hosting; waar worden de data-analyseprocessen uitgevoerd (bijvoorbeeld Amazon EC2)? En waar staat de data (cloud computing, distributed hosting (bijvoorbeeld Amazon S3)?

Big Data

Open Data

'Open' Data zijn datasets die met een open licentie beschikbaar worden gesteld zodat toegang en hergebruik zonder beperkingen mogelijk is. De voorwaarden waaronder de data beschikbaar zijn, worden beschreven in licenties en gebruiksvoorwaarden. Het idee van open data is om de beperkingen in hergebruik tot een minimum te limiteren. Hierdoor wordt het delen en hergebruik van data bevordert. Vaak wordt data van de overheid op deze manier gedeeld om transparantie te vergroten en economische activiteit te bevorderen. Verder geeft open data derde partijen de kans om met data te innoveren.

Veelal wordt ook onder open data verstaan dat data (indien relevant) 'machine readable' moet zijn en in een open formaat beschikbaar moet worden gesteld. In principe kunnen tekstdocumenten ook geprint worden, gescand worden en in jpg als open data beschikbaar worden gesteld. Het doel van open data, hergebruik, wordt dan wel lastig. Hetzelfde geldt voor data die beschikbaar worden

Open Data

gesteld in een gesloten formaat dat alleen met dure software is in te lezen; ook dat belemmert het hergebruik. Vandaar: open licentie, machine readable, en open formaat. Tot slot wordt regelmatig vergeten dat onderhoud, kwaliteit en governance op de open data ook belangrijke aspecten zijn. Datasets kunnen snel verouderen, dus alleen eenmalig op Internet zetten is niet voldoende. Wat is de kwaliteit, en wie bepaalt welke correcties wanneer doorgevoerd gaan worden?

Linked Data

Linked Data

'Linked' Data is een essentieel onderdeel van het Semantische Web. Door Tim Berners-Lee in 2006 al beschreven als een component van 'Web 3.0', een trend waarbij Internettoepassingen goed op elkaar afgestemd en soms zelf geïntegreerd kunnen worden. De ontwikkeling van het Internet wordt dan als volgt beschreven: van het web van documenten (Web 1.0) via Web 2.0 waar het Internet als interactief communicatiemedium beschouwd wordt en gebruikers informatie kunnen uploaden naar Web 3.0: het web van Linked Data, waar Internettoepassingen en data bijvoorbeeld via webservices aan elkaar gelinkt kunnen worden. Deze slimme toepassingen kunnen dan de links volgen tussen datasets. Dit is de basis van het semantic web.


Linked Data, Semantic Web, Web 3.0 worden weleens als synoniemen gebruikt, alhoewel specifiek Linked Data gebruikt wordt als term voor een methode voor het publiceren van data in een structuur zodat het linkbaar wordt en daarmee bruikbaar.



Linked Open Data

Linked Open Data is een praktische manier om een bijdrage te leveren aan het Semantische Web. Semantisch wil zoveel zeggen als 'de betekenis lerend'. Het semantisch web zou je grofweg kunnen definiëren als een web van verbanden. Verbanden gelegd tussen informatie op het internet, waardoor nieuwe inzichten kunnen ontstaan. Informatie gepubliceerd als Linked Open Data stimuleert het hergebruik van je data, omdat je zelf zoveel mogelijk verwijzingen aanbrengt naar kennisbronnen elders en omdat anderen gemakkelijk naar jouw informatie kunnen verwijzen.

Als je in een willekeurige webbrowser naar 's Gravenhage zoekt, vind je geen resultaten waarin 'Den Haag' voorkomt, terwijl beide woorden naar dezelfde stad verwijzen. Dat komt doordat webdocumenten met elkaar verlinkt zijn, maar de inhoud zelf niet. Een zoekmachine kan zo alleen maar op woorden zoeken. Linked Data biedt een oplossing voor dit probleem door woorden als concepten uniek te maken en te beschrijven in één of liefst meerdere subject – predicaat – objectrelaties. Een stad wordt daarmee een concept en kan meerdere attributen krijgen, waarvan elk attribuut ook weer een eigen concept is. Zowel subject, als predicaat en object zijn dus op zichzelf weer unieke concepten. Elk concept wint aan betekenis naarmate er meer beschrijvingen aan gelinkt worden. Op deze manier wordt de inhoud



van webdocumenten betekenisvol en worden zoekresultaten nauwkeuriger. Uiteindelijk bereik je zo taalonafhankelijkheid omdat het dus niet meer uitmaakt of je naar 'The Hague' of 'Den Haag' zoekt.

Zoals gezegd staat 'Linked Data' voor het idee dat informatie (resources) op het web aan elkaar gelinkt wordt zodat je informatie maar één keer hoeft te beschrijven, en er vervolgens vanuit verschillende bronnen naar kunt verwijzen. Open data gaat over het ontsluiten van data op een zodanige manier dat eenieder die data kan gebruiken zonder vast te zitten aan licenties die het gebruiksrecht beperken. Als we beide combineren dan komen we op Linked Open Data, oftewel: data die via het web op een open manier aangeboden wordt, en die onderling verbonden is.

Het belang van Linked Data

Om nauwkeurig te kunnen zoeken in de enorme hoeveelheid informatie op het web is zoeken op basis van betekenis onontkoombaar. Het semantisch web is een verzamelnaam voor technieken die computers in staat stellen de betekenis van de informatie op het web te begrijpen zónder menselijke tussenkomst. Een complicerende factor hierbij is dat de betekenis van mensen, dingen, gebeurtenissen etc. niet constant is, maar kan variëren. Zo heeft de koningin van Nederland een wisselende betekenis die o.a. afhankelijk is van de tijd: in 2010 verwijst het begrip naar Beatrix, maar in 1970 was het Juliana, en in 2013 naar Willem-Alexander. Menselijke informatieverwerkers zijn gewend om contextuele factoren, zoals tijd, mee te nemen bij het toekennen

van betekenis. Voor machines geldt dit niet. Om computers toch in staat te stellen de juiste betekenis toe te kennen, is het aanbieden van relevante context van groot belang. Linked Data is een techniek om machine-leesbare context te genereren.

We sluiten deze paragraaf af met een quote uit het OECD-rapport: 'The linking and use of data across sectors drive innovation, socioeconomic development and growth.' [4]

Linked Open Data concepten

Om data op een betekenisvolle manier aan elkaar te linken wordt gebruik gemaakt van 'semantic web' technologie. Dit is een verzamelnaam voor verschillende standaarden en technologieën die gebruikt kunnen worden om te komen tot een web van 'Linked Open Data' die op een betekenisvolle manier gebruikt (en door computers geïnterpreteerd) kan worden.

De basis voor het semantic web is RDF (Resource Description Framework). Een resource kan van alles zijn: een persoon (bijvoorbeeld Erwin, een auteur van dit hoofdstuk), een organisatie (Geonovum, een werkgever van Erwin), een boek (bijvoorbeeld dit boek), een artikel, etc. Elke resource heeft een unieke URI in de vorm van een URL. Op de locatie van deze URL wordt een representatie van de resource aangeboden.

Het linken van data gaat via zogenaamde 'triples'. Een triple is een getypeerde relatie tussen twee resources, en bestaat uit een subject, een predicate en een object. Hiermee kun je bijvoorbeeld uitdrukken Erwin (subject) werkt voor (predicate) Geonovum (object). Erwin en Geonovum zijn beide resources waarnaar verwezen wordt middels een URI. Van het boek Pilot Linked Open Data kan ook een resource gemaakt worden. Het is nu heel eenvoudig om aan te geven wie de auteur van dit boek is door een triple toe te voegen: Pilot Linked Open Data (subject) is geschreven door (predicate) Erwin (object). Omdat elke keer verwezen wordt naar dezelfde resource hoeft deze maar één keer beschreven te worden.

Het hergebruik van data wordt pas echt waardevol door niet alleen data te linken, maar door ook betekenis toe te voegen aan de link. Oftewel: door expliciet vast te leggen wat de relatie (de 'predicate') 'werkt voor' betekent. Ook hiervoor wordt de hetzelfde mechanisme gebruikt: de relatie wordt beschreven in RDF-formaat, en ontsloten via een URI.

In aanvulling op RDF kan RDFS gebruikt worden. Met behulp van RDFS kunnen 'klassen' van resources aangemaakt worden, en tevens beperkingen gelegd worden op de verschillende relaties die mogelijk zijn tussen instanties van deze klassen. Hiermee zou je bijvoorbeeld vast kunnen leggen dat de relatie (lees: predicate) 'is getrouwd met' alleen gelegd kan worden tussen twee resources van het type 'persoon'.

