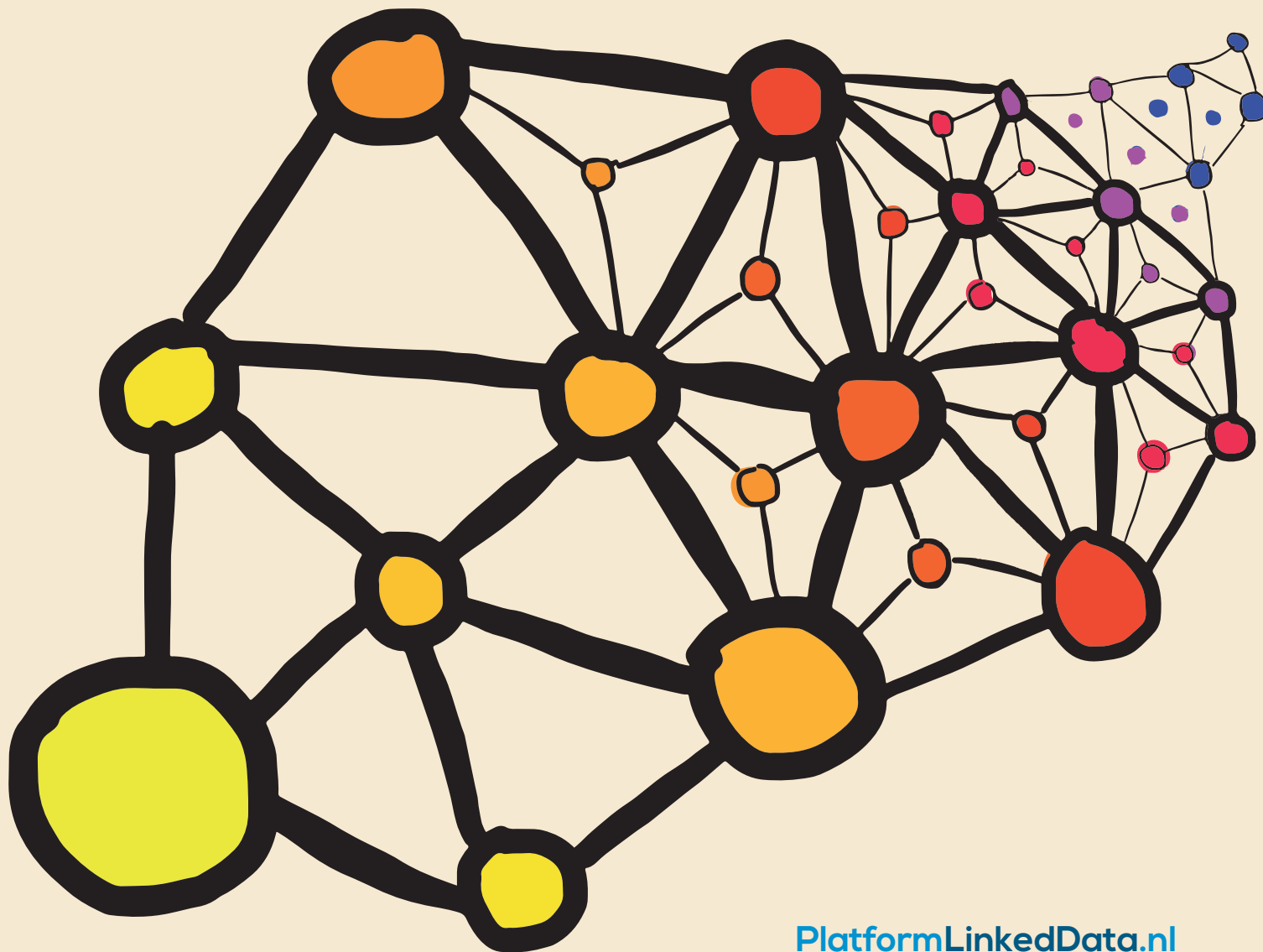
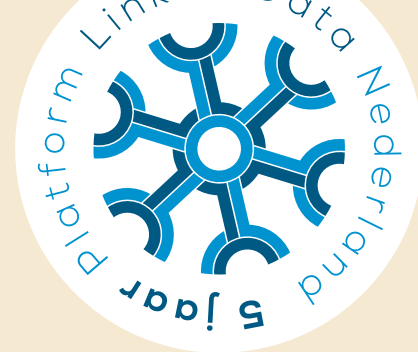


Linked Data in Nederland

Van pilot tot platform



PlatformLinkedData.nl

4	Voorwoord Alexander Rinnooy Kan	49	Blog: Rinke Hoekstra - linked data, zonder hoofdletters
5	Achtergrond: Linked Data: het hoe en waarom	50	Blog: Paul Hermans - Groenten uit Vlaanderen
11	Achtergrond: Van pilot naar platform	51	Ontwikkeling: #Geo4web testbed
12	Achtergrond: Linked Data in Nederland	52	Blog: Martin Kaltenböck - Experiences on the PLDN 2013 - 2017
16	Achtergrond: Linked Data in een dozijn	53	Case: Kadaster
18	Interview: Noud Hooyman (Ministerie van I&M)	54	Blog: Phil Archer - Standardisation is a community effort
22	Blog: Ed Parsons - Amplifying your spatial data on the web	55	Ontwikkeling: LOD Laundromat
23	Achtergrond: Open data vraagt om open standaarden	56	Case: NXP & SEMAKU
24	Interview: Frank Tierolff (Kadaster)	57	Blog: Marc de Vries - Privacy en Linked Open Data: aan tafel!
29	Case: Belastingdienst	58	Ontwikkeling: Linked Open Erfgoed Data
30	Interview: Frank van Harmelen (Vrije Universiteit Amsterdam)	59	Case: Open PHACTS
36	Thema: Semantic Data in de agrisector	60	Blog: Paul Hermans - Doen we het goed of doen we het goed?
38	Thema: De rol van Linked Data en REST API's bij de totstandkoming van de Omgevingswet	61	Ontwikkeling: CLARIAH
40	Thema: Beter onderwijs op maat met de Proeftuinen Linked Data	62	Linked Partners
43	Thema: KOI zorgt voor sectoroverstijgende gegevensuitwisseling	63	Nawoord Erwin Folmer
46	Thema: Digitale geletterdheid in het onderwijs en de impact van Linked Data	64	Colofon

Voorwoord

Sinds 2011 kent de Nederlandse overheid een open-databeleid. Het moet ertoe leiden dat steeds meer overheidsgegevens vindbaar en uitwisselbaar zijn en bovendien (gratis of tegen marginale kosten) beschikbaar worden gesteld. Die stap van de overheid heeft inmiddels al veel opgeleverd. Met name als het gaat om (open) geodata zijn er de afgelopen jaren flinke stappen gezet. Maar we zijn er nog lang niet. Vooral wat betreft de vindbaarheid en de bruikbaarheid van gegevens, kan en moet er nog wel het nodige gebeuren. Linked Data kan daar een voorname rol in spelen.

De maatschappelijke en economische waarde van (open) data is groot en zal alleen maar toenemen. Daarover is iedereen het eens. Maar data heeft alleen écht waarde als het ook bruikbaar is. Dat is nog geen gemeengoed. Linked Data gaat over bruikbare (open) data, gepubliceerd conform open standaarden op basis van de filosofie van het internet. Linked Data zorgt ervoor dat gegevens bruikbaar zijn en voorziet die gegevens ook van context.

Met Linked Data kan niet alleen economische en maatschappelijke waarde worden gecreëerd, maar kan ook de dienstverlening van de overheid wezenlijk worden verbeterd. Te denken valt aan het toekennen van huurtoeslagen, waarbij niet alleen de samenstelling van het gezin en inkomensgegevens worden meegenomen, maar ook aan juridische uitspraken (bijvoorbeeld over scheidingen) en energierekeningen. Hiermee kunnen onder meer onterechte afwijzingen en onjuiste berekeningen worden voorkomen.

Linked Data kan ook een grote rol spelen bij de kwaliteit en daarmee het gebruik van de basisregistraties. Van die basisregistraties bestaan momenteel honderden (lokale) kopieën en het zal enorm schelen als gebruikers, dankzij Linked Data, gegevens direct vanuit de bron kunnen gebruiken. Gelukkig worden óók op dat gebied flinke stappen gezet. Inmiddels zijn de eerste basisregistraties beschikbaar als Linked Data, staat het onderwerp op de agenda van de Digicommissaris en zijn veel organisaties (in de publieke en de private sector) druk doende met het publiceren van Linked Data.

De ontwikkeling van Linked Data kan niet zonder een goed georganiseerde en gemotiveerde achterban. We mogen ons gelukkig prijzen dat het Platform Linked Data Nederland daarin een centrale en verbindende rol speelt. In deze uitgave leest u meer over en van deze community. Ik beveel het u van harte aan, en hoop dat het zal bijdragen tot de bloei en groei van dit kansrijke werkterrein.

Alexander Rinnooy Kan
Hoogleraar aan de Universiteit
van Amsterdam en onder meer
voorzitter van de Big Data Alliance



Linked Data: het hoe en waarom

In 2012 is een pilot opgestart rondom Linked Data, een aanpak om informatie uit verschillende bronnen op een betekenisvolle manier met elkaar te verbinden en te gebruiken. Ten tijde van de start van de pilot was Linked Data voor veel organisaties een buzzwoord met interessante potentie. Maar wat is Linked Data eigenlijk, hoe is het ontstaan, wat is de kracht ervan en hoe interessant is het voor overheid, bedrijfsleven en wetenschap?

Het was Sir Tim Berners-Lee, bedenker en grondlegger van het wereldwijde web (www), die eind vorige eeuw als eerste over de term 'het semantisch web' sprak. Volgens hem is het semantisch web de visie dat informatie verbonden moet zijn en daar heb je technieken voor nodig. Een complicerende factor daarbij is dat de betekenis van mensen, dingen, gebeurtenissen, et cetera, niet constant is. Die betekenis kan variëren. Zo kan iets meerdere benamingen hebben. Zoals bij 's-Hertogenbosch, Den Bosch en Oeteldonk, waarbij het om dezelfde stad gaat. Of dezelfde benaming kan gebruikt worden binnen verschillende contexten met een verschillende betekenis. Zoals bij Bastille, wat een monument, een fort of een gevangenis kan zijn. Dit kan bij onjuist of slordig gebruik tot verwarring leiden. Mensen zijn gewend om contextuele factoren mee te nemen bij het toekennen van betekenis aan informatie. 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. Daar komt Linked Data om de hoek kijken. Linked Data is een techniek om met machine-leesbare context om te gaan, deze te genereren en te interpreteren.

De basis voor het semantisch web is RDF (Resource Description Framework), een data modellering standaard van het World Wide Web Consortium (W3C). In aanvulling op RDF kan RDFS (RDF Schema) gebruikt worden. Met behulp van RDFS kunnen 'klassen' van resources aangemaakt worden en ook beperkingen worden gelegd op de verschillende relaties die mogelijk zijn tussen instanties van deze klassen. Inmiddels is er een flink aantal RDF-standaarddefinities beschikbaar van resource- en relatietypes. Deze zijn vastgelegd in zogenoemde vocabulaires. Voorbeelden van veelgebruikte vocabulaires zijn SKOS (Simple Knowledge Organization System), voor het opstellen van begrippenkaders, gegevenswoordenboeken, taxonomieën en thesauri, en FOAF (Friend of a Friend), een vocabulaire dat gebruikmaakt van RDF om personen te beschrijven, hun relaties met andere personen en voorwerpen, en hun interacties.

In de visie van Berners-Lee past het semantisch web in de ontwikkeling van het internet: van het web van documenten (Web 1.0) via Web 2.0, waar het internet als interactief en sociaal communicatiemedium beschouwd wordt en waarbij gebruikers informatie kunnen uploaden naar Web 3.0: het web van Linked Data, waarbij links kunnen worden gebruikt en gevolgd en waarbij aanvullende informatie kan worden gepresenteerd.

De kracht van Linked Data

Binnen de (Nederlandse) overheid wordt niet alleen veel en steeds vaker data uitgewisseld, ook wordt veel data beschikbaar gesteld. Bijvoorbeeld via de verschillende basisregistraties, zoals de Basisregistraties Adressen en Gebouwen (BAG). Om die data goed te kunnen gebruiken, is een 'bijsluiter' wenselijk. Vergelijkbaar met een bijsluiter die je aantreft bij het verkrijgen van medicijnen. De bijsluiter als het gaat om data, bevat twee onderdelen. Het eerste onderdeel is semantiek. Met semantiek leg je de betekenis van gegevens vast. Het tweede onderdeel is provenance. Provenance gaat over hoe de data is ingewonnen, wordt aangegeven wat er met de data is gebeurd voordat het gepubliceerd is, wat de kwaliteit is van de data, waarvoor het wel of niet gebruikt kan worden, et cetera. De bijsluiter, met semantiek en provenance, helpt voorkomen dat er beslissingen worden genomen op basis van een verkeerde interpretatie van data.

Kenmerkende principes van Linked Data zijn:

- Alle data wordt vastgelegd in triples. Deze triples vormen graafs die er uitzien als ketens en netwerken van datawolken.
- Alle informatie wordt belicht vanuit een bepaalde invalshoek (viewpoints). Voor een volledig beeld van de situatie kunnen viewpoints gecombineerd worden tot een overzichtelijk en samenhangend verhaal.
- Linked Data werkt met een 'open world assumption'. Er kan altijd meer data beschikbaar komen en gebrek aan data betekent niet dat iets niet waar kan zijn. Een antwoord op een vraag kan dan 'misschien' of 'onbekend' zijn.
- In de wereld van Linked Data kan iedereen een gegeven verrijken met eigen informatie. Zo komen verschillende perspectieven bij elkaar. Deze eigenschap wordt ook wel de AAA-slogan genoemd, Anybody can say Anything about Any topic.
- Door gebruik te maken van bestaande vocabulaires wordt de interoperabiliteit tussen gegevens vergroot, waardoor data uit verschillende bronnen makkelijker met elkaar gecombineerd kunnen worden. De verschillen tussen data uit verschillende bronnen zijn makkelijker te overbruggen met Linked Data, omdat meer dezelfde 'taal' gesproken wordt. Ook kunnen datasilo's uit hun isolement worden gehaald en data beter hergebruikt worden zonder dat data onnodig gekopieerd wordt. Er kan gelinkt worden naar één leidende bron.

- Met Linked Data is het mogelijk om een grote hoeveelheid en verscheidenheid aan data met elkaar in verband te brengen. Daarbij maakt het niet uit of het nu kaartmateriaal is, illustraties, informatie op een webpagina, of gegevens uit een database.
- Met Linked Data krijgen gegevens context. Een context die digitaal te verwerken is: de gegevens vertellen een verhaal. Hierdoor kunnen enorme hoeveelheden gegevens met elkaar worden verbonden en verwerkt.
- Met Linked Data is verregaande data discovery mogelijk. Je kunt nieuwe data vinden, waarvan je het bestaan vooraf niet wist, maar die gelinkt is aan de databron waarmee je je zoekvraag begon. Zoekacties worden daarmee gericht.
- Databronnen kunnen op verschillende locaties staan en toch met één SPARQL query bevroegd worden, door gebruik te maken van federated queries.
- Met Linked Data zijn oplossingen mogelijk die voorheen niet of niet makkelijk te realiseren waren. Voorbeelden zijn oplossingen waarbij data uit vele bronnen met elkaar gecombineerd worden. Zoals bij de fraudedetectieoplossing bij de Belastingdienst, waarbij zowel het model van de data als de data zelf op dagelijkse basis vernieuwd en gevisualiseerd kan worden om zo mogelijke fraudegevallen beter en sneller te kunnen detecteren.

Bij een dialogtafel met stakeholders rond risico's van aardbevingen door gasboringen in de provincie Groningen zitten mensen met verschillende achtergronden, kennis en informatie aan tafel. Met behulp van simulaties is het mogelijk inzicht te krijgen in de waarschijnlijke gevolgen voor huizen en gebouwen, de schade aan dijken en de bewoonbaarheid van het gebied na een beving. Voor het maken van zo'n simulatie is het nodig om informatie uit diverse bronnen bijeen te brengen. Denk aan data over grondsamenstelling, breuklijnen, meetgegevens van sensoren of kwaliteit van dijken. Wanneer je een dergelijke dialogtafel ondersteunt met Linked Data, wordt het mogelijk om relatief snel een grote diversiteit aan relevante gegevens te verwerken. Hierdoor kunnen de deelnemers de reële risico's van aardbevingen in beeld brengen op basis van een grote diversiteit aan kennisbronnen en invalshoeken. Zo ontstaat een goed beeld van de waarschijnlijke gevolgen voor huizen en gebouwen, de schade aan dijken en de bewoonbaarheid van het gebied na een beving.

Big data vs. open data vs. Linked Data

Er zijn vele benaderingen mogelijk op 'data', maar de drie belangrijkste van de laatste jaren zijn Big, Open en Linked. De drie benaderingen hebben overlap. Vaak gaan initiatieven van Linked Data over vrij beschikbare (open) data, en spelen issues gerelateerd aan het Bigviewpoint ook een rol. Als we hebben over de BAG als Linked Data, dan hebben we het automatisch ook over open data (immers, de BAG is open data).

Big data gaat vooral over het bij elkaar brengen van data in alle soorten en maten om daar analyses op te doen. Niet voor niets wordt big data soms ook wel smart data genoemd. Open data zijn datasets die met een open licentie beschikbaar worden gesteld zodat toegang en hergebruik zonder beperkingen mogelijk is. Het idee van open data is om de beperkingen in hergebruik tot een minimum te limiteren. Hierdoor wordt het delen en hergebruik van data bevorderd. Veel data van de overheid is open om transparantie te vergroten en economische activiteit te bevorderen. Onder open data wordt inmiddels ook verstaan dat data (indien relevant) 'machine readable' moet zijn en in een open formaat beschikbaar moet worden gesteld (Zie ook: het vijfsterrenmodel).

Linked Data is gebaseerd op de concepten van het wereldwijde web: de data is daar te vinden, in een gestructureerd open formaat, er is naar te linken dankzij de toepassing van URI's en de data biedt zelf ook verdere relevante verwijzingen.

Vier principes

Voor Linked Data gelden er vier principes:

- Geef alle dingen waaraan je wilt kunnen linken, een uniek adres op internet (Uniform Resource Identifiers; URIs), conform de Nederlandse URLstrategie.
- Gebruik HTTP-URI's zodat er naar deze dingen kan worden verwezen en ze kunnen worden opgezocht door mensen en machines.
- 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, SPARQL, OWL, JSON-LD of SKOS.
- Neem links naar andere gerelateerde, open data-concepten op in de beschrijving om het ontdekken van gerelateerde informatie op het web te verbeteren.

Vijfsterrenmodel

Als het gaat over Linked Data, dan gaat het onherroepelijk ook over het vijfsterrenmodel, zoals dat is uitgewerkt door Tim Berners-Lee. Het model is een hulpmiddel aan de hand waarvan organisaties toe kunnen werken naar machine-readable herbruikbare Linked Data. In het model is informatie zo ingedeeld dat computers deze automatisch kunnen verwerken. Hoe meer sterren data op het web hebben, hoe meer zij geschikt zijn voor automatische verwerking. En hoe hoger het aantal sterren, hoe beter de bruikbaarheid en de openheid van de dataset. Datasets van de overheid, zoals basisregistraties, hebben bij voorkeur vijf sterren.

De eerste drie sterren betreffen open data en vanaf de vierde en vijfde ster wordt gesproken over Linked (Open) Data.

In het Platform Linked Data Nederland is, met een knipoog naar het vijfsterrenmodel van Tim Berners-Lee, een variant verschenen met zes sterren. Deze variant heeft als extra dat data in de vierde stap wordt aangeboden als een service via een JSON API. Het beschikbaar stellen van data via een JSON API vraagt van de aanbieder van data minder inspanning dan het beschikbaar stellen van data in RDF, zoals Berners-Lee dat omschrijft. Daarmee wordt een drempel verlaagd en wordt tegemoet gekomen aan de wensen van ontwikkelaars.



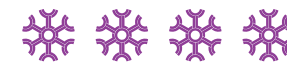
De informatie is beschikbaar op het internet, in welk formaat dan ook.



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, zodat anderen makkelijk naar de dataobjecten kunnen verwijzen. Je zou kunnen zeggen: bij vier sterren is de data 'linkable' (URIs, RDF, etc.) en bij vijf sterren ook daadwerkelijk gelinkt.



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

Overheden en Linked Data

De roep om digitalisering van diensten en services heeft ervoor gezorgd dat (open) data, en daarmee Linked Data, ook voor de Nederlandse overheid steeds belangrijker zijn geworden. Sinds een aantal jaar volgt de overheid daarbij een internationale trend (met name zichtbaar in de Verenigde Staten, Australië en het Verenigd Koninkrijk), dat informatie van overheden actief openbaar beschikbaar, toegankelijk én herbruikbaar moet worden gesteld.

De beschikbaarheid en openheid van data en de mogelijkheid om verbanden te leggen tussen informatie op internet, speelde ook een voorname rol in het rapport 'Open Overheid – Internationale beleidsanalyse en aanbevelingen voor Nederlands beleid' (2011), dat het onderzoeksinstituut TNO in opdracht van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) heeft opgesteld. In dat onderzoek is niet alleen de nadruk gelegd op informatie die 'open' is, maar is ook duidelijk gemaakt dat de samenhang in de informatie voor publiek duidelijk zichtbaar zou moeten zijn. 'Omdat veel informatie tot stand komt binnen processen krijgt zij aanzienlijk meer betekenis door de samenhang van de informatie te laten zien. Daarmee wordt de informatie pas echt toegankelijk en herbruikbaar. Linked Data kan hierbij een rol spelen' (pagina 17 – rapport Open Overheid).

Meerwaarde

In de afgelopen jaren hebben we gezien dat digitale dataverwerking bij overheid, bedrijfsleven en wetenschap een steeds belangrijker rol is gaan spelen. Maar er is nog veel meer mogelijk. Daarvoor is het wel nodig dat de basis, de onderlaag als het ware, in orde is. In plaats van dat ieder zijn digitale 'taal' gebruikt, zou je meer crosssectoraal (geowereld, zorg, cultuur, onderwijs, et cetera) willen werken. Door standaarden te gebruiken, bij voorkeur (open) standaarden die toch al gebruikt worden op het web, kan je tot meer interoperabiliteit komen. Dat kan al leiden tot een grote vermindering van ergernis en tot een forse toename van de maatschappelijke en economische waarde van data. Maar échte meerwaarde bereik je als niet alleen dezelfde (open) standaarden worden gebruikt, maar ook URIs worden toegepast, vocabulaires worden gemaakt en gebruikt en er wordt gelinkt. Dan kunnen we de volledige kracht van Linked Data benutten. Alleen daarom al is het bestaan van een brede en sterke community in Nederland rondom Linked Data, geen overbodige luxe.

Van pilot naar platform

Om de vele mogelijkheden van Linked Data voor de overheid te onderzoeken, is in 2012 de Pilot Linked Open Data (PiLOD) gestart. Het doel van de pilot was vooral om te onderzoeken wat Linked Data is, wat het voor overheden zou kunnen betekenen en welke stappen er in dat kader gezet zouden moeten worden.

In de pilot werd al snel duidelijk dat er grote vraag was naar een brede, uniforme strategie en aanpak met betrekking tot Linked Data. Ook werd een sterke behoefte geconstateerd naar een community, waarin het verzamelen en uitwisselen van ervaringen en kennis met betrekking tot Linked Data centraal zou moeten komen te staan. De gewenste aanpak en de community kregen gestalte in het Platform Linked Data Nederland (PLDN), dat na twee pilotfasen in 2014 het licht zag.

Met de komst van het Platform Linked Data Nederland werd ook de scope verbreed. Zo werd de focus vooral gelegd op de kenniscommunity, gericht op het toepassen van Linked Data, en lag de aandacht voortaan op alle data in plaats van alleen op open data. Verder werd breder gekeken dan alleen de overheid en werd in het vervolg ook ingezoomd op het bedrijfsleven en de wetenschap, waarmee nieuwe kennis en ervaring werd ingebracht.



Platform Linked Data Nederland

In de afgelopen jaren heeft de PLDN-community het aantal specialisten in Nederland op het gebied van Linked Data enorm vergroot én ook geleid tot een prettige 'bijvangst'. Tal van organisaties, zoals het Kadaster, Kennisnet, vooraanstaande banken, de Belastingdienst en de Nederlandse Politie, hebben in de afgelopen jaren via de PLDN-community specialisten ingehuurd om hun beleid en innovaties met betrekking tot Linked Data vorm te geven.

Wat vijf jaar geleden begon met een pilot over het hoe en waarom van Linked Data, is uitgegroeid tot een succesvolle en gevarieerde community. Die community heeft ervoor gezorgd dat Linked Data in Nederland niet alleen maar een buzzwoord is gebleken, maar onderdeel is geworden van operationele kennis en een uitgebreid internationaal kennisnetwerk.

Linked Data in Nederland

Als het gaat om de ontwikkeling van Linked Data in Nederland, is er vanaf 2012 veel gebeurd. Presentaties, het ontwikkelen van ‘producten’, evenementen, het (digitaal) samenkomen van experts en specialisten, praktijktoepassingen, et cetera. Te veel mijlpalen om allemaal te benoemen. Daarom een bloemlezing met highlights uit vijf jaar Platform Linked Data Nederland.

LOD Laudromat, NXP en Open PHACTS

Op 29 september 2015 werden de winnaars bekendgemaakt van de beste Nederlandse Linked Open en Linked Enterprise Data toepassing van 2015. Voor de prijs van beste toepassing in Nederland, waren zestien toepassingen ingezonden. In de categorie Linked Open Data ging de eerste prijs naar de LOD Laudromat van de VU Amsterdam, een ‘wasmachine’ die ervoor zorgt dat alle ongewilde ‘vlekjes’ worden verwijderd en er ‘schone’ Linked Data overblijft. De beste Linked Enterprise toepassing van 2015 werd de NXP Enterprise Data Hub. NXP sleepte ook in de Europese Linked Data Contest de eerste prijs binnen in de categorie ‘Linked Enterprise Data’. In de categorie Open Data was de ELDC-award voor Open PHACTS, een Europees project in de farmaceutische industrie met een stevige Nederlandse inbreng.
www.geonovum.nl/onderwerpen/linked-data/nieuws/nxp-en-lod-laudromat-beste-linked-datatoepassingen-2015

BGT als Linked Data: een proof of concept

In opdracht van het ministerie van Infrastructuur en Milieu heeft Geonovum begin 2015 onderzocht wat Linked Data kan betekenen voor de Basisregistratie Grootschalige Topografie (BGT). In het onderzoek is gezocht naar antwoorden op vragen als ‘Is Linked Data een goede technologie om data in het Stelsel van Basisregistraties over de grenzen van registraties heen te gaan verbinden, op zo’n manier dat deze data kan worden gebruikt om allerlei vragen te beantwoorden?’ en ‘Waar liggen de grenzen op het gebied van zaken als performance en opslag waar het grote hoeveelheden data betreft?’.
www.pilod.nl/wiki/BGT_Linked_Data

Publicaties

In de afgelopen jaren heeft het Platform verschillende publicaties opgeleverd:

- Parels van Linked Data Toepassingen www.pilod.nl/w/images/e/ee/LinkedDataParels.pdf
- In 9 stappen naar Linked Data www.pilod.nl/wiki/BoekTNO/stappenplan
- Linked Data in Beeld www.pilod.nl/w/images/3/32/Linked_Data_in_beeld_2014.pdf
- Pilot Linked Open Data Nederland (Management-overzicht & Verdieping) www.pilod.nl/wiki/Boek
- Betekenisvol Verbinden met BP4mc2 www.bp4mc2.org
- Concept URI-strategie www.pilod.nl/wiki/Concept_URI-strategie

Big Data Value Center

In april 2014 heeft de Nederlandse Linked Datacommunity de beschikking gekregen over een eigen server, die gehost wordt bij het Big Data Value Center in Almere. De server is de primaire leer- en experimenteeromgeving en is ook een omgeving waar organisaties Linked Datasets kunnen publiceren en onderhouden. Ook leveranciers kunnen op de server hun tooling beschikbaar stellen voor experimenten, waardoor er een reikwijdte aan tools te gebruiken is. De omgeving is verder te gebruiken in het onderwijs om studenten (bijvoorbeeld via OpenRefine) zelf Linked Data te laten maken. De server wordt beheerd door vrijwilligers van het Platform Linked Data Nederland.
almere.pilod.nl

SKOS op ‘Pas toe of leg uit’-lijst

In juni 2015 is SKOS (Simple Knowledge Organization System), namens PLDN, toegevoegd aan de ‘Pas toe of leg uit’-lijst van Forum Standaardisatie. Voorstandaarden op deze lijst geldt dat organisaties binnen de Nederlandse overheid deze verplicht moeten toepassen. Het basisidee van SKOS is dat elk begrip een eigen webadres krijgt. Voorbeeld: het begrip inkomsten wordt gebruikt door de Belastingdienst, maar ook door het gevangeniswezen (nieuwe bajesklanten). Beide begripshanteringen krijgen een eigen webadres. Aan het webadres is te zien welke organisatie verantwoordelijk is voor de definitie van het begrip.
www.w3.org/2004/02/skos
www.forumstandaardisatie.nl/standaard/skos

Van 0 naar 1160

abonnees op de
nieuwsbrief van PLDN

Geo-datasets als Linked Data

Op 27 juni 2016 heeft het Kadaster twee basisregistraties als Linked Open Datasets gelanceerd. Het gaat om ‘de Digitale kadastrale kaart’ (Basisregistratie Kadaster) en ‘TOPIONL’ (Basisregistratie Topografie). Met de lancering van de twee datasets is het Kadaster de eerste (overheids)organisatie die geo-datasets als vijfsterren Linked Data aanbiedt. Het Linked Datateam bij het Kadaster bestaat overigens voor een groot deel uit experts uit PLDN-community.
www.kadaster.nl/Dataplatform-voor-de-toekomst