Gelukkig hoeven we het wiel niet opnieuw uit te vinden. Er zijn al veel standaard definities beschikbaar van resource- en relatietypes. Deze zijn vastgelegd in zogenaamde ontologieën. Een ontologie is, eenvoudig gesteld, een neerslag van concepten en hun onderlinge verbanden die door een groep betrokkenen gezien wordt als representatie van hun gedeelde werkelijkheid. Afhankelijk van de toepassing kan een ontologie daarmee de vorm krijgen van bijvoorbeeld een definitielijst, een

RDF-graph, een UML-model, een taxonomie, een zuiver logisch model en vele andere. Voorbeelden van veelgebruikte ontologieën zijn SKOS (Simple Knowledge Organization System) voor het opstellen van vocabulaires en thesauri, en FOAF (Friend-of-a-friend). SKOS bevat zaken als 'broader', 'narrower' en 'alternative label'. FOAF bevat zaken als 'Person' en 'email'.

RDF beschrijft het algemene principe achter triples, en schrijft bijvoorbeeld het gebruik van URI's voor. Het technische formaat (de syntax) waarin triples worden vastgelegd en uitgewisseld wordt echter vrijgelaten. Hiervoor zijn verschillende technische formaten beschikbaar zoals RDF/XML, Turtle, N3 en JSON. Met elk van deze formaten kunnen triples vastgelegd worden. De formaten verschillen enerzijds in syntax, en anderzijds in een aantal toevoegingen bovenop RDF. Triples kunnen ook opgeslagen worden in een database, ook wel 'triple store' genoemd. Om een triple store te bevragen is een speciale taal ontwikkeld (analoog aan SQL voor relationele database): SPARQL (SPARQL Protocol and RDF Query Language).

De vier principes van Linked Open Data

Linked Data moet aan vier principes (vrij naar Tim Berners Lee) voldoen [1]. Deze principes zijn:

Principe 1

Gebruik URI's om dingen te identificeren.

Principe 2

Gebruik HTTP-URI's zodat er naar deze dingen kan worden verwezen en dat ze kunnen worden opgezocht door zowel mensen als machines.

Principe 3

Leg de informatie over het concept vast in een 'triple' (subject-predicaat-objectrelatie), leg die triple vast en maak het beschikbaar op basis van standaarden zoals RDF en SPARQL.

Principe 4

Neem links naar andere, gerelateerde, open-dataconcepten op in de beschrijving om het ontdekken van gerelateerde informatie op het web te verbeteren.

Een URI betekent in feite het toekennen van een unieke string om data-objecten uniek identificeerbaar te maken. Door middel van HTTP-URI's wordt er verwezen naar een

unieke plek op het internet waardoor informatie over het object vindbaar wordt. Linked Open Data zijn datasets in RDF-formaat, met URI's en met links tussen de datasets.

De weg naar Linked Open Data: het vijfsterrenmodel

Organisaties zullen vaak niet direct met Linked Open Data beginnen. Vaak worden er verschillende stappen gezet voordat men kan spreken over Linked Open Data. De eerste stappen betreffen veelal puur het open maken van Data. Dit heeft Tim Berners-Lee verwerkt in het vijfsterrenmodel van Linked Open Data [12]. De vijf sterren in dit model zijn hieronder beschreven.

De eerste drie sterren betreffen open data, en pas bij de laatste twee sterren spreken we van Linked Open Data. In 2013 maken leiders zoals Kroes en Obama duidelijk een beleid gericht op drie sterren, terwijl in het verleden één ster veelal voldoende was. Als die lijn zich doorzet, zitten we binnen een paar jaar op vijf sterren (overheids)data!



De informatie is beschikbaar op het internet, in welk formaat dan ook.



De informatie is online beschikbaar in een gestructureerd formaat, dat geschikt is voor automatisch hergebruik (zoals Excel in plaats van een plaatje van een tabel).



De informatie is online beschikbaar in een open bestandsformaat (zoals CSV in plaats van Excel).



Al het bovenstaande, en bovendien wordt gebruikgemaakt van de open standaarden Resource Description Framework (RDF) en SPARQL, zodat anderen makkelijk naar de dataobjecten kunnen verwijzen.



Al het bovenstaande, en bovendien wordt er naar data van anderen verwezen voor meer context van de data.

Referenties

[1] Statements Kroes:
<http://bit.ly/19bpC0q>



[5] Roadmap ICT:
<http://bit.ly/ShQGE4>



[2] Boek Big Data: Viktor Mayer-Schönberger & Kenneth Cukier (2013), De Big Data Revolutie, Hoe de data-explosie al onze vragen gaat beantwoorden, Maven Publishing.



[6] Small Data:
<http://bit.ly/17O3Z88>



[7] Long Data:
<http://bit.ly/Wd9FCK>



[3] Statements Obama:
<http://bit.ly/197zCG2>



[8] Smart Data:
<http://bit.ly/ZUgNoV>



<http://1.usa.gov/15NKxby>



[9] Boek Linked Data: Tom Heath & Christian Bizer (2011), Linked Data, Evolving the Web into a Global Data Space, Morgan & Claypool Publishers.
<http://bit.ly/13Qsz2W>



[4] OECD rapport: OECD (2013), 'Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data'', OECD Digital Economy Papers, No. 222, OECD Publishing.
<http://bit.ly/14DIB2E>



[10] Boek Linked Open Data: Florian Bauer & Martin Kaltenböck (Semantic Web Company) Linked Open



Data: The Essentials – A Quick Start Guide for Decision Makers. De eerste drie hoofdstukken uit dit boek zijn opgenomen in deel 2.

Het volledige boek is beschikbaar als PDF op <http://bit.ly/AC4EIz>

[11] Vier principes van Linked Open Data: <http://bit.ly/FOVNV>



[12] Het vijfsterrenmodel: <http://www.5stardata.info>



Linked Open Data in Nederland: een snelgroeiend gewas in vruchtbare bodem





Linked Open Data in Nederland: een snelgroeiend gewas in vruchtbare bodem

Auteur

Frank van Harmelen
(Netwerk Instituut, Vrije
Universiteit Amsterdam)

Deze bundel is in zekere zin een mijlpaal. Er zijn al talloze internationale publicaties verschenen over Linked Open Data maar er is geen enkele die zo uitgebreid verschillende aspecten van Linked Open Data behandelt, variërend van technische principes en praktische toepassingen tot maatschappelijke en zelfs filosofische beschouwingen.

In deze korte beschouwing geef ik een kort synthetisch overzicht van de inhoud van deel 2 van dit boek. Dat overzicht leidt tot een aantal conclusies over de stand van Linked Open Data in Nederland, en die conclusies zijn voornamelijk optimistisch van aard. Tot slot probeer ik naar de korte termijn toekomst te kijken, om te bepalen welke volgende stappen nu het meest nodig zijn.

Observaties over dit boek

Diversiteit in Toepassingen

Het eerste dat elke lezer zal opvallen, is de grote diversiteit in de toepassingen van Linked Open Data die in dit boek besproken worden. Geurts, Brentjens en Van Aart maken informatie beschikbaar over talloze monumenten voor de gemeente Nijmegen, Francissen en Van Echtelt ontsluiten kadastrale gegevens, van Leeuwen schrijft uit eigen ervaring met zijn co-auteurs Pattje en Folmer over het nut van open data voor hulpverleningsdiensten als de brandweer, en Salters illustreert haar

betoog met behulp van bestuurlijke informatie.

Het brede scala aan toepassingen wordt voortgezet door het consortium van Hamers, Van der Arend, Verhoeff, Nijstad en Van Vuuren die onderwijskundige kennis openbaar maken met Linked Open Data, met de bijdrage van Walker en Nelissen betreft productdata voor NXP, en sluit af met Vanderheiden, Hendriks en Verbeij die laten zien hoe het Nederlands cultureel erfgoed baat heeft bij het publiceren van een gemeenschappelijk vocabulaire dat vervolgens gekoppeld wordt aan diverse bronnen. Dit alles beslaat een zeer brede waaier aan informatie uit de publieke, semi-publieke sector en private sector, en laat goed zien hoe breed de impact van Linked Open Data kan zijn.

Diversiteit van samenwerkende partijen

Naast de diversiteit in toepassingen is ook opvallend hoe heteroog de teams zijn die voor deze toepassingen samenwerken. Een gemeentebestuur werkt samen met een hightech startup, een brandweerman werkt samen met een adviesorganisatie, onderwijskundigen en erfgoeddeskundigen werken samen met een ICT-consultant. Andere bijdragen in deze bundel komen van technologieleveranciers, van universitair onderzoekers, van het Kadaster, van medewerkers bij de landelijke overheid etc. Zo'n diversiteit van achtergronden laat zien dat Linked Open Data zich bevindt in een heel rijk 'ecosysteem' van partijen. En zo'n rijk ecosysteem is zeer gunstig voor de snelle absorptie van nieuwe technologie in de praktijk.