UX Linked Data Challenge

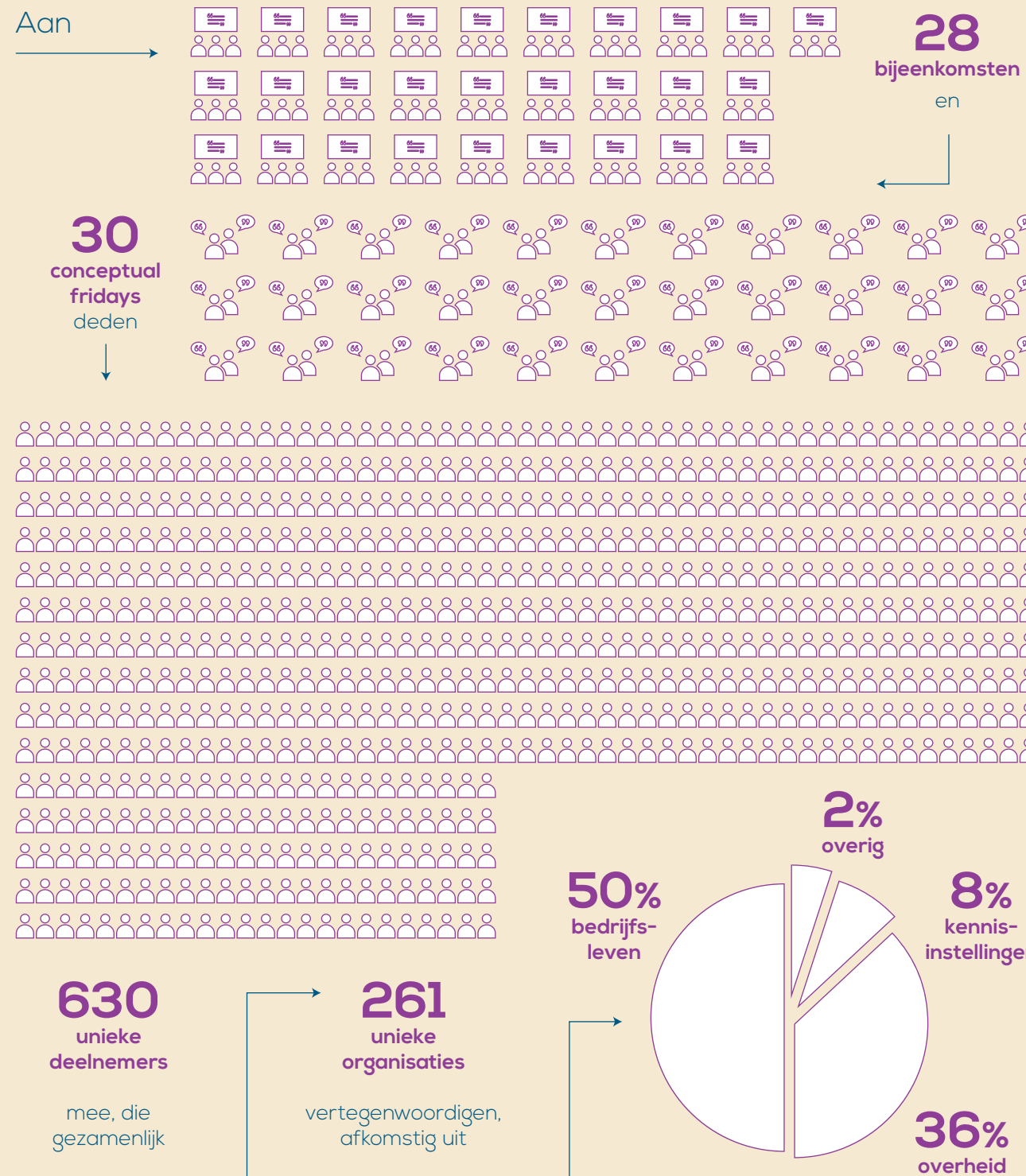
Op 3 november 2016 kwamen in CGI's Spark Innovation Center in Rotterdam experts bijeen op het gebied van visualisatie en Linked Data. Hun uitdaging was: maak een toepassing die de kracht van Linked Data goed benut, die eenvoudig te bedienen is en die uitnodigt tot meer en beter gebruik van Linked Databronnen. Winnaar werd de toepassing waarbij Linked Data van toegevoegde waarde is voor datajournalistiek. De gerealiseerde oplossing ondersteunt datajournalisten bij het doen van onderzoek op basis van data, afkomstig uit verschillende bronnen in verschillende formaten. Vertrekpunt waren twee datasets over drone-aanval- len in Pakistan en Jemen. Deze datasets zijn omgezet naar Linked Data en verrijkt met gegevens over coördinaten en links naar andere databronnen. www.geonovum.nl/nieuws/ux-data-challenge

URI-strategie

De Werkgroep URI-strategie is vanuit het Platform Linked Data Nederland ingesteld om te onderzoeken of een nationale URI-strategie voor Linked Data van de overheid zinvol zou zijn en zo ja, wat de ingrediënten voor zo'n strategie zouden moeten zijn. De leden van de werkgroep concludeerden dat een URI-strategie, of beter een Linked Data-strategie, zinvol zou kunnen zijn omdat Linked Data een nieuw vakgebied is, de nodige complexiteit kent en domein overstijgende vraagstukken behelst. Technische, maar ook organisatorische. www.pilod.nl/wiki/Boek/URI-strategie

Community

Een van de succesfactoren van Linked Data in Nederland, is de community. Uit een in het najaar van 2016 gehouden enquête blijkt dat meer dan de helft van de communityleden bovenmatig ervaren is als het gaat om Linked Data. Opvallend is verder dat de community een sterke vertegenwoordiging heeft in het geodomein. Acht op de tien leden vindt het een goede of uitstekende zaak dat de PLDN actief bezig is om Linked Data-standaarden op te nemen op de 'Pas toe of leg uit'-lijst van de Nederlandse overheid. Datzelfde percentage komt terug op de vraag of beïnvloeding op het Nederlandse overheidsbeleid door de PLDN gewenst is. Ruim acht op de tien PLDN-leden juicht die beïnvloeding toe. Uit de enquête is ook gebleken dat er verbeterpunten zijn. Zo zou de PLDN zich meer moeten focussen op aanwezigheid (bijvoorbeeld op universiteiten), het uitdragen van goede voorbeelden buiten de community en communicatie voor leken op het gebied van Linked Data. In dat kader stelt een van de leden voor om met een item aanwezig te zijn in De Wereld Draait Door... ☺ www.platformlinkeddata.nl



Linked Data in een dozijn

Sinds de eerste ervaringen binnen proeftuin Omgevingen wordt Linked Data nu meer en meer in de dagelijkse praktijk gebruikt bij een groeiend aantal organisaties. Het is belangrijk om de mogelijkheden van Linked Data structureel in te passen in de visie en het beleid op het gebied van ketensamenwerking en het slim gebruik van data over grenzen heen. Wat Linked Data voor uw organisatie kan betekenen.

10 Linked Data biedt uitkomst bij het doorzien van complexe vraagstukken

Linked Data leent zich uitstekend voor het maken van visualisaties om zo verdachte patronen, afhankelijkheden en feiten 'on-the-fly' sneller en beter te kunnen doorzien. Iets dat met meer rigide traditionele oplossingen vaak veel moeilijker, tijdrovender en duurder is.

11 Linked Data zorgt voor kosten-efficiënte oplossingen

Linked Data is een kosten-efficiënte en meer flexibele data-integratie oplossing ten opzichte van meer traditionele data-integratie oplossingen. Het leent zich uitstekend voor een aanpak waarbij men kleinschalig kan beginnen en stap-voor-stap oplossingen kan uitbreiden, conform de best practices die er zijn (W3C/OGC best practices, ook in relatie tot geodata).

1 Linked Data richt zich op open standaarden, conform overheidsbeleid

Linked Data bouwt voort op breed geaccepteerde open internetstandaarden (W3C) en standaarden zoals die voorkomen op de 'Pas toe of leg uit'-lijst van de Nederlandse overheid.

8 Linked Data biedt flexibiliteit

Linked Data biedt de flexibiliteit om naast standaard vocabulaires ook zelf vocabulaires te definiëren als dat nodig is, conform hetzelfde uniforme modelleringsprincipe. Daardoor zijn knowledge graphs (kennismodellen) oneindig uitbreidbaar op een consistente en praktische manier en kan een eigen kennisdomein in detail gemodelleerd worden.

12 Linked Data kan ook aangeboden worden via API's

Inmiddels is relatief veel ervaring opgedaan met het beschikbaar maken van Linked Data via API's, waardoor het werken met Linked Data voor webontwikkelaars laagdrempeliger wordt.

2 Linked Data zorgt voor een web van data

Naast het web van documenten, waarbij hyperlinks geen betekenis hebben (de eerste versie van het internet), ontstaat met Linked Data een web van data, waarbij hyperlinks wel betekenis hebben (de verbeterde versie van het internet, waarbij data betekenisvol verbonden kan worden).

7 Linked Data interoperabiliteit door hergebruik vocabulaires

Er wordt gebruikgemaakt van standaard vocabulaires, waardoor:

- Het uitwisselen van gegevens vereenvoudigd wordt (interoperabiliteit);
- Hergebruik van gegevens bevordert wordt, omdat meer dezelfde taal wordt gesproken;
- Kloven tussen datasilo's makkelijker te overbruggen zijn, omdat er minder detail mappings nodig zijn en gemakkelijker de link gemaakt kan worden tussen dataobjecten uit verschillende bronnen en er meer sprake is van harmonisatie en orkestratie dan van integratie. Bovendien kan data uit verschillende bronnen met één query bevraagd worden door middel van een federated query;
- Zoekopdrachten tot betere resultaten leiden, omdat data uit verschillende bronnen makkelijker met elkaar gecombineerd kan worden, beter gefilterd kan worden en beter geaggregeerd kan worden in slimme datatoepassingen (smart data).

6 Linked Data zorgt voor één uniform formaat

Bij Linked Data wordt alles uitgedrukt met RDF triples (subject-predicte-object constructies) in knowledge graphs (kennismodellen), waardoor er één uniforme manier is om data te modelleren en te definiëren. Omdat er sprake is van één uniform formaat, behoren conversieproblemen tot het verleden.

3 Linked Data zorgt voor beter vindbare en linkbare data

Door het toevoegen van unieke identifiers (URI's) aan data, wordt data beter vindbaar. Niet voor eindgebruikers, maar voor machines (machine-leesbare data). Daardoor kunnen computers het zoekwerk doen en wordt data linkbaar over het internet, sector-, domein- en organisatie overstijgend (federatieve architectuur en connectivity).

4 Linked Data zorgt voor begrijpelijke data

Door het toevoegen van betekenis (semantiek) aan data, wordt data beter begrijpbaar voor eindgebruikers, zodat besluitvorming op basis van deze data beter kan zijn.

5 Linked Data bevordert de kwaliteit en hergebruik van data

Linked Data is zelf-beschrijvend, oftewel metadata en provenance data zijn een integraal onderdeel van Linked Data en worden op dezelfde uniforme manier vastgelegd. Daarmee wordt de kwaliteit van de data verder verhoogd.



Heeft Linked Data meerwaarde voor de overheid?

Als opdrachtgever van een aantal basisregistraties is Noud Hooyman van het ministerie van Infrastructuur en Milieu geen onbekende in de wereld van data. Hij ziet weliswaar de voordelen van Linked Data, bijvoorbeeld als het gaat om de BAG of de BGT, maar is nog een beetje zoekende hoe het ingezet zou moeten worden in het brede overheidslandschap.

In de afgelopen jaren is digitale dataverwerking bij overheden een steeds belangrijker rol gaan spelen. Een rol, waarbij Linked Data steeds vaker om de hoek komt kijken. Maar ondanks die toenemende rol, is het 'begrip' Linked Data lastig te vatten. Ook voor mensen die toch goed in de digitale datamaterie zitten. Zoals Noud Hooyman, afdelingshoofd beleid geo-informatie op het ministerie van Infrastructuur en Milieu (I&M). Als opdrachtgever voor onder meer de Basisregistraties Adressen en Gebouwen (BAG) en de Basisregistratie Grootchalige Topografie (BGT) vindt hij Linked Data "een moeilijk concept". Hij vindt met name moeilijk in te schatten wat het belang en de meerwaarde is van Linked Data.

"Ik word soms benaderd door bedrijven die vinden dat bepaalde data open moet zijn, maar er zijn eigenlijk nooit bedrijven die vinden dat de data ook 'linked' moet zijn. Soms vraag ik mij af of wij, als het gaat om Linked Data, niet te veel voor de troepen vooruitlopen? 'Ik vind het fantastisch dat het Kadaster (waarvan het ministerie van I&M opdrachtgever is) de BAG als Linked Data beschikbaar heeft gemaakt. Tegelijkertijd ben ik nog een beetje aan het zoeken in welk tempo de overheid als geheel iets zou moeten

met Linked Data. Als overheid willen wij het Stelsel van Basisregistraties veel meer een stelsel laten worden en het zou mooi zijn als Linked Data daar een rol in kan spelen. Als door middel van het gebruik van Linked Data er een basis wordt gecreëerd, waardoor de voordelen ook beter zichtbaar worden."

Vijfsterrenmodel

Wie Linked Data zegt, refereert aan het vijfsterrenmodel, een door Tim Berners-Lee ontwikkeld hulpmiddel aan de hand waarvan organisaties toe kunnen werken open naar machine-readable herbruikbare Linked Data. Hooyman denkt intuïtief

"Ik vind het fantastisch dat het Kadaster de BAG als Linked Data beschikbaar heeft gemaakt"

dat basisregistraties, zoals de BAG of BGT, moeten voldoen aan die vijf sterren, maar wat hem betreft zijn die vijf sterren niet alleszeggend. Hij is er zelf een voorstander van om met name te kijken naar de waarde van data. Die data moet een bepaald maatschappelijk of economisch probleem oplossen en



“Mogelijk is het een oplossing om het geen Linked Data te noemen, maar FAIR-data”

daarna begint, wat Hooyman betreft, pas het traject waarbij die data vijf sterren wordt gemaakt. “Het is volstrekt nutteloos om links tussen willekeurige zaken te maken, alleen omdat het kan.” Een punt waar Noud Hooyman nog niet uit is, of de vijf sterren voor alle overheidsdata moet gelden. “Gevoelsmatig kan ik mij prima voorstellen dat basisregistraties aan elkaar gelinkt moeten worden, maar het is de vraag of dat ook een verplichting moet worden voor alle andere overheidsdata.”

Open data vs. Linked Data

Als je het hebt over Linked Data, dan is de stap naar open data dichtbij. Voor wat betreft open data is het allemaal een stuk simpeler, zo stelt Noud Hooyman.

“Het is vanuit de overheid beleidsmatig en politiek bepaald om in te steken op open data. Als open data niet mogelijk zou zijn zonder Linked Data, dan zou ik er al heel snel een voorstander van zijn om te investeren in Linked Data. Ik heb een beetje het gevoel dat Linked Data een manier is, en misschien wel de beste manier, om gegevens op een kwalitatief goede manier ter beschikking te stellen zodat andere partijen daar hun voordeel mee kunnen doen. Ik zie die voordelen ook. Tegelijkertijd denk ik: misschien is een simpele API op een kwalitatief mindere manier ook goed genoeg. Dat is het grote dilemma. Mogelijk is het een oplossing om het geen Linked Data te noemen, maar FAIR-data (Findable, Accessible, Interoperable, Reusable). Dan zou je mij honderd procent mee hebben.”

BAG hét verbindende bestand

Als opdrachtgever is Noud Hooyman nauw betrokken bij de ontwikkeling van de BAG. Als er één basisregistratie uitermate geschikt is voor Linked Data, dan is het wat hem betreft de BAG. “De BAG is de spil, de combinatie tussen geografie en administratieve automatisering van de overheid. Het is bij uitstek het verbindende bestand en het is dan ook goed dat het Kadaster de BAG als Linked Data beschikbaar heeft gemaakt. Misschien kunnen wij, als het gaat om het leggen van verbindingen, als overheid nog een stapje verder gaan. Bijvoorbeeld door het gemakkelijker te maken dat private organisaties, of zoiets als Wikipedia, de BAG straks ook gaan gebruiken. Dat past ook goed in de discussie die wij nu aan het voeren zijn rondom open data. In eerste instantie ‘gooien’ wij die open data over de schutting, terwijl je door vanuit de overheid een paar dingen te veranderen, ervoor zorgt dat de data veel beter (her)bruikbaar is. Dat moeten we zorgvuldig doen, maar tegelijkertijd moeten we daarin niet te bang zijn. Sowieso heb ik niet de illusie dat ik het hele dataverkeer van de overheid kan sturen en regelen. Daarvoor wordt er te veel data van alle kanten heen en weer gebruikt. Maar als Linked Data ervoor kan zorgen dat gebruikers op een rechtstreekse manier data kunnen betrekken, dan vind ik dat erg interessant.”

PLDN-community

Wat in 2012 ooit begon als een pilot om te verkennen wat de mogelijkheden en kansen zijn van Linked Data in Nederland, is inmiddels uitgegroeid tot een volwaardig platform waar kennis en kunde rondom het thema wordt verzameld en uitgewisseld. Noud Hooyman draagt de community een warm hart toe, maar heeft tegelijkertijd ook wel een aantal ‘tips’ voor de PLDN-community. Met name als het gaat om het ‘verkopen’ van Linked Data richting bestuurders en beleidsmakers. “Die groep heeft vooral behoefte om te weten waar de voordelen van Linked Data zitten, waarbij er geredeneerd wordt vanuit de gebruiker. Waarom zouden bestuurders en beleidsmakers moeten investeren in Linked Data? En wat is de rol van Linked Data bij andere technologische ontwikkeling, zoals robotica, het gebruik van sensoren of de toename van zelfsturende auto’s. Hoe zit daar Linked Data in? Ook komen er enorme beleidsopgaven op ons af: klimaatwijziging, energietransitie, technologische revolutie rond mobiliteit, de gigantische toename van data. Daar moeten bestuurders en beleidsmakers op inspelen en welke rol heeft Linked Data daarin? Ik denk dat het goed is dat de community ook daar over nadenkt en helpt om ons, beleidsmakers en bestuurders, daarbij richting te geven.”

Erwin Folmer en Frits de Jong

Amplifying your spatial data on the web

As a student I spent what little spare cash I had on hi-fi equipment to play what was the technological marvel of the time Compact Disks. As a hi-fi nerd I was always advised to spend most of my money on one component, my systems amplifier. The hifi amplifier served two roles really; to amplify or boost the relatively weak signal coming out of the other components of your system and to provide a mechnansion for lining those components together.

When it comes to the web the 'amplifier' is web search, the mechanism we all use to find information online, and spatial data like much other scientific data published on the web is not getting amplified and is not reaching the widest possible audience. The joint W3C/OGC Spatial Data on the Web (SDW) working group is later this year publishing some Best Practice Guidelines to improve the situation.

Building upon the earlier work of the W3C Data on the Web Best Practices Group the aim is to provide a set of recommendations that are applicable to the publication of spatial or data about places on the Web.

Existing attempts to publish using approaches often termed Spatial Data Infrastructures (SDI) meet only limited success outside of expert communities, as catalog services intended to allow for discovering spatial data cannot make use of the amplifying effect of general purpose search engines of the Web.

The catalog services only provide access to metadata - and in general metadata that is focused on the needs of expert users - not the data itself, in hi-fi terms we have collated and organised our CD cases but the CD's themselves are elsewhere !

In the draft version of the SDW Best Practice a few recommendations have been made to address these problems and improve ability for data published via SDI's to be discovered and accessed. These include:

- Best Practice 4: Make your spatial data indexable by search engines
- Best Practice 7: Use globally unique persistent HTTP URIs for spatial things
- Best Practice 11: Expose spatial data through 'convenience APIs'

The collection of Spatial Data around the world to help build transport infrastructure, monitor agriculture, plan cities and provide health care represent a massive investment economically and socially. By taking a 'web friendly' approach we can maximise this investment by turning up the volume and finally reaching the ears of both decision makers and the citizens who ultimately benefit from its collection.

Ed Parsons

Geospatial Technologist at Google
(www.edparsons.com + www.google.com)

Open data vraagt om open standaarden

Open overheidsdata schept transparantie en voedt innovatie. We zien daarom dat een groeiend aantal overheidsorganisaties hun data online beschikbaar stelt. Ook komt er steeds uitdrukkelijker wetgeving die de overheid verplicht om de data die zij verzamelt en produceert, beschikbaar en herbruikbaar te maken.

Om open data te kunnen gebruiken en uitwisselen, moet de overheid deze publiceren in formaten die voldoen aan open standaarden. Open standaarden zijn geen 'nice to have'. Zij maken deel uit van de definitie van 'open government data' (OGD).

Het Forum Standaardisatie steunt actief het Platform Linked Data Nederland (PLDN) omdat overheidsorganisaties het belang van open data inzien. Samen met het PLDN wil het Forum Standaardisatie ervoor zorgen dat de overheid haar data publiceert in formaten die voldoen aan open standaarden en aan de wetgeving rondom herbruikbare overheidsinformatie.

De Nederlandse overheid stelde het Forum Standaardisatie ruim tien jaar geleden in om het gebruik van open standaarden in de publieke sector te stimuleren.

Het Forum Standaardisatie beheert de lijst van standaarden die overheidsorganisaties verplicht moeten uitvragen bij investeringen in ICT-projecten.

Op deze 'Pas toe of leg uit'-lijst vinden we onder meer de Linked Open Data-standaard SKOS. Ook benoemt de 'Pas toe of leg uit'-lijst verschillende standaarden voor metadata, zoals OWMS, OAI-PMH, NL-LOM en geo-standaarden. Daarnaast hanteert het Forum Standaardisatie een lijst van aanbevolen standaarden. Voor deze standaarden bestaat geen verplichting, maar ze hebben een zo duidelijke toegevoegde waarde en ze worden zo algemeen toegepast, dat het Forum Standaardisatie het gebruik ervan aanbeveelt. Op de lijst aanbevolen standaarden staan open data-standaarden, zoals RDF 1.1, RDFa 1.1, DCAT en OWL.

Op Europees niveau heeft het Forum Standaardisatie de standaarden RDF 1.1 en SKOS aangemeld bij het European Multi-Stakeholders Platform on ICT Standardisation. Deze expertgroep van de Europese Commissie identificeert standaarden die bij aanbestedingen in Europa uitgevraagd kunnen worden.

Het Forum Standaardisatie peilt periodiek de adoptie van open standaarden bij de overheid en publiceert de resultaten in de jaarlijkse Monitor. Zo helpt het Forum Standaardisatie de overheid de vinger aan de pols te houden wat betreft de toepassing van open standaarden bij haar organisaties. Maar bovenal wil het Forum Standaardisatie het gebruik van open standaarden bij de overheid stimuleren en ondersteunen.

“Data bij de bron is voor ons het uitgangspunt”

Drie basisregistraties als Linked Data; een dataplatform voor de publicatie van datasets gericht op gebruikers; een Linked Data-ontwikkelteam: Als het gaat om de ontwikkeling en het gebruik van Linked Data in Nederland, dan is het Kadaster één van de voorlopers. Linked Data is voor het Kadaster de weg naar de kern: data bij de bron. “Waarop je moet kunnen vertrouwen”, aldus Frank Tierolff.

Voor Frank Tierolff, bestuurder van het Kadaster, is het gebruik van Linked Data een logische stap. “Enkele jaren geleden was ik met andere datarijke uitvoeringsorganisaties al in gesprek over hoe onze data het beste geïnterpreteerd zou kunnen worden. Dat zou op een zodanige wijze moeten zijn dat andere partijen met de juiste interpretatie in staat zijn tot correcte toepassingen en juiste keuzes te komen. Een mogelijk scenario was daarbij om verbeterde metadata aan de data toe te voegen. Juist op dat moment kwam Linked Data als technologie voorbij: Daar kwam veel samen. Voor ons als Kadaster is het cruciaal dat je kunt vertrouwen op data. Die dient actueel en live opvraagbaar te zijn. Dat betekent dat data bij de bron voor ons het uitgangspunt is. Linked Data is een goede weg daar naartoe.”

Om meerdere redenen is data bij de bron voor het Kadaster essentieel, zo schetst Tierolff. “Als data niet bij de bron gebruikt wordt, heeft de eindgebruiker vaak geen zekerheid op welke data hij of zij beslissingen neemt. Die data kan een dag oud zijn, een week, maar ook een jaar. Daarnaast heeft de eindgebruiker in die

gevallen ook geen duidelijkheid of duplicatie van data wel goed is gegaan, of er geen informatie verloren is gegaan of dat data anderszins is gewijzigd. Een derde reden voor ons om data bij de bron als uitgangspunt te nemen, heeft te maken met de kosten. Het kopiëren van data, dat lang noodzakelijk was, kost veel geld. Dat kan vele malen efficiënter door data bij de bron te gebruiken. Linked Data en API's maken het mogelijk dat die data ook daadwerkelijk bij de bron kan worden opgehaald. Bovendien stelt die aanpak ons in staat om op een goede manier metadata toe te voegen aan de data.”

Het Kadaster staat bekend om zijn betrouwbare data. Die betrouwbaarheid wordt bereikt door actieve samenwerking met onze partners. “Daar worden we ook steeds om gewaardeerd.” De rol van betrouwbare dataleverancier kunnen we uitbouwen door het inzetten van Linked Data, zo stelt Frank Tierolff. “In de wereld van Linked Data ontstaan nieuwe datasets die de verbanden tussen bestaande datasets weergeven. Dat zullen ook datasets van anderen zijn, bijvoorbeeld van overheden, bedrijfsleven of communities. Maar ook voor deze nieuwe gelinkte datasets zullen gebruikers een hoge betrouwbaarheid eisen; vanuit het Kadaster gaan we die leveren.”

Betrouwbaarheid van het Kadaster staat ook centraal in het eigen ontwikkelde dataplatform dat past in het landelijke platform PDOK (data.pdok.nl). “Dit dataplatform zet datasets om naar Linked Data; daarnaast publiceert het API's en biedt het een plek voor



kennisuitwisseling door de community. Het biedt gebruikers, waaronder ontwikkelaars en data-analisten, handvatten om succesvol met data aan de slag te gaan. Ook het live bevragen van de data is een functionaliteit van het platform.”

Omgevingswet

Eén van de grootste uitdagingen voor overheden de komende jaren, is de invoering van de Omgevingswet. Ook voor het Kadaster betekent de Omgevingswet veel. “Het is voor ons een speerpunt dat verankerd is in onze drie ambities: ‘platform’, ‘zekerheid’ en ‘partner’. Het dataplatform dat wij ontwikkeld hebben, geeft invulling aan de platformambitie. Een platform stelt gebruikers in staat, onze data effectief en efficiënt te

“Linked Data kan maatschappelijke vraagstukken ondersteunen”

gebruiken. Als je naar de ambitie ‘zekerheid’ gaat dan vind ik het meegeven van betekenis aan data belangrijk, maar ook dat je de data bij de bron direct kunt ophalen. Huidige maatschappelijke vraagstukken zijn doorgaans complex en kunnen niet door één organisatie opgelost worden; met onze partnerambitie willen we, samen met partners, werken aan maatschappelijke vraagstukken. Die komen terug in de Omgevingswet. De techniek achter Linked Data maakt het mogelijk om die maatschappelijke vraagstukken te on-

dersteunen. Dat betekent dat wij niet alleen datasets en kennis aanbieden of ondersteunen bij het maken van een link, maar ook dat we die data relatief simpel ontsluiten. Dat kan met behulp van het dataplatform.”

Kadaster als informatie-organisatie

Linked Data, big data, machine-learning, deep learning, blockchain: Er zijn veel technologische ontwikkelingen die mogelijk relevant zijn voor het Kadaster. Frank Tierolff ziet het Kadaster vooral als een informatie-organisatie. “Een informatie-organisatie valt in beginsel uiteen in inhoud en techniek; die zijn echter steeds meer met elkaar verbonden. Voorheen waren binnen het Kadaster de business en IT van elkaar gescheiden; nu zie ik steeds meer logische samenwerking met

een gezamenlijk resultaat; dat is snel gegaan. De introductie van Scrum Agile heeft daar enorm aan bijgedragen, inclusief de bewustwording. Het is eigenlijk niet logisch meer om aan de informatiekant nog traditioneel

te spreken over datasets en dergelijke; het gaat om inzet van technologie teneinde oplossingen voor maatschappelijke vraagstukken te realiseren. In toenemende mate regelen we de dingen samen: In de architectuur, in projecten, in productontwikkeling. Er is steeds minder onderscheid.”

Frank Tierolff is niet verrast over het tempo van al die technologische ontwikkelingen. “Het gaat gigantisch snel. Het is kenmerkend dat het meestal begint bij

de techniek; mensen zijn in het begin waanzinnig enthousiast, daarna komt een tijd dat het allemaal wat lang lijkt te duren en dan volgt weer een tijd van versnellen. Ongetwijfeld zal het bij Linked Data niet anders zijn. Linked Data hebben wij daadwerkelijk geïmplementeerd. Met blockchain-technologie zijn we aan het experimenteren; inmiddels hebben we kennis over het onderwerp opgedaan en we doen mee met pilots of we initiëren zelf pilots. Ook voeren we pilots uit op het gebied van kunstmatige intelligentie. Ik denk dat je in de toekomst, bij het ophalen van data bij de bron, ook iets zal krijgen waar je gestructureerde en niet gestructureerde data samen kan gaan bevragen. Maar voor welke techniek je ook gaat, je zult in ieder geval voor jezelf de basisstappen moeten zetten. In onze ambities en in de realisatie van het dataplatform is Linked Data een belangrijke technologie, maar het zal nooit de enige technologie zijn. We houden alle mogelijkheden open.”

Een andere reden waardoor ontwikkelingen in een versnelling kunnen raken is als de business de potentie van kansen ziet. “Dat hebben we gezien met Linked Data. Als Kadaster hebben we daar ook aan meegegaan door vanuit de PLDN-community veel experts aan boord te halen. Dat is gecombineerd met de Scrum Agile-aanpak en de kennis die al in huis was. Een stip op de horizon geplaatst door de business, heeft business en IT in beweging gezet; samenwerking en actiegerichtheid zorgden voor resultaat. Zo ontstaat synergie in de eigen organisatie.

Nog voordat Linked Data in zicht was, waren we al bezig om semantiek toe te voegen, om met andere woorden, beter te beschrijven wat die data nou precies inhield. Door een uniforme (en kwalitatief solide) interpretatie binnen de organisatie te hanteren, kon daarmee ook naar buiten toe uniformiteit in definitie worden bewerkstelligd.”

Trots op behaalde resultaten

Momenteel biedt het Kadaster de datasets ‘de Digitale kadastrale kaart’ (Basisregistratie Kadaster), de ‘TOP10NL’ (Basisregistratie Topografie) en ook de BAG (Basisregistraties Adressen en Gebouwen) aan als Linked Data met API’s. Daarnaast is het Kadaster initiatiefnemer geweest van het dataplatform waar geodata van de overheid wordt gepubliceerd én omgezet naar Linked Data en API’s. Frank Tierolff is super trots op al die resultaten. “Ik vind het heel mooi dat we in zo korte tijd zoveel hebben bereikt. Ik houd regelmatig presentaties waar het gaat over de combinatie van business en informatiehuishouding en dan noem ik Linked Data ook steevast. Ook op dat punt is het Kadaster innovatief. Als organisatie beseffen wij dat misschien niet altijd zo, waardoor het ook niet zo snel naar buiten wordt gebracht. Maar het is wel iets waar wij als organisatie trots op mogen zijn. Naast Linked Data mogen we ook trots zijn op de automatische generalisatie van de topografische kaart en de Ketenintegratie Inschrijving Kadaster (KIK), het geautomatiseerde akteproces. Op drukke dagen wordt vijftig procent van het aanbod automatisch verwerkt. Het Kadaster heeft van oudsher dat betrouwbare imago,

maar het innovatieve karakter van onze organisatie is wat minder bekend. Daar mogen wij wel explicieter over zijn. Ik denk dat in de professionele wereld, door onze stakeholders, wel gezien wordt wat wij doen. Dat blijkt ook uit het feit dat we regelmatig benaderd worden voor taken, zoals de Landelijke Voorziening WOZ.”

Privacydiscussie

Voor Frank Tierolff is privacy erg belangrijk, maar niet per definitie aan Linked Data gekoppeld. “Het vraagstuk is overal waar data leidt of kan leiden tot het identificeren van een natuurlijke persoon. Dit speelt niet alleen als je data met elkaar combineert, maar zelfs als het gaat om de data van één set, heb je te maken met het privacyvraagstuk. Daarbij zijn er verschillende normen. Er is een wettelijke norm en een beleveningsnorm, waarbij de laatste soms bepalender is. Er worden wel discussies hierover gevoerd en dat zit toch nog ver uit elkaar. Als Kadaster voldoen wij uiteraard aan het wettelijk kader en dat wordt ook gecheckt. Zoals dat bijvoorbeeld is gedaan bij de kadastrale kaart op het moment dat die beschikbaar kwam als open data. Elke dataset die je combineert en die extra waarde toevoegt, levert potentieel een grotere kans op dat het privacygevoelige gegevens worden, ongeacht de techniek. Dat zal ook gebeuren op het moment ik meer papieren registers bij elkaar leg. Met Linked Data gaat het makkelijker en is het sneller zichtbaar; uiteindelijk is het vraagstuk min of meer hetzelfde. Nogmaals: Het hangt niet aan Linked Data, het speelt bij alle facetten van het verstrekken van data.”