Overzicht van essentiële basistechnologieën

Naast een mooie bloemlezing van Linked Open Data toepassingen, is deze bundel ook een zeer nuttig overzicht van allerlei essentiële basistechnologieën. In de delen C en D komen vrijwel alle essentiële elementen van Linked Open Data aan bod: het publiceren van legacy data (Hermans), de W3C-principes (Guéret), het ontwerpen van semantische modellen (Verhelst, en elders Stap), het aanbrengen van links tussen datasets (Guéret), het gebruik van triple-stores (Aasman en Stap), een strategie voor URI-ontwerp (Van den Brink, Overbeek en Brentjens), de keuze tussen verschillende dataformaten (Brentjens, Reuvers en Portele), de te gebruiken netwerkprotocollen (idem), een procesmodel voor het publiceren en beheren van Linked Open Data (Van den Broek, Van Veenstra en Folmer), en richtlijnen voor het publiceren van ruimtelijke data (Portele, en ook Van den Brink, Janssen en Quak) en statistische data (Van Hage). Niet onbelangrijk is dat deze bundel hiermee de eerste (deels) Nederlandstalige beschrijving is van een zo compleet arsenaal aan technologieën van Linked Open Data.

Filosofische en ideologische beschouwingen

Zoals elke nieuwe, ver reikende technische vernieuwing is Linked Open Data niet alleen maar technologie. Het is een combinatie van technologie, ideologie en filosofie. Ook die filosofische aspecten komen aan bod. En dat is goed, en zelfs noodzakelijk. Behalve overeenstemming over technische specificaties voor formaten en protocollen is ook een gedeeld 'wereldbeeld' (of zelfs: een gedeelde filosofie, een gedeelde ideologie) noodzakelijk als grondslag voor daadwerkelijke maatschappelijke vernieuwing. Een interessant vergezicht naar de toekomst is de bijdrage van Geerdink en Verhoeven, die handelt over de gevolgen van Linked Data voor de organisatievorm van ondernemingen, leidend tot wat ze noemen de 'extended enterprise'. De bijdrage van Van Rijn en Santema is een soortgelijke beschouwing, maar nu juist gericht op de overheid. En bijna filosofisch van aard is de bijdrage van Van Mackelenbergh en Hoekstra, die nieuw Linked Open Data licht laten schijnen over aloude informatie-filosofische vragen: is informatie te begrijpen en gebruiken zonder context? Kan informatie überhaupt hergebruikt worden in een andere context dan waarin ze gemaakt werd? En evenzo bespreekt de bijdrage van Brandt en Daniele een andere filosofische

klassieker die raakt aan de kern van Linked Open Data: de balans tussen formele correctheid en pragmatische werkbaarheid. Zoals gezegd is dit niet slechts filosofische luchtfietserij. Overeenstemming over zulke ontastbare zaken is net zo van belang als overeenstemming over concrete formaten en protocollen.

Uitdagingen voor de nabije toekomst

Uit deze bundel blijken ook enkele concrete uitdagingen die in de nabije toekomst zullen moeten worden opgelost om Linked Open Data tot volle bloei te laten komen. De belangrijkste twee daarvan zijn mijns inziens: bredere kennisoverdracht, en verlaging van technische drempels.

Deze bundel levert een belangrijke bijdrage aan de kennisoverdracht van de eerstelijns pioniers naar een bredere groep toekomstige gebruikers en ontwikkelaars. Maar zoals elke publicatie zal ook deze bundel slechts een beperkte reikwijdte hebben. We zullen nog hard moeten werken aan kennisoverdracht naar brede groepen in de maatschappij: ontwikkelaars, gebruikers, beleidsmaker, bij de overheid (variërend van lokaal tot provinciaal en landelijk), bij de semi-overheid, en bij het bedrijfsleven, variërend van ZZP'ers en MKB tot multinationals. Er ligt hier een duidelijke taak voor organisaties

in het hoger onderwijs, maar er ligt ook een grote markt voor commerciële partijen. Zulke brede kennisoverdracht is nodig als Linked Open Data gemeengoed wil worden, op hetzelfde niveau als bijvoorbeeld databases of web-technologiedata zijn.

De tweede uitdaging ligt mijns inziens in het verlagen van technische drempels. Enerzijds is het mooi om te zien dat zoveel verschillende partijen in deze bundel rapporteren over het succesvol gebruik van Linked Open Data technologie. Anderzijds betreft het hier wel de pioniers in Nederland, en is het aantal mensen dat in staat is om zulke toepassingen te bouwen in Nederland beperkt tot enkele honderden (en dat lijkt me een optimistische schatting). Met andere woorden: Open Data is nu toch nog teveel een kunst die slechts door weinigen verstaan wordt. Natuurlijk is kennisoverdracht een manier om dat aantal te vergroten, maar nog beter is het om ervoor te zorgen dat veel van die kennis niet meer nodig is, doordat de techniek beter verpakt wordt in laagdrempelige tools en applicaties. Ook op dat gebied is nog veel te winnen.

Conclusie

Al met al levert deze bundel een zeer bemoedigend beeld op over de toekomst van Linked Open Data in Nederland.

Ten eerste is het goed om te zien dat technologie niet langer de beperkende factor is. Er zijn, mede door de bijdragen in deze bundel, nu heldere (deels Nederlandstalige) uiteenzettingen over vrijwel alle technologische ingrediënten. En wat belangrijk is: die uiteenzettingen zijn niet geschreven door de oorspronkelijke initiators, bedenkers en uitvinders van Linked Open Data, maar juist door de eerste generatie ontwikkelaars en gebruikers ervan. En die eerste generatie pioniers, die de bijdragen aan deze bundel hebben geschreven, zijn veel beter in staat om die kennis op de juiste wijze door te geven aan de veel grotere groep gebruikers en ontwikkelaars die klaar staan om hun voorbeeld te volgen. Het is ook een goed teken dat de bundel zelfs een stuk voor beleidsmakers en beslisser bevat (de bijdrage van Bauer en Kaltenböck). Als er eenmaal stukken voor beleidsmakers verschijnen, is dat een teken dat het met de basistechnologie wel goed zit.

Ook bemoedigend is om te zien wat er niet in de bundel staat. Er wordt (gelukkig) maar zeer weinig geschreven over problemen bij het verzamelen van data, en ik lees dit als een signaal dat het met de beschikbaarheid van die ruwe grondstof, de data, in Nederland in orde is. Nederland is in een bijzonder goede positie vanwege het relatief goed georganiseerde datalandschap in de publieke en semipublieke sector. Dat is dan ook de reden voor de titel van deze korte beschouwing. Linked Open Data in Nederland is een snelgroeiend gewas, dat in de vruchtbare bodem van het Nederlandse datalandschap snel tot bloei kan komen.

Artikelen Deel 2

De samenvattingen





Artikelen Deel 2

De samenvattingen

We eindigen deel 1 van deze publicatie met een overzicht van alle bijdrages die in deel 2 zijn opgenomen. We hopen dat deze samenvattingen uitnodigen om het volledige paper te gaan lezen. Dat kan enerzijds door deel 2 (het boek) erbij te pakken, maar anderzijds kunt u ook de QR-code met uw mobiele telefoon inscannen en dan wordt u doorgelinkt naar de online versie.

Deel A - Linked Open Data – The Essentials

Dit deel beschrijft een algemene introductie in Linked Open Data.

A1 - Linked Open Data: The Essentials

Auteurs

Florian Bauer (REEEP)

Martin Kaltenböck

(Semantic Web Company)

The publication Linked Open Data: The Essentials gives decision makers a good overview of Open Government, Open Government Data, Open Data and Linked Open Data (LOD). It highlights the potentials and benefits of LOD, providing a quick guide with the most important steps for LOD publishing, a consumption strategy for your organisation and three best practice examples of LOD in use.



The original book has been written in 2011 to support the event 'Linking Open Data to Accelerate Low-Carbon Development,' a workshop for decision makers in clean energy organisations that was held at the Masdar Institute in Abu Dhabi in January 2012 in the course of the World Future Energy Summit, organised by the Renewable Energy and Energy Efficiency Partnership (REEEP).

But since then it has been widely used as resource on Linked Open Data, especially for decision makers to understand the potentials and benefits as well as the concepts of Linked Open Data, but also by project leaders, or technical experts, that need a starting point for Linked Open Data.

This reprint contains the first three chapters of the original book, and is the ideal starting point for the remainder of this publication.

Deel B - Linked Open Data – De Toepassingen

Dit deel beschrijft toepassingen en casus beschrijvingen op basis van Linked Open Data in verschillende toepassingsdomeinen.

B1 - 5 sterren toepassingen

Auteurs

Paul Geurts (Gemeente Nijmegen)

Thijs Brentjens (Brentjens Geo-ICT)

Chris van Aart (2CoolMonkeys)

Bij het maken van toepassingen zijn we er in de loop van de Pilot tegenaan gelopen dat er nog een brug te slaan



is tussen de aanbieders van slimme Linked Data en de app-bouwers. Waar app-bouwers graag tegen een JSON of REST service aanpraten, of de data zelf verzamelen en bewerken, heeft de wereld van Linked Data een eigen taal en logica. En die kosten wat moeite om eigen te maken. Je loopt er als app-bouwer tegenaan dat je een nieuwe semantische wereld instapt, waarin je moet investeren om hiervan kennis op te doen. Uiteindelijk krijgt het zijn meerwaarde, de link tussen data ligt in de DNA van de data in tegenstelling tot de huidige werkwijze waar de links door de app-bouwer worden gelegd door de data bij elkaar te sprokkelen, dan wel door webservices aan elkaar te knopen. In de app-wereld zou je dezelfde sterrenbenadering van Linked Data kunnen toepassen als die bij de data al geïntroduceerd is.

Harvesting van Data in open data formaten

App-bouwers verzamelen data en lezen dit zelf in (Harvesting). De data is in principe al verouderd na het inlezen. Data wordt meegeladen in de app. Links worden gelegd in de eigen database.



Koppelen van dataservices

App-bouwers koppelen dataservices binnen de app. Voor elke dataset maakt of configureert hij schermen. Logica en vertaling wordt dus binnen de app gelegd. Data is altijd actueel. Links worden binnen de app gelegd.



Linked Data

App-bouwers koppelen Linked Data Services, zijn altijd actueel, en maken slechts één scherm voor het presenteren van de data. Links liggen in het DNA van de data.



Dit artikel gaat over het toegankelijk maken van deze nieuwe semantische wereld van aanbod, via een Linked Data viewer en een Linked Data API, naar een toepassing.

B2 - De Huiskluis, Open Samenwerken met Linked Data

Auteurs

Paul Francissen (Envolve BV)

John van Echtelt (CGI)

De 'Huiskluis' is een praktijkcase die ontwikkeld is tijdens de landelijke Pilot Linked Open Data.



De Huiskluis 'weet alles over je huis'. Hij bevat links naar digitale informatie van overheden (basisadministraties), bedrijven en instellingen over woningen en bedrijven in Nederland. Daarnaast kunnen bewoners en eigenaren zelf informatie toevoegen aan hun Huiskluis. Daarmee creëren zij een online informatiedossier over hun eigen woning of bedrijfspand. De Huiskluis biedt bovendien de mogelijkheid om verzamelde informatie te delen met anderen; met burens, hulpverleners, winkels en anderen. Dit kan op een eenvoudige en veilige manier. Deze manier van koppelen en delen van informatie maakt vele nieuwe toepassingen mogelijk. Denk aan overheidsinformatie op maat voor elk huis of bedrijf, historisch materiaal dat gedeeld wordt met burens en archiefdiensten, of specifieke diensten rond energiegebruik, woninginrichting of hypotheekverstrekking.

Het initiatief laat zien dat informatie over elk huis of bedrijf in Nederland op maat kan worden geleverd. Dit sluit aan bij het doel van de Pilot Linked Open Data om te verkennen hoe informatie uit verschillende bronnen op het web bijeengebracht kan worden en ontsloten kan worden voor concrete toepassingen.

B3 - Branden blussen met Linked Open Data

Auteurs

Bart van Leeuwen (Netage)

Lian Pattje (Geonovum)

Erwin Folmer (Geonovum)

Informatie is van cruciaal belang voor de brandweer.

Beslissingen moeten in zeer korte tijd en



onder hoge druk gemaakt worden. Daarbij wordt gebruik gemaakt van alle informatie die voor handen is. In het huidige tijdperk, mede in het kader van Open Data, is er zeer veel informatie over onder meer gebouwen, verkeersinformatie, chemische stoffen, etc. beschikbaar, maar die wordt niet altijd gebruikt. De nachtmerrie van een brandweerman is dat hij verkeerde, noodlottige keuzes maakt, doordat beschikbare informatie niet is meegenomen in de besluitvorming. De brandweer (in samenwerking met Netage)

is daarom volop bezig om de mogelijkheden van Linked Open Data in de praktijk te testen. Dit om beschikbare en relevante open data aan elkaar te linken, gericht op het inzetten ervan bij een calamiteit (zoals een brand). Dit paper geeft inzicht in de problematiek en de uitdagingen van een brandweerkorps, en hoe Linked Data ingezet kan worden voor betere ontsluiting van relevante databronnen.

B4 - Linked Data & semantiek: we begrijpen elkaar steeds beter

Auteur

Marijke Salters
(Bureau Forum Standaardisatie)

De Linked Data methode biedt als open methode meerwaarde voor interoperabiliteit op de semantische laag. De e-overheid heeft steeds meer behoefte aan betekenis van begrippen. Steeds meer partijen wisselen data uit, ook om te komen tot beslissingen over uitkeringen, toeslagen etc.. Voor eenduidige uitwisseling van data tussen partijen is de juiste betekenis essentieel. De centrale vraag voor dit artikel is: Is Linked Data een oplossing voor het probleem van semantische interoperabiliteit? Om deze vraag te beantwoorden wordt in dit artikel de vergelijking gemaakt tussen een woordenboek in onze fysieke wereld en het gebruik van deze methodiek. Aan de hand van twee cases, Juriconnect en de Stelselcatalogus, wordt de meerwaarde van de Linked Data methode, aangetoond. Wel worden kanttekeningen geplaatst bij de haalbaarheid op korte termijn. Bovendien is een conclusie dat verkeerd gebruik van deze methode grote gevolgen heeft.



B5 - Casus Onderwijs- BegrippenKader: de basis onder Educational Linkedscope

Auteurs

Jeroen Hamers
Jos van der Arend
Leonie Verhoeff
Henk Nijstad
**(allen Kennisnet en Bureau
EduStandaard)**
Jeroen van Vuuren
(Verdonck, Klooster en Associates)

Het Onderwijs
Begrippenkader
(OBK) is dé
gemeenschappelijke
online database
met alle onderwijs-



begrippen en hun onderlinge
relaties – te beginnen met de
niveaus en het curriculum. Het OBK
is ontstaan in 2012 als oplossing
voor knelpunten in de Educatieve
Contentketen (ECK) bij het zoeken
en vinden van leermateriaal. Doel
van het OBK is om er voor te zorgen
dat alle informatiesystemen in het
Nederlandse onderwijs dezelfde
onderwijstaal spreken door middel
van het uniek identificeren en
definiëren van de begrippen met hun
onderlinge relaties. De inhoud van
het OBK is volgens de semantisch
web principes opgeslagen in een
RDF-store en wordt als open Linked
Data beschikbaar gesteld. De RDF-
store kan op meerdere manieren

worden uitgelezen: Direct uitvragen
door middel van een SPARQL end
point, via een specifieke API of via
een specifieke webbased 'browser'.
Voor de uitbreidingen van OBK is
een beheerorganisatie opgezet.
Het toepassingsgebied van de
uitbreidingen van de inhoud wordt in
toenemende mate breder dan alleen
de Educatieve Contentketen: het OBK
zal als generieke voorziening voor
eenduidigheid en duurzaamheid in
alle onderwijssemantiek een rol gaan
spelen. Het OBK wordt zodoende de
basis onder alle uitwisselingen binnen
het 'Educational Linkedscope'.

B6 - Casus De erfgoedthesaurus in de kennisketen van het cultureel erfgoed

Auteurs

Joop Vanderheiden (RCE)
Kees Hendriks (RCE)
**Nico Verbeij (Verdonck Klooster &
Associates, VKA)**

De Rijksdienst
voor het Cultureel
Erfgoed (RCE) wil
haar elektronische
dienstverlening
vernieuwen én de
eigen openbare gegevens zoveel
mogelijk ter beschikking stellen voor
verrijking en hergebruik. Daarom
koppelt de RCE zijn bronnen door
middel van een gemeenschappelijke
erfgoedthesaurus aan andere



bronnen in het erfgoedveld. Het geheel wordt voor verder gebruik als Linked Open Data beschikbaar gesteld aan private en publieke partijen. Deskundige partners in het erfgoedveld krijgen de gelegenheid mee te bouwen aan dit gemeenschappelijke erfgoed begrippenkader en kunnen met behulp hiervan hun eigen gedigitaliseerde collecties beter vindbaar maken. Leveranciers van informatieproducten (websites, apps, widgets, enz.) krijgen direct toegang tot de thesaurus en de daaraan gekoppelde informatie over allerlei erfgoedobjecten.