Platform Linked Data Nederland

Het Kadaster draagt actief bij aan het Platform Linked Data Nederland (PLDN). Frank Tierolff is van mening dat het succes van Linked Data in Nederland afhangt van de mate waarin kennis gedeeld wordt en het tempo waarin de community groeit. Dat zijn belangrijke vliegwielen. “Het aantal deelnemers met datasets groeit en de community ook en ik denk dat PLDN daar een grote rol in speelt. Als je met tien man iets wil verbeteren, is de kans veel groter dat de kwaliteit van je oplossing toeneemt dan als je dat in je eentje doet. Het gaat om kennisdeling.”

Tierolff is er van overtuigd dat Linked Data op den duur meer mainstream wordt. “Dat komt vanzelf. Voor een groot deel gaat het om bewustwording. Als je nu eens begint met het beter uitdragen van de doelen, dat wil zeggen dat je beter uitlegt wat het belang is van het principe van “data bij de bron”. Dan kom je vanzelf op een punt dat data aan elkaar gelinkt zou moeten worden. Richt je daarbij zowel op informatie-architecten en technici als op gebruikers en besluitvormers. Dat zijn meestal niet dezelfde groepen. Als je dat kunt combineren, dan wordt je bereik veel groter. Je zou het ook kunnen koppelen aan maatschappelijke vraagstukken. Laat zien wat je aan een maatschappelijk vraagstuk met deze Linked Data-aanpak hebt kunnen toevoegen. Als Kadaster hebben wij in ieder geval die behoefte.”

Erwin Folmer en Frits de Jong

Belastingdienst - Fraud detection

Linked Data geeft een onverwachte boost aan het opsporen van fraude en fraudeurs. Spuurwerk dat een inspecteur nu twee weken kost, kan met behulp van Linked Data worden gedaan in vijf minuten.

Linked Data maakt gebruik van ‘identifiers’. Een identifier is een uniek kenmerk van een ding of een persoon. Dit unieke kenmerk wordt gebruikt om het ding of de persoon mee aan te duiden. Deze identifiers kunnen ook gebruikt worden als knooppunten van data. Verschillende databronnen kunnen op die manier met elkaar worden gecombineerd. De Basisregistraties Adressen en Gebouwen kan bijvoorbeeld gecombineerd worden met het Handelsregister, omdat dezelfde identifier voor het adres wordt gebruikt. Ook het Burgerservicenummer (BSN, of het RSIN - Rechtspersonen en Samenwerkingsverbanden Informatienummer - voor bedrijven) is een krachtig knooppunt.

De fraude-experts bij de Belastingdienst gebruiken in hun werk een visualisatie van de gelinkte dataknooppunten. Ieder knooppunt kan worden aangeklikt, waarna je toegang krijgt tot de daaraan gelinkte knooppunten. Dit geeft de fraude-expert

oneindig veel mogelijkheden voor het uitvoeren van zijn analyse. De visualisatie van het netwerk ondersteunt het denken van de fraude-expert. Hij behoudt overzicht, maar kan tevens ontdekken waar er mogelijk een nieuw verband is.

Ter ondersteuning van het spuurwerk naar fraude wordt gebruik gemaakt van de interface AllegroGraph. Deze interface wordt alleen gebruikt om te lezen. Input en update van data vindt plaats op de RDF-store. Om de relaties tussen gegevens te leggen, wordt gebruik gemaakt van standaarden voor identifiers.

De belangrijkste uitdaging in dit project ligt op het organisatorische vlak: toestemming om te investeren in deze technologie en mensen om het werk te doen. Inhoudelijke uitdaging ligt op het vlak van betekenis geven: welke knooppunten en welke relaties zijn relevant voor medewerkers die toezicht uitvoeren?

Naast de focus op het opsporen van fraude en fraudeurs, wordt Linked Data door de Belastingdienst vandaag de dag ook steeds vaker gebruikt om overzichten te genereren van bedrijfsstructuren, zonder dat er sprake is van fraude.



“Wat betreft Linked Data doet Nederland het zeker goed”

Een uitgave over Linked Data in Nederland kan niet zonder een gesprek met de meest geciteerde wetenschapper in ons land op dat gebied: Frank van Harmelen. Hij vindt dat wij hier in Nederland het zeker goed doen als het gaat om Linked Data, maar heeft en passant nog wel wat wensen.

Het zal rond 2007 zijn geweest dat Frank van Harmelen, informatica-hoogleraar aan de Vrije Universiteit Amsterdam en onder meer lid van de Europese Academie van Wetenschappen, pas echt in de gaten kreeg dat Linked Data groter was dan hij tot dan toe had gedacht en ook echt zou kunnen gaan werken. Aanleiding was de International Semantic Web Conference, die dat jaar gehouden werd gehouden in Busan, Zuid-Korea. “Vanuit de VU waren we sinds 1999 al wel bezig met het onderwerp, maar de eerste jaren was dat vooral academisch georiënteerd. Het was vooral een onderzoeksproject. Pas na 2007 kreeg ik eigenlijk pas goed in de gaten dat ook externe partijen zogenoemde knowledge graphs begonnen te maken. Met name in de hoek van de cultuur en de media gebeurde er veel. Ik weet nog dat ik ademloos heb zitten kijken naar wat de Britse omroep BBC er mee deed en ook Beeld en Geluid in Nederland was wat dat betreft goed bezig.”

Inmiddels zijn we tien jaar verder maar is Linked Data nog niet echt mainstream. Frank van Harmelen vindt dat dat vooral te maken heeft met het ontbreken van voldoende trainings- en communicatiemogelijkheden. “Toch ben ik steeds weer verbaasd als ik hoor en zie

welke organisaties inmiddels Linked Data gebruiken. Dat aantal stijgt de laatste tijd zo snel, dat ik het niet meer kan bijhouden.”

FAIR-data

Dat er, ondanks de schijnbaar hoge drempel, toch vaart zit in de ontwikkeling van Linked Data, bewijzen ook de stappen die in internationaal verband worden gemaakt. Opmerkelijk is dat daarbij steeds vaker de term FAIR-data (Findable, Accessible, Interoperable, Reusable) wordt gebezigd. Ook Frank van Harmelen ziet dat. “De Europese Commissie heeft het in hun beleidsstukken nu ook over FAIR-data. Zo staat in de nieuwe richtlijnen voor onderzoeksdata, die halverwege 2016 zijn opgesteld, bijvoorbeeld zwart op wit dat onderzoekers moeten gaan werken volgens de principes van FAIR-data. En ook tijdens de laatste top van de G20, de economische grootmachten in de wereld, werd het onderwerp behandeld en afgetikt. Dat zegt wel wat, want daar staan wel de handtekeningen onder van mensen als Obama en Merkel. Oké, zij zullen het zelf niet snappen, maar de mensen die er vlak onder zitten, snappen het blijkbaar wel. Wat mij betreft gaat FAIR-data over de doelen en Linked Data is de meest volwassen manier om die doelen te bereiken.”

Als het gaat over Linked Data, is er één ding dat Frank van Harmelen specifiek wil benoemen. Dat is de maatschappelijke discussie die gevoerd zou moeten worden over het delen van data en privacy. Privacy is niet specifiek voor Linked Data, maar Linked Data maakt de discussie erover wel zichtbaarder en daar

mee mogelijk gemakkelijker. “Dat betekent dat je niet met droge ogen met Linked Data bezig kunt zijn en dan zeggen dat je met die privacydiscussie niks te maken hebt.”

LOD Laundromat

Op 27 juni 2016 hield Frank van Harmelen een presentatie op het PLDN-congres te Eindhoven, waar hij de balans opmaakte van de ontwikkeling van Linked Data. Wat is er gerealiseerd en waar liggen de uitdagingen? Dat resulteerde in een Done-lijstje en een ToDo-lijstje (zie verderop). Een item dat zowel op het ToDo-lijstje van Frank van Harmelen staat, maar ook op het Done-lijstje, is de wens om meer Linked Data online te krijgen. Op dat gebied zijn er de afgelopen jaren

“De PLDN-community is iets waar veel landen jaloers op zijn”

in Nederland grote stappen gezet. Debet daaraan is onder meer LOD Laundromat, een toepassing die uit de koker komt van het onderzoeksteam van de VU Amsterdam en ervoor zorgt dat bestaande Linked Data beter vindbaar is en wordt verbeterd. Van Harmelen is supertrots op de toepassing. “Als er iets is wat lijkt op de Google van Linked Data, dan hebben wij die in Amsterdam. De LOD Laundromat. Een volgende ontwikkeling is het combineren van Linked Data met machine-learning. Die twee zijn complementair aan

elkaar, en naar beide kanten toe: met de semantische kennis in Linked Data kan machine-learning slimmer worden gemaakt, met betere resultaten tot gevolg. Aan de andere kant kan met machine-learning meer Linked Data worden gegenereerd.”

Streaming data

Eén van de dingen die nog op het ToDo-lijstje staat van Frank van Harmelen, zijn tools voor streaming data. “Met streaming data heb je het over data die continu binnenkomt en verandert. Uit sensornetwerken bijvoorbeeld. Met hoge frequentie komen daar duizenden, of honderdduizenden gegevens, per seconde binnen en daar wil je een semantische interessante analyse op kunnen doen. Smart cities is typisch

zo’n toepassing waar je hoogfrequente gegevens, bijvoorbeeld over weer of verkeer, binnenkrijgt en waarbij je al die gegevens ook wilt combineren. Dat is niet meer data waar je elke dag een update op doet, maar dat is data die constant stroomt. Op dit moment zitten wij nog erg in een denkkader van

betrekkelijk statische knowledge graphs, die met een lage frequentie veranderen. De vraag is wel hoe breed de uitdaging van die streaming data is. Er zijn sommige gebieden waar dat echt telt. Bijvoorbeeld als het gaat om smart cities of als je het hebt over Internet of Things. Dat zijn typisch toepassingsgebieden waar die streaming data erg belangrijk is. Maar in mijn visie is het niet een universeel probleem. Het is voor sommige gebieden wel een probleem en voor andere niet. Dat zullen we moeten uitzoeken.”

Nederland vs. de rest

Als gekeken wordt naar de rest van Europa, en de rest van de wereld, dan doet Nederland het niet slecht als het gaat om Linked Data. Zeker niet als gekeken wordt naar het academisch niveau. Dat vindt Frank van Harmelen ook. “Nederland doet het zeker goed. Sowieso doen wij het in Europa goed. Je zou ook kunnen zeggen dat het een van de weinige gebieden in de informatica is waar Europa vooroploopt. Als het gaat om Linked Data zijn ze ook in Engeland goed bezig op academisch- en onderzoeksniveau, maar dat geldt ook voor landen als Slovenië, Duitsland, Italië en Spanje. Het ligt iets anders bij de mensen die geld verdienen met de tooling die gebruikt wordt. Je ziet met name dat er nog ‘witte vlekken’ liggen bij de tools voor het maken van links tussen Linked Data. Ook wat betreft de overheid worden er, volgens Van Harmelen, in Nederland de nodige serieuze stappen gemaakt. “De (semi-)overheid loopt voorop. Een goed voorbeeld vind ik het Kadaster, dat de laatste jaren flink aan de weg heeft getimmerd heeft met het publiceren van basisregistraties op het gewenste vijfsterreniveau. Een groot aantal andere overheden bevinden zich nog op drie sterren niveau, maar dat is het mooie van Linked Data. Je kan stap voor stap zetten in plaats van de big bang-benadering. Het Nederlandse bedrijfsleven loopt in het geheel nog wat achter, hoewel met name in de sector van de logistiek, media, pharma en de bouwwereld er al het nodige gebeurt op het gebied van Linked Data.”

Platform Linked Data Nederland

Dat Nederland het zo goed doet op Europees niveau is, wat betreft Frank van Harmelen, voor een erg groot deel te danken aan het Platform Linked Data Nederland. “Het platform is cruciaal in die ontwikkeling. Ik weet nog de eerste keer dat ik uitgenodigd was bij het platform en toen dacht: hier zit de zaal vol met mensen van onder andere gemeenten en provincies die allemaal met Linked Data aan de gang zijn en ook nog eens actief zijn met het uitwisselen van kennis en kunde. De PLDN-community is iets waar veel landen jaloers op zijn. Dat maakt het allemaal ook zo mooi en compleet. De academische wereld heeft wel de kennis, maat niet de mogelijkheid of de tijd om dat te organiseren en dan is het goed dat het gebeurt door mensen die uit de praktijk komen. Zij weten wat er werkt en wat niet en weten ook waar de moeilijkheden, mogelijkheden en kansen zitten. Ik ben er inmiddels van overtuigd dat de community het meest waardevolle is dat wij hebben als het gaat om Linked Data. En er zal de komende jaren, zolang het onderwerp nog niet echt breed is geland, ook wel behoefte blijven aan een dergelijke community. Al is het alleen maar vanwege de brugfunctie. Een brugfunctie tussen de kennisoverdracht vanuit academie naar praktijk, maar ook kennisuitwisseling binnen die praktijk. Daarnaast zijn er nog genoeg partijen die behoefte hebben om uitgelegd te krijgen hoe de Linked Data-technologie werkt. Het platform kan die partijen helpen om daarbij stappen te maken en het geheel concreet te maken.”

Het wensenlijstje van Frank

Frank van Harmelen maakte tijdens het PLDN-congres te Eindhoven van juni 2016 de balans op van waar we staan met Linked Data. Wat is er gerealiseerd en waar liggen de uitdagingen?

"Waar staan we op dit moment? We kunnen concluderen dat Linked Data een breed gebruikt datamodel (RDF) heeft. Er is overeenstemming over hoe de semantiek met Linked Data beschreven kan worden. Ook wordt het belang van semantiek alom omarmd en zijn verschillende syntaxen succesvol gestandaardiseerd. De sterke mate van standaardisatie is sowieso de kracht van Linked Data. Verder zijn er tools beschikbaar voor het converteren van legacy data, en ook schaalbare opslag is gerealiseerd, waardoor er inmiddels al veel Linked Data online beschikbaar is.

Klagen kunnen we ook niet over de hoeveelheid toepassingen. Inmiddels zijn er te veel om bij te houden. Ook zijn er op bescheiden schaal tools beschikbaar om links aan te brengen tussen data, toch wel de kern van Linked Data. Daarnaast wordt binnen de overheid het publiceren van open data gestimuleerd, waarbij een Linked Data-aanpak wordt gebruikt op weg naar de vijf sterren.

Maar we zijn er nog niet... Zo zijn er te weinig kenners/toepassers, een significante drempel voor organisaties om aan de slag te gaan met

Linked Data. Dat betekent dat er meer en betere opleidingsmogelijkheden moeten komen met meer en beter opleidingsmateriaal. Daarbij is een betere user interfaces nodig bovenop de Linked Data, waarbij specifiek rekening moet worden gehouden met de eindgebruiker. Verder zien we dat er al veel Linked Data online beschikbaar is, maar er kan echt nog veel meer bij. Bijvoorbeeld op het gebied van streaming data. Belangrijk is ook dat er op het gebied van tools voor het linken van data, een professionaliseringsslag plaats moet vinden. De huidige tools zijn veelal academisch van aard en daarmee te beperkt. Een nieuw aandachtsgebied vormt verder gesloten data, al dan niet in combinatie met open data. Het is goed om te zien dat ook daar de techniek van Linked Data terrein aan het winnen is.