In deze casusbeschrijving gaan de schrijvers in op de visie van RCE met betrekking tot Linked Open Data in het erfgoedveld. Vervolgens beschrijven ze de activiteiten en bevindingen tot nu toe op het terrein van de ontwikkeling van de Erfgoedthesaurus, de gekozen informatiearchitectuur, de standaarden en de datamodellering die worden gehanteerd, de ontsloten gegevensbronnen en de eerste voorbeeldapplicaties.

Er is extra aandacht voor de Erfgoedsuite, een certificering van online tools die erfgoedpartijen helpen bij het bouwen en onderhouden van een eigen webpresentatie en het duurzaam opslaan van de eigen

collectiedata. Door de inzet van de erfgoedthesaurus worden deze lokale collecties verbonden en nationaal toegankelijk gemaakt. Een doorkijk naar de toekomst besluit deze bijdrage.

B7 - OpenInc, Ondernemende mensen die zich met Linked Open Data op de toekomst voorbereiden

Auteurs

John van Echtelt (CGI)

Pieter van Everdingen

Het potentieel van Linked Open Data is enorm groot. De komende jaren zal worden ontdekt welke toepassingen succesvol zijn, wat de winstgevende verdienmodellen blijken te zijn en welke maatschappelijke thema's het meest baat hebben met het slim koppelen van gegevens.



De Pilot Linked Open Data heeft ons geleerd dat het delen van kennis over organisatiegrenzen heen een hele goede manier is om maatschappelijke en economische waarde met open data te creëren. Mensen die hun kennis en talenten inzetten om te leren en om mooie nieuwe dingen (zoals de Huiskluis) te creëren.

Deze ervaring is een van de redenen waarom in Utrecht de open samenwerkingsomgeving OpenInc wordt opgezet. OpenInc is een coöperatieve netwerkorganisatie van en voor ondernemende mensen die nieuwe toepassingen willen ontwikkelen met (Linked) open data. OpenInc is vooral bedoeld voor studenten en werkzoekenden die hun kennis en ervaring op het gebied van (Linked) open data willen verbeteren en zich klaar willen stomen voor de beroepen van de toekomst.

B8 - Linked Open Data in electronics industry

Auteurs

John Walker (NXP)

Tim Nelissen (NXP)

In the electronics industry, the ability to get accurate, timely product data in front of the customer is a very important factor in the overall business process. Furthermore, enabling the customer to easily compare and select the right product for their application from the choice of literally hundreds, or even thousands, of candidates can reduce the overall time and costs involved in the purchasing process. Opening up access to the data is a key component, whether this is to free the data from existing silos for use within the organization, or making the data available to third parties. Also to facilitate the aggregation of data from multiple parties, it is very important to agree on a common schema that can be used to describe the products and enable easy mapping between schemata. In this paper we describe the approach we have used to manage and publish the data and discuss the questions that arise from a publishers perspective.



Deel C - Linked Open Data – How to

De bijdragen in dit deel zijn gericht om toepassers van Linked Open Data te ondersteunen bij het uitvoeren van bepaalde activiteiten.

CI - Walking the extra byte: A lifecycle model for Linked Open Data

Auteurs

Tijs van den Broek (TNO)

Anne Fleur van Veenstra (TNO)

Erwin Folmer (TNO)

More and more public organizations publish their data in an open format to increase transparency and foster economic activity. In this modern gold rush, organizations strive to open up as many datasets as possible, without considering the strategic importance of open data. Especially linking data to other datasets can lead to the creation of innovative services. One issue that is often not explicitly addressed before opening up data is the format of the dataset. Central to open data is that the format is machine readable. But to allow for effortless linking of datasets, data being merely machine readable is not sufficient. Lifecycle models can guide the process of



publishing Linked Open Data. Current Linked Open Data lifecycle models focus on the technical steps that need to be taken by the internal IT organization and often forget to include actions to be taken after publication. The effectiveness of Linked Open Data, however, depends on how much the data is used. Hence, this paper develops a Linked Open Data lifecycle model that takes the multiple disciplines and stakeholders within and outside the organization into account as well as the steps to be taken after publication of the datasets. First, using existing Linked Open Data lifecycle models, this paper identifies generic phases of opening up Linked Data: identification, preparation, publication, re-use and evaluation. Then, investigating the process of opening up Linked Data in a semi-public organization in the Netherlands, the lifecycle model is refined and detailed. This case study shows that the involvement of relevant stakeholders, both within and outside the organization and of various disciplines, is essential to realize the support for the process and stimulate re-use.

C2 - How to publish Open Data on the Web

Auteur

Christophe Guéret

(DANS, Free University Amsterdam)

There is an increasing interest towards publishing data open on the Web. Opening what use to be closed data brings a lots of opportunities in terms of social and economic development, participatory governance and improvement of research processes. Although the motivation is clear the publication is in practice open to a lot of different options. Even for one motivated in publishing his data online, it remains still unclear _how_ to do it. This chapter discusses the wide range of data publication approaches and pays a particular attention to the Linked Open Data publication principles advocated by the W3C.



C3 - Legacy dataformaten omzetten naar RDF met behulp van LODRefine

Auteur

Paul Hermans (ProXML)

De meeste data zijn nog niet beschikbaar als triples, wat wel gewenst is voor Linked Data. Zij zitten in tabellen, spreadsheets en zijn soms beschikbaar als XML of JSON. De vraag is hoe deze data efficiënt naar RDF-triples kunnen omgezet worden. Een mogelijke tool om hierbij te helpen is LODRefine, een gespecialiseerde versie van de initieel door Google ontwikkelde 'messy' data cleaning tool 'Open Refine'. Dit artikel schetst de mogelijkheden en manier van werken.



C4 - Semantiek en Linked Data

Auteur

Lieke Verhelst (Linked Data Factory)

De waarde van data, ook Linked Data, wordt in grote mate bepaald door de aanwezige beschrijvende informatie. Bij Linked Data is deze beschrijving vastgelegd in een semantisch model of vocabulaire. Het vastleggen van de betekenis van data door semantische modellen gaat beter dan bij tabulaire data. Dit komt doordat een semantisch model niet beperkt is, maar juist oneindig. Iedereen kan een semantisch model maken waarmee een onderwerp wordt beschreven. Soms worden daardoor onderwerpen door meerdere partijen vastgelegd. Dit is geen probleem omdat elke partij het onderwerp vanuit zijn eigen context beschrijft. In dit hoofdstuk wordt verder uitgelegd hoe je zelf een semantisch model maakt, en hoe je bestaande vocabulaires integreert in je eigen model.



C5 - How to: Linking resources from two datasets

Auteur

Christophe Guéret (DANS, Free University Amsterdam)

The Linked Data publication principles define a way to share data while using the Web as a publication platform. Just like the Web is used to publish documents it can also be used to publish factual data. URIs are used to design resources with are described with property/ values pairs and connected to other resources with typed links. These links between resources makes it possible to browse Linked Data and jump from resource to resource. They are also used to enrich datasets by connecting the resources with related datasets. But the best part is that `_anyone_` can create and publish these links!



In this article we will describe how two relate resources from two public datasets using a tool called 'SiLK'. The links will be published by a third party which is the linker tool provided by the semantic search engine 'Sindice'.

C6 - How to use AllegroGraph to work with Linked Open Data

Auteurs

Jans Aasman (Franz Inc.)

Roel Stap (Data for Use)

Most articles in this publication focus on interesting applications of Linked Open Data. However, for most people to get going with their first application with triples and Linked Open Data is not an easy task, especially when you are not a trained programmer or computer scientist. Therefore, in this article we'll describe some simple steps on how to use a triple store, how to load in some Linked Open Data and then how to write SPARQL queries with a graphical query builder.



C7 - BigData4Apps: Van Linked (Open Big) Data naar Contextualized Little Data

Auteurs

Chris van Aart (2CoolMonkeys)

Reind van Olst (2CoolMonkeys)

Michiel van Dijk (2CoolMonkeys)

Dit artikel beschrijft de ervaringen opgedaan tijdens de ontwikkeling van het BigData4Apps platform. Het BigData4Apps platform is in staat om van (semi)gestructureerde data, mobile data te maken. Aan het begin van 2013 zijn er honderden Nederlandse data bronnen publiekelijk beschikbaar en via het brede aanbod van App Wedstrijden zijn er diverse App concepten ontstaan. Veel van deze apps maken gebruik van een aantal geïsoleerde (open)data bronnen. Met de belofte: 'if you build it, they will come' (uit de film Field of Dream) zijn er diverse dataregisters opgericht, waarbij 'you' de data-aanbieder is en 'they' de derde partijen zijn die de daadwerkelijke apps bouwen en zo uit de data toegevoegde waarde kunnen creëren. Deze derde partijen worden op dit moment geconfronteerd met gefragmenteerde en niet gecontextualiseerde databronnen. Dit betekent dat er dus pas waarde kan worden toegevoegd als de



bronnen geharmonieerd, verrijkt en gecontextualiseerd zijn.

App-bouwers willen niet geconfronteerd worden met complexe datamodellen of vreemde opslagformaten, maar juist de beschikking hebben over app-componenten die de data ongefragmenteerd en gecontextualiseerd aanbieden. Uit diverse professionele app-projecten is dan ook het BigData4Apps-platform ontstaan: een combinatie tussen gedistribueerde data-, semantic web-, agent-, geo- en mobiele technologie. 'BigData4Apps' bestaat uit de volgende vier onderdelen: (1) De data-stekker, (2) De data-linker, (3) De query-agent; en de (4) De app-agent. De data-stekker koppelt met API's of leest data in een centrale data repository. De data-linker verrijkt deze data met semantic web URI's. De query-agent leeft boven op de verrijkte data en onderhandelt met de app-agent, die op het target device leeft. Deze app-agent kent de taak van een gebruiker. De app agent weet bijvoorbeeld dat de gebruiker een archeoloog is op zoek naar referentieopgravingen, een toerist op zoek naar een gerelateerd gebouw, een constructeur op zoek naar een leiding in de grond, een officier van dienst (brandweer) op zoek naar alle relevante gegevens van brandend pand, etc. De query-agent verstuurd in een compact

en gecontextualiseerd formaat de gegevens naar de app-agent. De app-agent onderhandelt vervolgens met de app-software over welke gegevens getoond moeten worden en wat er moet worden verwerkt. De Testaccio-app en de Spui-app laten de huidige (on)mogelijkheden zien van de Linked Data en BigData4Apps technologie.

C8 - Linked Open Data en 'BI': vervagende grenzen in de 'extended enterprise'

Auteurs

Bas Geerdink (CSC)

John Verhoeven (Inspiring Data)

Met het explosief groeien van de hoeveelheid Linked Open Data (LOD) worden bedrijven, overheden en

kennisinstellingen uitgedaagd tot het ontwikkelen van nieuwe en krachtige instrumenten om zicht te krijgen op deze reusachtige bron van nieuwe kennis. Daarbij gaat het om het vinden van toegevoegde waarde in Open Data, maar tevens om het realiseren van verrassende koppelingen van open en gesloten datasets. Een derde element dat kwaliteit toevoegt aan data, is het gebruik ervan zelf: het toepassen en toevoegen van LOD kan, mits goed gestructureerd en geborgd



in de organisatie, ook resulteren in 'verrijking' door de gebruikers.

Tegelijkertijd is een tweede grote trend zichtbaar: de opkomst van de 'extended enterprise'. In een 'extended enterprise' komt informatie niet meer alleen van interne, en gescreende bronnen, maar ook van 'buiten'. Deze externe informatiebronnen kunnen zowel professionele partners zijn als tijdelijke of incidentele allianties met partijen die over informatie beschikken (andere bedrijven, overheden, kennisinstellingen) of die kennis zelf genereren (de individuele burger). Deze externe informatiestroom kan zowel big data, small data, en gestructureerde of ongestructureerde data zijn. Deze bronnen van kennis bestaan nu nog goeddeels los van elkaar, waarbij van nature meer waarde wordt toegekend aan de 'eigen' enterprise data dan aan informatie van buiten. 'Eigen data eerst', zo zou je deze visie kunnen vertalen.

In een innovatieve, professionele omgeving is dit onderscheid achterhaald. Linked Open Data en 'enterprise data' komen samen in een vruchtbare mix waarbij de toegevoegde waarde meer is dan de som der delen. 'Joining and sharing' zijn de nieuwe sleutelbegrippen zodra het gaat om data. Een bedrijf dat op zoek is naar meer kennis en inzicht

over klanten, producten en nieuwe markten, zal hier een competitief voordeel kunnen behalen. Centraal hierbij staat het vermogen van de organisatie om data goed te beoordelen, te linken en te verrijken. Zowel inhoudelijk als technisch vergt dit een nieuwe aanpak. Om dit proces handen en voeten te geven is binnen de 'Open Data Value' Workshop een 'maturity scan' ontwikkeld die zichtbaar maakt in welke mate een instelling of bedrijf klaar is om open data en bedrijfsdata aan elkaar te linken en de toegevoegde waarde te vinden en vervolgens te optimaliseren. De scan maakt tevens inzichtelijk in welke mate een bedrijf op weg is een 'extended enterprise' te worden.

Dit proces van integratie van open en gesloten Linked Data bezien we vanuit de consequenties voor de organisatie, voor de (eisen aan) de data zelf, en vanuit de consequenties voor de geleverde diensten en producten.

Deel D - Linked Open Data – Technical Considerations

De bijdragen in deel D zijn een verdere verdieping in Linked Open Data met antwoorden op inhoudelijke uitdagingen.

DI - Designing A URI Strategy For The Dutch Public Sector

Auteurs

Linda van den Brink (Geonovum)

Hans Overbeek (KOOP)

Tijs Brentjens (Geonovum)

An important part of any Linked Data initiative is coming up with good identifiers for the objects that are subject of this Linked Data. But what is a 'good' URI identifier? Over the last year, several Dutch organizations have begun to feel the need for a national URI strategy to answer this question. One of the drivers for this is the European INSPIRE directive, which requires that each spatial object has a unique and persistent identifier. INSPIRE recommends using URIs for this and moreover recommends each European member state to establish a national URI strategy for the creation of these URIs.



Geonovum, responsible for the implementation of INSPIRE in the Netherlands, took the initiative to develop a URI strategy together with other Dutch stakeholders, finding each other through the Pilot Linked Open Data. In the context of this pilot a draft NL URI strategy was developed, which could be used to investigate the consequences of the choices we made. Involved parties were, among others: Geonovum, the Knowledge Center for Official Government Publications and the Tax Service. The result is a draft URI-strategy not only for spatial objects, but for all government information. A NL URI-strategy should lead to recognizable, intuitive clear URIs that identify concepts of standard models, and reference objects from registers. A framework which leads to trustworthy URIs will stimulate organizations to re-use URIs and refer to them instead of creating copies and/or creating their own.

This paper describes the background and goal of the NL URI strategy, focusing primarily on the technical aspects of a strategy; the organizational approach is to be settled in a next phase. Three categories of information resources are distinguished: standards, registers and applications. An

important insight gained during the pilot was that identifiers can only be properly managed by a register and that a lot of registers already exist, and can be re-used for the purpose of minting good URIs. The insight 'no register, no identifier' resulted in the choice to let an indication of the register be part of a URI, which results in the need of a 'register of registers'.

W3C recommends http-URIs as identifiers and we chose to follow this recommendation. The URI pattern suggested by the draft NL URI strategy is described in the paper. Four prerequisites for a 'good' URI identifier are given: it must be human readable and short, have a trustworthy appearance, be intuitive, and be persistent. The use of a generic upper-domain – we suggest data.gov.nl – should give the URIs trustworthy-ness and recognisability. The paper describes the four parts of the URI pattern: register, URI type, concept, and object reference. The URI pattern is applicable, in a slightly different way, for both reference objects and vocabulary terms. Rules, suggestions and naming conventions for each part of the URI pattern are given.

D2 - Geschiktheid van formaten voor Linked Open Data

Auteurs

Thijs Brentjens (Geonovum)

Marcel Reuvers (Geonovum)

Clemens Portele

(Interactive Instruments)

Bij het publiceren of gebruiken van Linked Open Data krijgt men met verscheidene technische



aspecten te maken. Een belangrijk aspect is het gegevensformaat. Het W3C gaat sterk uit van RDF-formaten voor Linked Open Data. Voor velen zijn andere formaten echter handiger voor publicatie en/of in gebruik. Dit paper behandelt daarom een scala aan veel gebruikte formaten en encodings (zoals PDF, JSON, RDF, CSV, GML en metadataformaten) en hun geschiktheid voor Linked Open Data. Geschiktheid uitgedrukt in termen van leesbaarheid voor machine/software en mens, mate van structurering en openheid van een formaat. Ook is het essentieel om te kunnen verwijzen met links naar andere gegevens en andersom: dat anderen kunnen linken naar de gegevens. Kijkend naar deze aspecten, geeft dit paper een overzicht van de geschiktheid van veel gebruikte formaten voor Linked Open Data.