Stappen kunnen er ook worden gezet op het gebied van privacy. Op dat gebied is het noodzakelijk dat er de nodige garanties afgegeven kunnen worden. Ook als het gaat om provenance (herkomst) kunnen er stappen worden gezet. Provenance wordt een steeds belangrijker onderdeel van metadata. De standaarden zijn er inmiddels en nu het gebruik nog. Tot slot de kwaliteit, beschikbaarheid en betrouwbaarheid van de beschikbare Linked Data. Die moet beter. Maar desondanks kunnen we zeggen dat het goed gaat met de ontwikkeling van Linked Data."

to do list

~~Widely shared datamodel~~

~~Agreed upon semantics~~

~~Standardised syntax~~

~~Conversion tools for legacy data~~

~~Scalable storage and retrieval~~

~~Lots of Linked Data online~~

~~A wide variety of applications~~

~~Same tools for entity linking~~

~~Business models & incentives for Linked Open Data~~

Better educational material

Better user interface

More Linked Data online

Tools for streaming data

Better tools for entity linking

Business models & incentives for Linked non-Open data

Privacy guarantees

Provenance

Data quality and trust

Semantic Data in de agrisector

In de afgelopen jaren heeft TNO diepgaande kennis opgebouwd op het gebied van Linked (Open) Data en standaardisatie van datauitwisseling. Zo heeft het bijvoorbeeld, in het kader van het Smart Dairy Farming-project, bijgedragen aan het koppelen van verschillende databronnen binnen de melkveesector die nu als Linked Data bevestigd kunnen worden.

De Nederlandse melkveesector bestaat uit meer dan vijftienduizend bedrijven, waar in totaal meer dan anderhalf miljoen melkkoeien gehouden worden. Binnen de melkveesector wordt veel data verzameld. Deze data bevat een grote hoeveelheid informatie, waarmee het mogelijk is om de kwaliteit van leven van de koeien te verhogen. Zo kunnen, wanneer er op een juiste manier naar de data gekeken wordt, ziektebeelden eerder ontdekt en behandeld worden. Ook is het mogelijk om factoren te identificeren die invloed hebben op de kwaliteit en de kwantiteit van de melkgift van de koeien.

In de huidige situatie wordt al deze data niet voldoende benut. Het is voor de individuele veehouder niet rendabel om de data op het niveau van het individuele dier te analyseren. Toch bleek er vanuit de veeteeltsector wel behoefte aan een dergelijk systeem waarbij melkveehouders de data aanleveren, en vanuit dat systeem werkinstructies ontvangen op het niveau van het individuele dier. In het kader van het Smart Dairy Farming-project is zo'n systeem ontwikkeld. Binnen dit project deden zeven melkveehouders mee, die voor dit project data beschikbaar hebben gesteld. Zij

hebben meegedacht over de beoogde functionaliteit van de applicatie. Werkinstructies op het niveau van het individuele dier komen niet alleen het dierenwelzijn ten goede, maar ook de kosteneffectiviteit van het veeteeltbedrijf.

Een andere uitkomst van het meedenken was dat de veeteelthouder het belangrijk vindt dat hij de regie over zijn data heeft én blijft houden.

Infobroker

Om aan deze eisen te voldoen, is gekozen voor de ontwikkeling van een infobroker architectuur. Deze architectuur zorgt ervoor dat de data bij de veehouder blijft, waardoor hij de regie heeft en houdt over zijn eigen data. De veehouder zorgt voor de opslag van de data en kan zelf bepalen welke toegangsrestricties hij op deze data wil toepassen. Hij registreert deze databron bij de infobroker (of laat dit doen door een derde partij) en deze infobroker weet vervolgens welke data beschikbaar is. Deze data is vervolgens via de infobroker toegankelijk, door middel van een API.

Het idee is dat derde partijen diensten kunnen aanbieden aan deze melkveehouders, door bijvoorbeeld de data te analyseren en werkinstructies aan te bieden. Een andere mogelijke toepassing is dat de veehouder zijn data verkoopt aan onderzoeksinstellingen. Dit maakt het voor die instellingen relatief eenvoudig om toegang te krijgen tot een grote hoeveelheid data, en maakt het voor de veehouder mogelijk te verdienen aan het beschikbaar stellen van zijn data.

Dit alles was het doel van het eerste deel van Smart Dairy Farming. Het tweede deel van Smart Dairy Farming had als doel om het aantal deelnemende veehouders te vergroten (van zeven naar circa zestig, met de mogelijkheid om op te schalen naar vijftienghonderd deelnemende partijen). Tevens was het wenselijk om meer sensoren aan te sluiten op de infobroker, en meer partijen te betrekken die analyses op deze data zouden kunnen uitvoeren. Een ander doel was het toevoegen van semantiek aan deze databronnen. Dit zou gerealiseerd kunnen worden door de data aan de hand van een OWL-ontologie te beschrijven en als Linked Data beschikbaar te stellen. De data zou dan rechtstreeks als RDF te benaderen moeten zijn, en bevestigbaar door middel van een SPARQL endpoint.

Tweetal ontologieën

Het ontsluiten en beschikbaar stellen van data als Linked (Open) Data gaat via een door TNO ontwikkeld stappenplan, dat de kwaliteit van de resulterende data waarborgt. Dit plan standaardiseert het proces van het omzetten van data naar Linked Data. Het plan is toegepast bij het omzetten van de door de infobroker beschikbaar gestelde data.

Voor deze toepassing is een tweetal ontologieën ontwikkeld. De eerste was een eenvoudige sensor/meting-ontologie, die de sensoren en gemeten waar-

des beschrijft. Deze ontologie kan gebruikt worden om de ruwe sensor-metingdata te bevestigen. Daar bovenop is een melkvee-ontologie ontwikkeld, die het eenvoudiger moet maken om de melkveespecifieke vraagstukken te beantwoorden. Deze ontologieën zijn complementair aan elkaar. De melkvee-ontologie gebruikt concepten uit de sensormetingontologie.

Na het doorlopen van het Linked Data-stappenplan is het mogelijk om de infobroker-data door een SPARQL endpoint te bevestigen. Het grote voordeel van het beschikbaar stellen van data als Linked Data, is dat deze data eenvoudig te koppelen is aan andere relevante databronnen. Zo is het nu bijvoorbeeld mogelijk om een vraag te beantwoorden die ook de publieke databron van het KNMI bevestigt. Daarmee worden weer- en klimaatdata gebruikt bij het beantwoorden van vragen over melkproductie en dierenwelzijn.

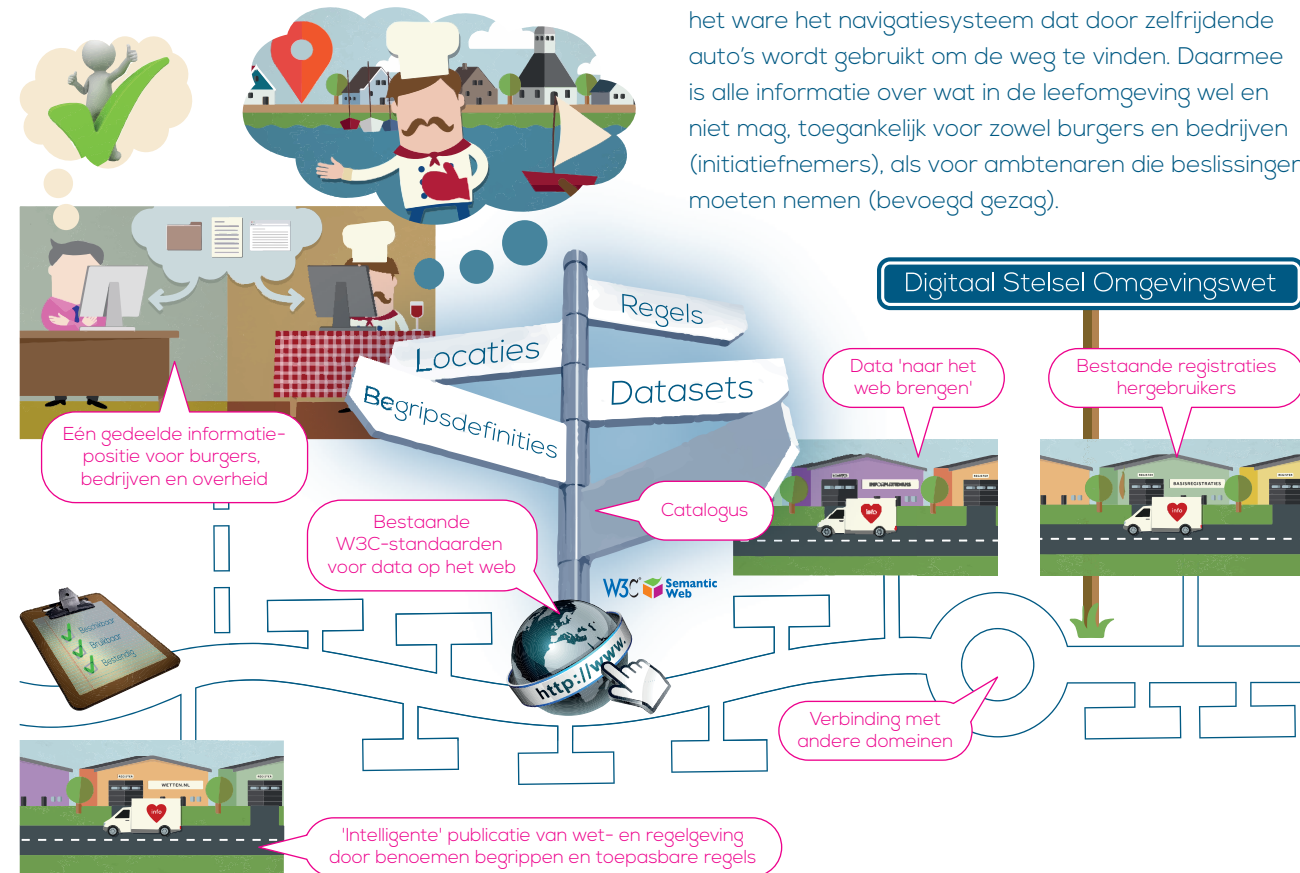
TNO werkt ook aan de implementatie van een Linked Data Platform. Het is een platform waarin alle applicaties die nodig zijn bij het uitvoeren van het stappenplan, geïntegreerd zijn. Dit platform moet het proces van het transformeren naar en het koppelen met Linked Data ondersteunen en verder standaardiseren.

Jacco Spek
TNO

De rol van Linked Data en REST API's bij de totstandkoming van de Omgevingswet

De Omgevingswet heeft als doel om het leven van burgers, bedrijven en overheden inzichtelijker, begrijpelijker en eenvoudiger te maken als het gaat om alle wetten en regels die gelden in de leefomgeving. Het Digitaal Stelsel Omgevingswet (DSO) zorgt ervoor dat alle informatie die hiervoor nodig is, ook daadwerkelijk beschikbaar is.

Om de weg te vinden in dit stelsel, is de Stelselcatalogus voorzien. Dit is een wegwijzer die de weg wijst langs alle begrippen, regels, datasets en locaties die van belang zijn voor de vragen waar burgers, bedrijven en overheden mee zitten. De catalogus volgt de best practices voor het publiceren van data op het web: Linked Data en REST API's. Hiermee is de catalogus ook leesbaar voor computers. De Catalogus is als het ware het navigatiesysteem dat door zelfrijdende auto's wordt gebruikt om de weg te vinden. Daarmee is alle informatie over wat in de leefomgeving wel en niet mag, toegankelijk voor zowel burgers en bedrijven (initiatiefnemers), als voor ambtenaren die beslissingen moeten nemen (bevoegd gezag).



Neem Bert. Hij wil een restaurant beginnen aan het water. Hij wil ook graag gasten lokken die met een boot komen en hen liefst ook een overnachtingsgelegenheid bieden aan een steiger of een haventje. Bert gaat op zoek naar de mogelijkheden. Hij zoekt op internet naar 'regels steiger jachthaven'. Omdat het digitaal stelsel de webstandaarden volgt, is alle informatie ook vindbaar voor zoekmachines. De catalogus is daarbij, zoals gezegd, het navigatiesysteem in een open stelsel.

De zoekmachine zal Bert doorleiden naar een officieel informatiepunt of informatiepunten die door derden zijn ingericht. Die informatiepunten kunnen bijvoorbeeld 'regelsindeleefomgeving.nl' of 'vergunningshulp.nl' heten. Vergunningshulp.nl kan gelijk een overzichtje van verwijzingen geven, waaronder relevante begrippen en regels in de wet. Bert vindt ook een kaartje waarop hij kan inzoomen en de plek kan aanwijzen waar hij een haventje wil beginnen. Vervolgens krijgt hij de aanvullende begrippen en regels van het betreffende waterschap, die zijn gepubliceerd in het Informatiehuis Ruimte. Er is nogal wat regelgeving rond 'milieubelastende en lozingsactiviteiten' voor jachthavens, maar Bert ziet al gauw dat die voor zijn situatie niet van belang is. Hij wil een recreatiehaventje en hij wil werkzaamheden aan boten verbieden.

Bert ziet ook een aantal regels over het 'afvoeren van vuil water'. Daar wil hij wat meer van weten. Hij

verwacht ook gasten van het strandje direct naast het perceel dat hij op het oog heeft. Hoe schoner het zwemwater, hoe meer badgasten. Bert ziet dat hij voor een vuilwaterafvoer een aansluiting op het vuilwaterriool moet maken als binnen veertig meter van de perceelgrens een vuilwaterriool ligt. Op basis van de coördinaten van het punt dat Bert in het begin heeft aangeklikt, is via de BRK (Basisregistratie Kadaster) de betreffende perceelgrens bekend. Via KLIC (Kabels en Leidingen Informatie Centrum) is direct zichtbaar dat binnen de vastgestelde afstand van de perceelgrens een vuilwaterriool loopt.

Bert's conclusie is dat hij een vergunning moet aanvragen voor zijn haventje en vult meteen het aanvraagformulier in, dat via een REST API ook beschikbaar is op vergunningshulp.nl. De meeste informatie wordt automatisch ingevuld, omdat die al in andere registraties bekend is. Zoals in de BRK en in de KLIC.

In de route op weg naar een nieuw restaurant aan het water, krijgt Bert ook te maken met Hans. Die werkt bij het bevoegd gezag dat de vergunning verleent. Hans kan, net als Bert, direct checken wat mag en niet mag en ziet in één oogopslag dat Bert alles goed heeft ingevuld en komt tot een voorgenomen besluit.

Marco Brattinga Ordina
Arjen Santema Kadaster

Beter onderwijs op maat met de Proeftuinen Linked Data

Het bieden van onderwijs op maat is een breed gedeelde wens van scholen. Dit betekent dat leraren de voortgang van leerlingen ten opzichte van leerdoelen moeten bijhouden en dat zij op zoek moeten naar het juiste leermiddel voor het juiste moment. Dat handwerk is arbeidsintensief en vaak ook omslachtig.

Door data aan elkaar te knopen ('Linked Data') zij leraren steeds beter in staat leerlingen op maat te begeleiden. De mogelijkheden die Linked Data biedt voor het onderwijs kunnen de leraar onder andere veel tijdswinst opleveren. Maar om dit voor elkaar te krijgen moeten leermaterialen wel eerst aan leerdoelen verbonden worden.

Die meerwaarde van een verbinding tussen leermaterialen en leerdoelen wordt kritisch onderzocht in de Proeftuinen van Kennisnet en SLO. In het verleden is al onderzoek gedaan met de Proeftuin Eindexamens in het voortgezet onderwijs en momenteel loopt er in het basisonderwijs een onderzoek met de Proeftuin Rekenen.