D3 - Het ophalen van gegevens met HTTP, URIs en links

Auteurs

Thijs Brentjens (Geonovum)
Marcel Reuvers (Geonovum)
Clemens Portele (Interactive Instruments)

URI's zijn essentieel voor Linked Open Data. Niet alleen als identifiers, maar ook een mechanisme om via HTTP de gegevens op te vragen, al dan niet via web API's. Dit paper kijkt naar de technische werking van HTTP en URI's in de context van Linked Open Data. Daarbij komen de mogelijkheden van HTTP voor doorverwijzingen (HTTP 303) naar de informatie aan bod en het opvragen van verschillende formaten via Content negotiation. Webgebaseerde API's (Application Programming Interfaces) kunnen helpen bij het bevragen van data en opvragen van sets van gegeven. Het paper zet enkele API's naast elkaar. Links alleen zijn niet genoeg, voor mens en machine, om op voorhand te weten waar de link nu precies naar verwijst en zo te kunnen bepalen of het nuttig is de link te volgen. Semantiek over de link kan daarbij helpen. Deze onderwerpen passeren de revue, vanuit een technisch perspectief, inclusief kort aandacht voor (technische) testen hierop.

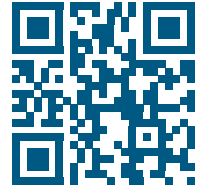


D4 - Een nieuwe wereld, een nieuwe informatiearchitectuur

Auteurs

Ria van Rijn (Atelier Helder Informatie Architecten)
Arjen Santema (Kadaster)

De techniek van Linked Open Data maakt het mogelijk, dat gegevens die door verschillende partijen zijn gepubliceerd op internet, worden geïnterpreteerd en gebruikt door derden. Dit leidt tot een interessante business case voor overheidsorganisaties. De eigenaar van gegevens kan met Linked Open Data voor een fractie van de kosten voldoen aan zijn plicht gegevens te publiceren. En de afnemers hebben niet alleen lagere kosten, maar ook veel meer vrijheid in de manier van afnemen. Bovendien kunnen de gegevens gecombineerd worden tot verrassende, ongedachte toepassingen. Informatiearchitectuur gaat over het aanbrengen van samenhang tussen ontwikkelingen in een organisatie rond de informatievoorziening. De techniek van Linked Open Data is daar een voorbeeld van. Inzetten van Linked Open Data beperkt zich namelijk beslist niet tot technische veranderingen.



Hoewel Linked Open Data nog in de kinderschoenen staat, gaan de ontwikkelingen momenteel ook erg snel. Daarom is het belangrijk om de consequenties ervan nu al in kaart te brengen. Optimaal gebruiken van deze nieuwe techniek vereist namelijk ook veranderingen in het omgaan met gegevens (door mensen maar ook door applicaties). Bovendien veranderen de verantwoordelijkheden van de organisatie bijvoorbeeld met betrekking tot veiligheid en privacy. Een organisatie moet hiermee bij voorbaat rekening houden, wil zij ten volle kunnen gaan profiteren van de voordelen van Linked Open Data.

D5 - Het contextdilemma hanteren met Linked Data

Auteurs

Marcel van Mackelenbergh
(Belastingdienst)

Rinke Hoekstra (Vrije Universiteit)

Hergebruik van gegevens kent een contextprobleem: een gegeven betekent iets anders in een



andere context. Als mensen gegevens creëren dan hebben ze daarmee een doel voor ogen. Ook hebben mensen een beeld van de wereld. Gegevens bevatten impliciet informatie die voortkomt uit dat doel en dat beeld. Geen rekening houden met deze impliciete informatie veroorzaakt belangrijke fouten in de uitvoering.

Linked Data erkent dat gegevens een context hebben en biedt een manier om hier mee om te gaan. Linked Data gaat uit van de 'dingen' die bestaan, zogenaamde resources. Iedere resource wordt aangeduid middels een identifier (URI). Doordat partijen dezelfde URI gebruiken voor het aanduiden van een resource, zijn ze in staat om gegevens over die resource met elkaar uit te wisselen. Een afnemer van gegevens ziet welk gegeven een leverancier van gegevens hanteert voor die resource.

De afnemer kan het gegeven van de leverancier overnemen maar hij kan ook altijd zijn eigen gegeven inzetten. Hierdoor ontstaat maximaal hergebruik met de flexibiliteit die de afnemer nodig heeft voor zijn eigen context.

Er is een nieuw soort gegevens nodig in dit web van gegevens. Er bestaat aandacht voor de zogenaamde provenance van gegevens. Provenance-gegevens beschrijven onder andere hoe, wanneer, door wie en waarom gegevens zijn gemaakt. Deze provenance is relevant omdat een afnemer van gegevens hiermee een inschatting kan maken of de gegevens (her)bruikbaar zijn voor zijn eigen toepassing. In dit artikel wordt een voorstel gedaan om naast de provenance ook de bruikbaarheid (usability) van gegevens te beheren. Bruikbaarheid van gegevens geeft aan in welke contexten een gegeven heeft geleid tot een succesvolle uitvoering.

D6 - Ontwerp van consistente domeinontologieën – hergebruik van bestaande sectorafspraken

Auteur

Roel Stap (Data for Use)

In veel bedrijfssectoren en toepassingsgebieden zijn, of worden, gemeenschappelijk informatieafspraken gemaakt en vastgelegd in zogenaamde informatiemodellen. Het doel van die afspraken is om te komen tot gemeenschappelijke betekenis van gebruikte terminologie, die binnen het betreffende domein gebruikt wordt om effectief tussen organisaties te kunnen communiceren. Door de toenemende samenwerking zowel binnen als buiten de eigen organisatie, is het belang van duidelijke informatie in de communicatie enorm toegenomen.



Het vastleggen van deze gemeenschappelijke afspraken gebeurt veelal op een formele maar technologie- en toepassings-onafhankelijke wijze. Dat wil zeggen

dat zij geen beperkingen kennen en toegepast kunnen worden in allerlei systemen en toepassingen waarin informatie wordt verwerkt.

Met de opkomst van het semantische web en de daarbij behorende middelen om informatie op een eenduidige manier in de vorm van ontologieën vast te leggen, ontstaat de vraag of en hoe we gebruik kunnen maken van deze bestaande informatieafspraken. Het hergebruik van deze afspraken zou de adoptie en toepassing van semantisch web applicaties vergemakkelijken en de aansluiting op bestaande informatiesystemen in de sector bevorderen.

Aan de hand van een IEC-informatiestandaard uit de energiesector zal in dit artikel de aanpak besproken worden hoe dit informatiemodel gebruikt kan worden om tot domeinspecifieke ontologieën te komen. Als voorbeeld wordt een energieverbruikontologie ontwikkeld die de basis is om energieverbruikgegevens eenduidig te publiceren, onafhankelijk welke organisatie de bron van de data is en hoe die is vastgelegd.

D7 - INSPIRE and Linked Data

Auteur

Clemens Portele
(interactive instruments)

'Linked Data' as such did not exist when the INSPIRE Directive or its technical foundation in Implementing



Rules were designed. However, as the focus of Linked Data is on principles how data is to be published on the Web, there is overlap in the goals of INSPIRE and of Linked Data, but there are also – sometimes subtle – differences. For example:

- INSPIRE did not limit the 'networks' on which spatial data is to be made available while Linked Data is explicitly focused on the Web. However, so far INSPIRE has limited its technical guidance to the Web, so in reality this is not a significant difference, but the network-neutrality requirements in the Directive and Implementing Rules have impact, even if all INSPIRE data is published on the Web. This affects in particular identifiers.

- The scope of INSPIRE is about publishing existing data sets and excludes the collection of new data. Typically publishing data as Linked Data goes beyond this as it strongly encourages adding links to other, related data.

The compatibility of INSPIRE concepts and Linked Data were first analysed in 2009 and presented at the INSPIRE conference 2010. In that paper it was highlighted that Linked Data is not a technical subject per se. It uses the principles of the Web, and technologies of the Semantic Web, to develop the Web of data. Linked Data can be seen as a philosophy about using the Web to create a unified set of data. A key aspect is to interlink the data through Web mechanisms.

All this is consistent with what is done in SDIs in general and INSPIRE in particular. There are some differences in technologies, but these are minor. The biggest issues are non-technical.