Proeftuin Eindexamens

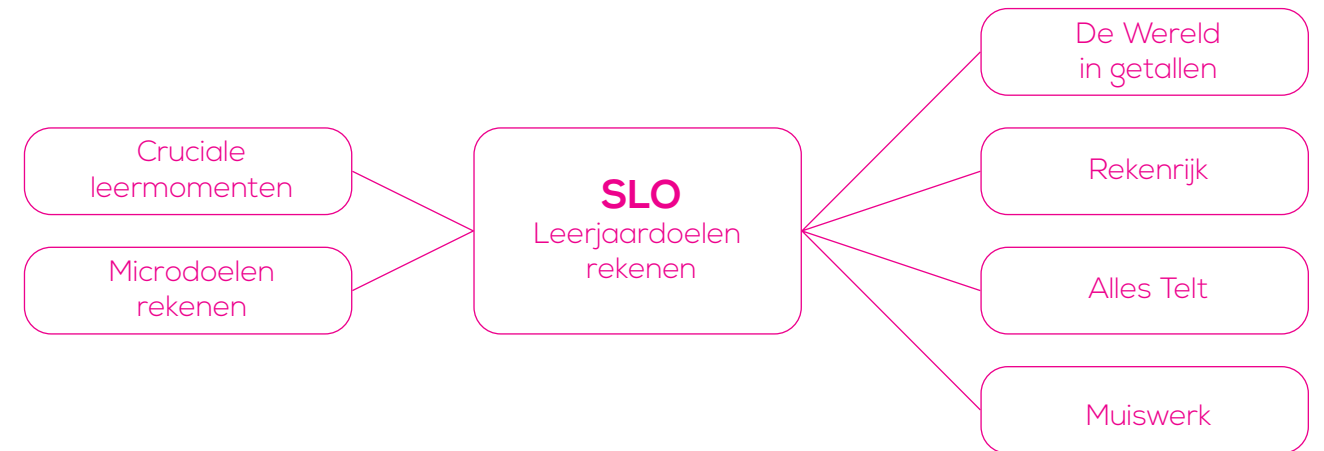
Hoe werkt het? De Proeftuin Eindexamens is een online applicatie die de leerling helpt te analyseren waar hij of zij staat ten opzichte van het eindexamen. De applicatie koppelt verschillende oefenexamens aan diverse methodes, waarbij de leerdoelen van kernprogramma's de verbindende schakel vormen. Door het maken van de oefenexamens krijgt de leerling inzicht in de mate waarin hij of zij onderwerpen en

vaardigheden beheerst ten opzichte van het landelijk gemiddelde. Op basis hiervan doet de applicatie een suggestie: deze hoofdstukken uit de methode, of uit zogenoemde 'open content', kun je doornemen om je kennis en vaardigheden bij te spijkeren.

De Proeftuin Eindexamens was beschikbaar voor leerlingen van havo 4 en 5, voor de vakken economie, aardrijkskunde, Engels, management & organisatie, biologie, scheikunde, wiskunde en natuurkunde.

Hieronder volgt een aantal bevindingen:

- Leerlingen kregen beter inzicht in de eigen sterktes en zwaktes waar het ging om beheersing van de leerstof.
- Leerlingen kregen een beter beeld van de veelvoorkomende onderdelen en onderwerpen in het eindexamen en van de onderdelen die belangrijk zijn.
- Docenten vonden de kernprogramma's bij het oefenen van eindexamens van voldoende kwaliteit. De begrippen en termen noemden zij herkenbaar.
- Docenten voelden zich geholpen bij het aanbieden van onderwijs op maat doordat ze nu per leerling inzicht hadden in de mate van beheersing van de leerstof.
- Docenten hadden behoefte aan een uitbreiding van de rapportage. Meer specifiek: aan een klassenoverzicht zodat meerdere leerlingen die een onderwerp niet goed begrepen, als groep begeleid zouden kunnen worden.



Proeftuin Rekenen

In het basisonderwijs wordt het onderzoek op dit moment vervolgd met de Proeftuin Rekenen. Hoe werkt het? In de Proeftuin Rekenen vormen de leerjaardoelen Rekenen PO van SLO de centrale kapstok. De leerjaardoelenstructuur bestaat uit ongeveer zevenhondervijftig leerdoelen en bevat een set doelen per leerjaar van het basisonderwijs. Op deze manier heeft de leraar een richtlijn die hij kan gebruiken om zijn onderwijs vorm te geven wanneer hij niet met de structuur van een lesmethode werkt of flexibel wil omgaan met de lesmethode. Aan de leerjaardoelen zijn twee andere leerdoelenstructuren en vier veelgebruikte rekenmethodes gekoppeld.

Stel: een leraar werkt met de methode 'Alles Telt!' en hij behandelt met zijn leerlingen de opdracht 'Dozen tellen'. Deze opdracht is gekoppeld aan het

leerjaardoel 'Ik kan tot tien tellen'. De leerlingen behalen resultaten door het maken van deze opdracht. De leraar kan vervolgens via het voortgangsoverzicht zien welke leerlingen moeite hebben met tot tien tellen. Via een klik bij de opdracht 'Dozen tellen' ziet hij bovendien dat dit leerjaardoel ook is gekoppeld aan drie andere rekenmethodes. De leraar kan op deze manier alternatieve opdrachten klaarzetten voor leerlingen die moeite hebben met het onderwerp, zodat zij extra kunnen oefenen.

Doordat er twee andere leerdoelenstructuren zijn verbonden aan de leerjaardoelen van SLO, kan de leraar er ook voor kiezen om met deze leerdoelen te werken. Op het moment dat hij bijvoorbeeld voor Cruciale Leermomenten kiest, zal die informatie zichtbaar worden als leerdoelinformatie bij de leermaterialen en in het voortgangsoverzicht.

KOI zorgt voor sectoroverstijgende gegevensuitwisseling

Onderzoek

Het onderzoek naar het gebruik van de Proeftuin Rekenen in de klas richt zich onder andere op de volgende vragen:

- Hoe zorg je ervoor dat leermaterialen bij leerdoelen passen?
- Hoe zorg je ervoor dat getoonde informatie in een voortgangsoverzicht betrouwbaar is?
- Wat is mogelijk met bestaande leerdoelstructuren, leermaterialen en huidige voorzieningen?
- Waar liggen knelpunten die moeten worden opgelost om te komen tot betrouwbare informatie?
- Hoe sluit het lesgeven vanuit leerdoelen in plaats van methodes aan bij de eisen van de onderwijsinspectie?

De lessen die uit de Proeftuin Rekenen worden getrokken, worden breed gedeeld zodat iedereen hiervan kan leren: scholen met betrekking tot vraagverheldering, leveranciers met betrekking tot product(door)ontwikkeling en de publieke sector met betrekking tot het aanscherpen van voorzieningen.

De pilots met de Proeftuin Rekenen lopen tot juli 2017. De resultaten van het onderzoek worden eind 2017 gedeeld via www.kennisnet.nl.

De sectie Onderwijs binnen het Platform Linked Data Nederland richt zich op het inventariseren en beschrijven van initiatieven binnen het onderwijs op het gebied van Linked Data. Hier gaat het in eerste instantie om het digitaal ontsluiten van het curriculum, maar ook de koppelingen naar bijvoorbeeld uitgever, leermiddelcollecties, onderzoeksgegevens, professionaliseringsmateriaal en geodata. De activiteiten richten zich op:

- Het beschrijven van projecten;
- Voorbeelden geven over het gebruik van Linked Data binnen het onderwijs;
- Werkwijze en methodiek beschrijven voor het ontwikkelen Linked Data;
- Bronnen verzamelen wat betreft gebruik van Linked Data in het onderwijs;
- Verbindingen aangeven tussen lopende initiatieven en de rol van onderwijs toepassingen daarin.

Lalibel Mohaupt Kennisnet

Kees Hoogland SLO

Binnen onderwijssectoren en ketenprocessen vindt veel gegevensuitwisseling plaats. Hoewel deze gegevens veel overeenkomsten hebben, zijn er nog wel sterke verschillen in gegevensaanduiding en -definities. Het Kernmodel Onderwijsinformatie (KOI) zorgt ervoor dat sectoroverstijgende gegevensuitwisseling mogelijk is.

In het onderwijs zijn veel verschillende organisaties bezig met het verwerken van informatie. Daarbij gaat het niet alleen om onderwijsinstellingen, maar bijvoorbeeld ook om educatieve uitgeverijen en leveranciers van elektronische leeromgevingen. Bij het uitwisselen van gegevens tussen deze organisaties is semantische overeenstemming noodzakelijk. Semantische overeenstemming houdt in: duidelijkheid over de overeenkomsten en over de verschillen tussen begrippen. Als je bijvoorbeeld op de basisschool zit, ben je een leerling. Op de middelbare school ben je een scholier. Afhankelijk van het soort onderwijs dat je daarna volgt, ben je een deelnemer of een student. Daarnaast kan het voorkomen dat dezelfde aanduidingen worden gebruikt voor wezenlijk andere betekenissen. Denk bijvoorbeeld aan het begrip 'begindatum van de opleiding'. Is dat de beoogde datum waarop iemand aan zijn opleiding begint, de daadwerkelijke startdatum of de inschrijvingsdatum?

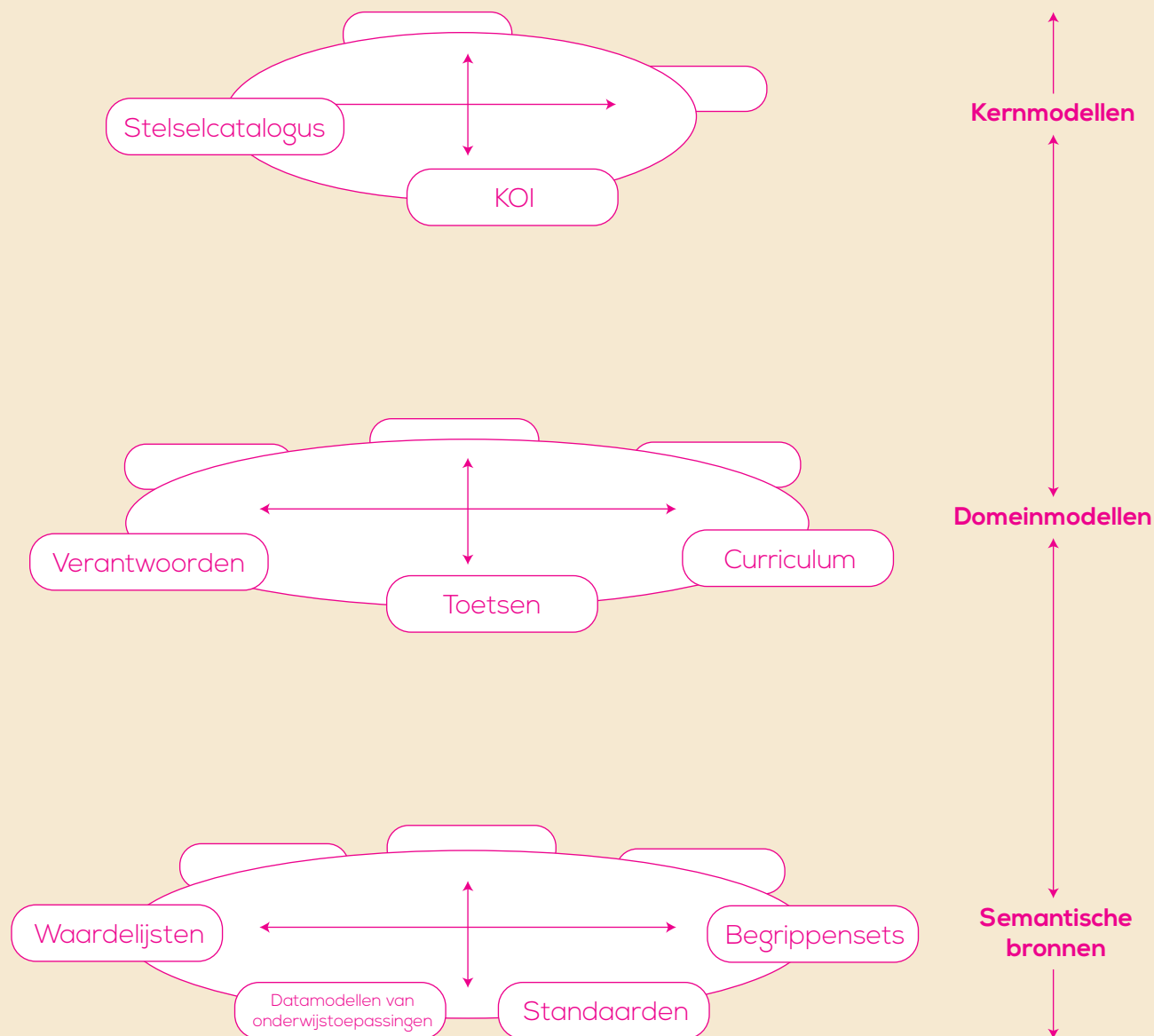
Modellen

Het gebruik van termen met verschillende betekenissen en verschillende betekenissen van een term kan een accurate en soepele

gegevensuitwisseling in de weg staan. Daarom werkt het onderwijs aan een aantal modellen. Een voorbeeld daarvan is het domeinmodel Toetsen. Dit model legt een gemeenschappelijke begrippenset vast die alle belangrijke toetsprocessen dekt: het construeren van een toets, het plannen daarvan, het afnemen van de toets, het verwerken van de verzameling afnamegegevens en het beschikbaar stellen van de resultaten. Het domeinmodel Toetsen is toepasbaar in alle onderwijssectoren en heeft betrekking op alle soorten digitale en papieren toetsen: van oefeningen tot landelijke centrale examens. Alle begrippen in domeinmodellen, zoals het domeinmodel Toetsen, zijn gerelateerd aan het Kernmodel Onderwijsinformatie.

Transparantie en inzicht

Het Kernmodel Onderwijsinformatie (KOI) zorgt ervoor dat organisaties in het onderwijs sectoroverstijgend gegevens kunnen uitwisselen. Het is een relationeel model waarin betekenis, gebruik en vorm van gegevens in samenhang worden gebracht. Het kernmodel biedt transparantie en inzicht in het bestaan van verschillende terminologie binnen de verschillende onderwijssectoren. Ook overbrugt het model die verschillen zonder een gemeenschappelijke taal voor te schrijven. Het KOI is een model in beweging. Dat houdt in dat het meegaat met nieuwe inzichten en veranderingen en er dus begrippen bij kan krijgen. Het kan er ook toe leiden dat er andere relaties tussen begrippen kunnen worden gelegd of dat begrippen weer uit het KOI worden gehaald. Het KOI is een onlosmakelijk onderdeel van de sectoroverstijgende



onderwijsarchitectuur ROSA. Met de begrippen van het KOI kun je de weg vinden naar begrippen in andere modellen en semantische bronnen. Door de sterke verwevenheid met ROSA is het KOI een doeltreffend hulpmiddel om semantische knelpunten en kansen te ontdekken. Zowel ROSA als het KOI vallen onder governance van Edustandaard, het platform waar alle partijen binnen het onderwijsveld bij elkaar komen om afspraken te maken. Zo wordt bereikt dat het model en de bijbehorende begrippen direct toepasbaar zijn in het onderwijs.

Semantisch landschap

Een van de doelen van Edustandaard voor 2017 is het semantische landschap en de onderlinge relaties tussen begrippen verder uitbreiden en inzichtelijk maken voor de betrokkenen. Daardoor zullen nieuwe ketenuitwisselingen geënt worden op bestaande, gestandaardiseerde begrippen. Immers, de kracht van Linked Data is dat delen van bestaande modellen uit verschillende domeinen en toepassingen met elkaar gecombineerd kunnen worden. In 2016 is Edustandaard gestart met het in kaart

brengen van de begrippen die aan de orde zijn bij onderwijsstandaarden. Een van die standaarden heeft betrekking op de uitwisseling van leerlinggegevens en resultaten. Als een school bijvoorbeeld gebruikmaakt van een applicatie van een uitgever waarmee leerlingen worden ondersteund bij het leren, moeten de resultaten daarvan kunnen worden overgedragen naar het leerlingvolgsysteem van de school. De digitale overdracht scheelt docenten veel tijd en bovendien voorkomt het fouten bij het overtypen van gegevens. Voordeel van een standaard voor deze uitwisseling van leerlinggegevens en resultaten is dat marktpartijen, zoals educatieve uitgeverijen en leveranciers van leerlingadministratiesystemen, minder koppelingen voor de uitwisseling hoeven te ontwikkelen en te onderhouden. Edustandaard heeft bekeken welke begrippen aan de orde zijn bij deze standaard en bij andere standaarden. Deze inventarisatie is een onderdeel van het inzichtelijk maken van het semantische onderwijslandschap met de centrale KOI-begrippen als 'linking pins'. Daarmee helpt het KOI de hele onderwijssector in Nederland om te komen tot een gemeenschappelijke informatiehuishouding.