D8 - From Geo-Data to Linked Data: Automated Transformation from GML to RDF

Auteurs

Linda van den Brink (Geonovum)

Paul Janssen (Geonovum)

Wilko Quak (TU-Delft)

Linked Data provide an alternative route for dissemination of spatial information as compared to the traditional



SOA-based SDI approach. Where the latter is built on predefined structuring of semantics within domains, Linked Data is open to linking information to any data over the Web. In this respect both are complementary. The traditional approach providing a mechanism for a basis of standardized and structured data within domains, and Linked Data providing an open mechanism for sharing and combining. GML as the ISO standard

for exchange of service based spatial data and RDF as the Linked Data format are therefore related. GML provides the format in which many spatial datasets are available and exchanged. This standardization process and effort has been realized on a large scale. Why not let the web of Linked Data take advantage of this effort? This article will focus on the use of GML structured data as a source for deriving RDF structured data.

The first part of the paper focusses on deriving Linked Data from GML data. The first version of GML, v1.0, was based on RDF. From version 2.0 onwards GML was based on XML and XML Schema, but the object-property structure was retained. We describe a transformation for translating any correctly structured GML to RDFS/OWL automatically, using XSLT. Because GML's object-property structure translates very well to triples, the transformation is straightforward. Well-known GML

content elements such as names and descriptions are mapped to their RDF equivalent. However, any semantics specific to the input GML data (a.k.a. the application schema) are ignored in this translation.

In the second part, we study how more meaningful RDF can be created from GML, given the underlying information model, by transforming it from UML to RDFS/OWL. There exists a straightforward mapping to convert a UML model into a RDFS/OWL vocabulary. However, the re-use of existing concepts in vocabularies takes a central role in RDFS/OWL while in UML the use of vocabularies is not supported. We describe how annotating the UML model could improve this translation. Furthermore we discuss how the difference in design philosophy between UML and RDFS/OWL (open world assumption vs. closed world assumption) results in a different approach of how the absence of a specific property should be handled.

D9 – Mapping Statistical Linked Data

Auteur

Willem Robert van Hage
(SynerScope B.V.)

In this paper we will demonstrate, step by step, how to analyze a combination of Linked Data and non-Linked Data with the R statistical programming language. We will show how to combine Linked Data from different sources, and how to join this data with tabular statistical data. Specifically, we will be working with Linked Data from the Land Registry (Kadaster), about addresses and buildings, the Basisregistratie voor Adressen en Gebouwen (BAG); Linked Data from DBpedia, and non-linked tabular statistical data from the Statistics Netherlands (CBS). We will investigate geographical correlations between building properties and the socio-economical status of their inhabitants. Topics covered in this tutorial are: simple SPARQL end-point interaction, joining of Linked Data, joining Linked Data with non-Linked Data by a shared key, basic statistics, and geographical visualization.



D10 – Pragmatism versus formalism: the relation between Linked Open Data, semantics and ontologies

Auteurs

Paul Brandt (TNO)
Laura Daniele (TNO)

This paper will present the relationship, or absence thereof, between Linked Open Data (LOD), Semantics and Ontologies. Clearly the central idea behind LOD as well as Ontologies is to facilitate data processing that is grounded in its semantics, e.g., establishing the meaning of data and act accordingly. Ontologies are engineering artifacts that represent the relevant state of affairs in the world, their interrelations and, of course, their significance to the application. Linked Open Data essentially represents a vision where applications can automatically consume and apply externally published data in a way similar to how humans consume information that has been published on the world wide web. Since both approaches can be considered a means to achieve a similar goal, and hence represent distinct tools,



it is necessary to become aware of their differences and similarities before one can decide when, and how, both tools can be applied best. To that end we will investigate the balance between the precision of formalism, coming from the ontological approach, and the speed and readiness of pragmatism, coming from the LOD approach.



Aan de Pilot Linked Open Data hebben bijgedragen:

Aad Kamsteeg - Aart van Sloten - Adriaan van Oosten - Alessandra Scotta -
Andreas Hoogeveen - Anne Fleur van Veenstra - Anneke Zuiderwijk -
Arian Nasiri - Arjan Borggreve - Arjan Loeffen - Arjan Stuurman -
Arjanna van der Plas - Arjen Santema - Arn van der Pluijm - Arnold Reinders -
Bart Broex - Bart Huijbers - Bart Knubben - Bart Kusse - Bart van Leeuwen -
Bas Geerdink - Bert Spaan - Bert ten Brinke - Bert van Rest - Bob Coret -
Bram van Beek - Camille van der Harten - Chris van Aart - Christophe Guéret -
Clemens Portele - Corstiaan Kanters - Dennis Krukkert - Dennis Ramondt -
Denny Boelens - Dimitri van Hees - Dirk Melenberg - Dirk van Barneveld -
Dorine Berkvens - Dylan de Jong - Eelke Jager - Elvin Martha - Emil Zegers -
Erik de Rooij - Erik Jonker - Erik Orbons - Erik van der Zee - Erik Vrind -
Erwin Folmer - Erwin Hoogman - Eugene Tjoa - Felix Faassen - Ferry de Groot -
Fiemke Griffioen - Florian Bauer - Floris de Bree - Frank van Harmelen -
Frans Knibbe - Frans van der Zande - Gabriel Hopmans - Gerard Houtman -
Gerard Kuys - Gerard Nienhuis - Gerard Wildenbeest - Gerrit Hendriksen -
Gijs van Duijn - Haico van der Vegt - Hans Overbeek - Hans van der Burg -
Harrie van der Werf - Hayo Schreijer - Heidy van Kaam - Hein Bisterbosch -
Henk Kuipers - Henk Nijstad - Henk van den Berk - Herko Coomans -
Herman Assink - Holger Peters - Huib Verweij - Huibert-Jan Lekkerkerk -
Hylke Muntinga - Imke Vrijling - Jaap van den Heuvel - Jack Ruibing -
Jan Boerstra - Jan Bruin - Jan Bruinenberg - Jan Hidders -
Jan Jelle Boomgaardt - Jan Kooijman - Jan Meijer - Jan Voskuil -
Jandirk Bulens - Jans Aasman - Jasper Roes - Jasper Soetendal -
Jean-Louis Roso - Jene van der Heide - Jeroen Baltussen - Jeroen Hamers -
Jeroen van Vuuren - John van Echtelt - John Verhoeven - John Walker -
Joop Kielema - Joop Rasser - Joop Vanderheiden - Joris Booman -

Jos van der Arend - Jurriaan van Diggelen - Just van den Broecke -
Karin Swets - Kees Hendriks - Kees van der Graaf - Kees Waterman -
Koen Rutten - Laura Daniele - Leen van Doorn - Leonie Verhoeff -
Lian Pattje - Lieke Verhelst - Linda van den Brink - Loek Pfundt -
Maarten Kroon - Makx Dekkers - Marc de Vries - Marcel Reuvers -
Marcel van Mackelenbergh - Marco Aarts - Marco Brattinga -
Marco Geuze - Marian de Vries - Marijke Salters - Marjan Vernooy-Gerritsen -
Mark Backer - Mark Peeters - Mark Wiseman - Marleen Peters -
Marten Terpstra - Martijn Snel - Martin Kaltenböck - Martin Kodde -
Martin Leenheer - Maurice Vanderfeesten - Michel de Ru - Michel Grothe -
Michiel Boelhouwer - Michiel van Dijk - Milco Wansleebeem - Nico Verbeij -
Noël Van Herreweghe - Oliver Bartlett - Pano Maria - Patrick Koetsier -
Paul Brandt - Paul Francissen - Paul Geurts - Paul Hermans - Paul Janssen -
Peter Keijzers - Peter Klaver - Pieter Bresters - Pieter van Everdingen -
Pim Wijna - Ramon de Louw - Raymond Sluiter - Reind van Olst -
Reinout van Schouwen - Rene Bakker - Reza Ahmadi - Ria van Rijn -
Richard Zijlstra - Rinke Hoekstra - Rob Schuurbijs - Rob van de Velde -
Rob van Dort - Rob van Winden - Roel Stap - Ron Wardenier -
Roos Groeneveld - Ruby Beltman - Ruud Steltenpool - Silja Eckartz -
Sjako ten Haken - Steven IJzer - Steven Mekking - Stijn Goedertier -
Theo Scholl - Theo Versteeg - Thies Mesdag - Thijs Brentjens -
Thomas van der Woude - Ties Blaauw - Tijs van den Broek - Tim Nelissen -
Tine van Nierop - Tom Vijlbrief - Victor Budke - Victor van Katwijk -
Wilko Quak - Wilko Stalknecht - Willem Koeman - Willem Melder -
Willem Robert van Hage - Wim van der Heiden - Yvette Ellenkamp -
Yvonne Verdonk



Linked Open Data