Elise Lustenhouwer en Jeroen Hamers
Kennisnet

Digitale geletterdheid in het onderwijs en de impact van Linked Data

De afgelopen jaren zijn door technische ontwikkelingen de mogelijkheden van Linked Data toegenomen. Binnen het onderwijs zijn verschillende pilots, zoals de adaptieve eindexamentrainer en de Proeftuin Linked Data – Rekenen PO uitgevoerd om de meerwaarde van Linked Data te onderzoeken. Door de onderzoeken naar de mogelijkheden worden ook de aandachtspunten duidelijker: wie is eigenaar van de gegevens, welke gegevens mogen gedeeld worden, wat is de waarde van de data, wat is de betekenis van de data, hoe moet de data geïnterpreteerd worden en welke mogelijkheden zijn er nu en welke zijn er nog te verwachten? Het koppelen van data en gegevens maakt het ook steeds beter mogelijk om daarmee te redeneren en vragen te beantwoorden die eerder te complex waren.

Ook het nadenken over de mogelijkheden van technologie en de grenzen van wat kan en mag wordt steeds belangrijker. In 2012 constateerde de KNAW al dat in het huidige onderwijs te weinig aandacht is voor digitale geletterdheid in het onderwijs. Digitale geletterdheid is een combinatie van ICT-basisvaardigheden, informatievaardigheden, mediawijsheid en computational thinking. Deze vaardigheden zijn onderdeel van de 21ste eeuwse vaardigheden die gezien worden als de basis voor het onderwijs van de toekomst. Niet alleen zijn deze vaardigheden voor leerlingen van belang, maar in toenemende mate ook voor leerkrachten. Met name omdat Linked Data in het vormgeven en verantwoorden van onderwijs een sterk toenemende rol gaat spelen.

ICT-basisvaardigheden

Onder ICT-basisvaardigheden wordt verstaan de kennis en vaardigheden die nodig zijn om de werking van computers en netwerken te begrijpen, om te kunnen omgaan met verschillende soorten technologieën en om de bediening, de mogelijkheden en de beperkingen van technologie te begrijpen. Het begrip computer wordt hier breed gebruikt. Niet alleen als personal computer, desktop of tablet, maar elke technologie waarin een microprocessor is gebruikt die op basis van ingevoerde gegevens volgens een programma een aantal logische handelingen verricht, met als uitvoer bepaalde algoritmes en tijdelijke opslag van gegevens.

Als het gaat om Linked Data, dan gaat het hierbij ook om de interactievormen en de wijze waarop we de data ontsluiten. Het om kunnen gaan met apparaten is daarbij een vaardigheid die veel verder gaat dan klikken met een muis en typen van een tekst met een toetsenbord. Door de hoeveelheid van de data en gegevens is het ook belangrijk om logisch te leren navigeren, functies en applicaties te leren kennen om data vorm te geven en te ontsluiten via spraak, interactieve schermen en bewegingen. De spraakherkenning van de mobiele telefoon heeft bijvoorbeeld (nu nog) een internetverbinding nodig om te werken. Het besef van infrastructuur om data te koppelen, te ontsluiten en te gebruiken is daarbij essentieel. Dat is ook een grote uitdaging voor de inrichting van scholen en het ontwikkelen van de gewenste vaardigheden bij leerkrachten.

Informatievaardigheden

De informatievaardigheid omvat het scherp kunnen formuleren en analyseren van informatie uit bronnen, het op basis hiervan kritisch en systematisch zoeken, selecteren, verwerken, gebruiken en verwijzen van relevante informatie en deze op bruikbaarheid en betrouwbaarheid beoordelen en evalueren.

De impact van Linked Data op de informatievaardigheden is groot. Informatie bestaat steeds vaker uit gecombineerde gegevens uit verschillende bronnen. Daarnaast wordt er steeds meer op basis van verschillende bronnen geredeneerd. Vragen in natuurlijke spraak voor het bedienen van de mobiele telefoon zijn normaal geworden. Bijvoorbeeld de vraag: 'Hoe oud is mijn vader' wordt in natuurlijke spraak beantwoord op basis van de 'lokale' gegevens op de telefoon. Context en relaties, geboortedata, namen in de adresboeken zijn data waarover geredeneerd wordt. De vraag of hiervoor Outlook, Facebook, LinkedIn, iCloud of Gmail als bron is gebruikt, is onduidelijk. Keuzes voor de gebruikte en samengestelde data liggen bij ontwikkelaars van de software en zijn voor gebruikers ondoorzichtig. Dit voorbeeld illustreert dat de consequenties voor zichtbaar logische informatie in een gebruikerscontext vraagt om een kritische en onderzoekende houding. Deze informatievaardigheden worden steeds relevanter bij meer samengestelde gegevens waar bronnen onduidelijk zijn of waarbij het aantal gelinkte componenten groot is.

Mediawijsheid

De term 'mediawijsheid' werd in 2005 geïntroduceerd in een advies van de Raad voor Cultuur ('Mediawijsheid: de ontwikkeling van nieuw burgerschap'). In het advies benadrukt de Raad hoe groot de impact van media op ons bestaan is: 'Weinig blijft onberoerd door het effect van media; media worden steeds meer context, inhoud en bemiddelaars van informatie, kennis en ervaring [...] Van elementen in een omgeving zijn media de omgeving zelf geworden.' In deze gemedialiseerde samenleving hebben burgers nieuwe competenties nodig om optimaal te kunnen functioneren, produceren en participeren. De Raad vat deze competenties samen in het begrip mediawijsheid, dat gedefinieerd wordt als 'het geheel van kennis, vaardigheden en mentaliteit waarmee burgers zich bewust, kritisch en actief kunnen bewegen in een complexe, veranderlijke en fundamenteel gemediatiseerd wereld'.

Met betrekking tot Linked Data is het van belang om te weten van wie de gegevens zijn die gebruikt worden. Hierbij speelt eigen privacy en dat van anderen een belangrijke rol. Welke gegevens mogen gebruikt worden en wat mag er wel en niet gekoppeld worden? Inzicht in eigen dataproductie en beheer van eigen profielen voor sociale media zijn hierbij aandachtspunten. Het bewust omgaan met privacygevoelige data is een vaardigheid die zeker voor Linked Data een belangrijk aspect is omdat het combineren van verschillende gegevens tot privacyproblemen kan leiden.

linked data, zonder hoofdletters

De afgelopen jaren heeft linked data een grote vlucht genomen. In Nederland niet in het minst door initiatieven zoals PLDN. Bedrijven en overheden kiezen steeds vaker het linked dataparadigma voor het communiceren van data en metadata.

Maar we zijn er nog niet. Het gebruik van linked data heeft veelal nog de vorm van datapublicatie: een selectie oorspronkelijke 'legacy' data wordt naar buiten toe blootgesteld als linked data. Hoewel dit een grote verbetering is ten opzichte van de vroegere situatie, met name door de globaal unieke identifiers, is linked data daarmee toch nog vaak de kers op de taart. Een extraatje, eerder dan een voorwaarde voor een efficiënte databedrijfsvoering.

Waarom is dit? Ik denk dat hier twee oorzaken aan ten grondslag liggen. In 1980 maakte Alan Newell een methodologisch onderscheid tussen vier verschillende niveaus waarop over computersystemen gesproken kan worden: het fysieke, het ontwerp-, het symbolen het kennisniveau. In linked data gaat het om een briljant idee op het kennisniveau - hoe informatie en kennis vervlochten is en betekenis ontstaat door onderlinge afhankelijkheid - dat een representatie kent op symboolniveau, waar we gebruikmaken van webtechnologie.

Communicatie over linked data vindt echter bij grote uitzondering op het kennisniveau plaats: zelden gaat er een presentatie, artikel of blog voorbij waarbij geen Turtle, SPARQL queries of URIs getoond worden.

De tweede oorzaak is een historische: het technologiehuis rond linked data is grotendeels naast bestaande technologie opgetuigd. Het datamodel van linked data zou speciale graph-databases vereisen, de huidige triple stores. Dit is waar, voor zover het noodzakelijk is om linked data grafen generiek te kunnen bevragen.

Het nadeel van deze aanpak is dat aansluiting met bestaande IT-infrastructuur stroef verloopt. De adoptie van het idee van linked data op kennisniveau lijkt te vereisen dat er tegelijkertijd een transitie op symboolniveau plaatsvindt: naast implementatiekosten is er een vrij steile leercurve en het nadeel van minder 'volwassen' technologie. Dit is niet waar.

De meeste toepassingen vragen een stabiel datamodel waarvoor de flexibiliteit van een graph store niet noodzakelijk is. Recent onderzoek van Minh-Duc Pham en Peter Boncz aan het CWI en de VU laat zien dat linked data latente, stabiele structuren bevat die gebruikt kunnen worden om de performance van query-answering significant te verbeteren.

Voor verdere adoptie van linked data is het essentieel dat linked data haar weg vindt binnen bestaande IT-infrastructuur. Dit vraagt flexibiliteit van twee kanten. Inderdaad, linked data, zonder hoofdletters.

Rinke Hoekstra

Department of Computer Science, Vrije Universiteit Amsterdam - Faculty of Law, University of Amsterdam (www.rinkehoekstra.nl)

Computational thinking

Computational thinking is het procesmatig (her)formuleren van problemen op een zodanige manier dat het mogelijk wordt om met computertechnologie het probleem op te lossen. Het gaat daarbij om een verzameling van denkprocessen waarbij probleemformulering, gegevensorganisatie, -analyse en -representatie worden gebruikt voor het oplossen van problemen met behulp van ICT-technieken en -gereedschappen.

De hoeveelheid Linked Data maakt het steeds beter mogelijk om problemen op te lossen. Ook vragen die eerst onmogelijk leken op te lossen, worden steeds verder omgezet in computervraagstukken die misschien niet zo moeilijk zijn. Steeds meer data is beschikbaar en er kan steeds verder geredeneerd worden met de data. Een vraag zoals: 'Hoe groot is dat transformatorhuisje' kan dan ook met de huidige datasets misschien wel een antwoord opleveren. Een transformatorhuisje is bekend bij het Kadaster, de afmetingen ook en het over 'dat' transformatorhuisje

hebben betekent dat het gaat om het dichtstbijzijnde transformatorhuisje hebben ten opzichte van de huidige locatie. Computational thinking gaat over de wijze hoe je problemen oplost met de computer, maar ook dat er besef is over welke gegevens beschikbaar zijn, hoe je deze ontsluit, koppelt engebruikt. De vraag rijst welke mate van computational thinking een leerkracht van de toekomst nodig heeft om verantwoord om te gaan met allerlei gelinkte onderwijsdata.

Linked Data en onderwijs

Om leerlingen goed toe te rusten voor de informatiemaatschappij, is digitale geletterdheid van groot belang. Maar niet minder zijn deze vaardigheden van belang voor leraren. In toenemende mate worden zij bij het vormgeven van hun onderwijs, al dan niet bewust, geconfronteerd met grote hoeveelheden Linked Data. Bij het vinden van geschikt materiaal, bij het volgen van leerlingen en bij het verantwoorden van de opbrengsten van hun onderwijs. Het toerusten van leraren om hier bewust mee om te gaan, is daarom voor het onderwijs essentieel.

Allard Strijker, Hans de Vries, Kees Hoogland, Frederik van Oorschot en Petra Fisser
SLO

Groenten uit Vlaanderen

In Vlaanderen wordt veel onderzoek gedaan naar Linked Data, maar het aantal applicaties op het gebied van Linked Open Data is nog beperkt. Toch zijn er wel voorbeelden.

De oertijd

De data van de KBO (Kruispuntbank voor Ondernemingen) werden in mei 2014 als open data gepubliceerd. Enkele LOD-fans hebben deze data verwerkt tot een Linked Datatoepassing met: dataset downloads, subject pages, SPARQL endpoint, Linked Data fragments endpoint, full text search en Reconciliation API.

Op federaal vlak

Fedstats België heeft de brondata van twee door hen beheerde nomenclaturen uitgebracht: de statistische nomenclatuur van de economische activiteiten in de Europese Unie en de NIS-codes (codes voor het aanduiden van de verschillende administratieve regio's) die door het Europees OpenCube project gepubliceerd werden als LOD. Verder worden in 2017 de eerste eigen federale stappen gezet op het vlak van Linked Open Data.

Westtoer

Westtoer is het autonoom provinciebedrijf voor toerisme en recreatie in West-Vlaanderen die hun (toeristische) data als Linked Data publiceren.

Op Vlaams vlak

In Vlaanderen wordt een zeer ambitieus plan uitgevoerd om alle basisregisters semantisch te modelleren en als LOD te publiceren. Momenteel wordt de laatste hand gelegd aan de Vlaamse URLstrategie. Eind maart (2017) wordt het Centraal Referentieadressenbestand (CRAB) als eerste dataset gepubliceerd. Het Agentschap Binnenlands Bestuur heeft een succesvolle pilot achter de rug met een Linked Open Data Editor voor gemeentebesluiten. Het departement Onroerend Erfgoed heeft zijn eerste stappen gezet.

World class research

In de semantische wereld is idlab, RUGent/imec het actiefst. Vanuit hun hoek komt het Linked Data Fragments idee om intelligentie van de server naar de cliënt te verhuizen als antwoord op het 'SPARQL endpoint down'-fenomeen. Verder werken zij aan RML, een mapping specificatietaal die naast RDBMS bronnen, ook CSV, XML en JSON, aankan en verder nog andere limitaties van bestaande mapping methodologieën adresseert.

Paul Hermans

Oprichter van ProXML (www.proxml.be)

P.S. Wie kan uitleggen waarvan de vreemde titel komt, wint een halve dag gratis vocabulary consulting.

#Geo4web testbed

Internet is een prachtig middel om digitale informatie op te delen. Die kracht van het web wilden we bij Geonovum graag benutten om geo-informatie beter toegankelijk te maken. Om te ontdekken op welke manier, organiseerde Geonovum in 2016 een testbed waarmee ervaring is opgedaan met diverse technieken, waaronder Linked Data.

Het testbed Geodata op het web (#geo4web testbed) draaide om een aantal onderzoeksvragen. Kort gezegd ging het erom geodata op zo'n manier te publiceren dat het voor webontwikkelaars en geautomatiseerde 'gebruikers' van data (zoals zoekmachines), beter te vinden en te gebruiken zou zijn. Daarbij is ook gekeken naar wat voor inspanning dat vraagt van data-eigenaren. In de geo-wereld zijn veel open-datasets en een volledig op open standaarden gebaseerde infrastructuur om die data te delen, maar de gebruikte standaarden zijn specifieke geo-standaarden, in plaats van webstandaarden.

Eén van de resultaten van het testbed was de tool LD-Proxy. Hiermee krijgt elk geo-object een eigen, stabiele URI. Dit is een fundamentele stap die ervoor zorgt dat er naar deze geo-objecten gelinkt kan worden. Simpelweg de link volgen levert informatie over deze objecten op. Bovendien kunnen zoekmachines nu de objecten ontdekken en opnemen

in hun index. Ook zijn de objecten nu via een API opvraagbaar en kunnen zij geleverd worden in bijvoorbeeld JSON of RDF, formaten die ook buiten de geo-wereld gebruikt kunnen worden.

LD-Proxy is ongetwijfeld het meest tastbare resultaat van het #geo4web testbed, maar misschien nog wel belangrijker is de set van aanbevelingen die het testbed heeft opgeleverd. Deze 'Lessons Learned' beschrijven onder meer hoe je geodata ook voor leken begrijpelijk kunt maken en hoe je ervoor kunt zorgen dat de tijd die het een ontwikkelaar kost om een data API te doorgronden, zo kort mogelijk is. Ook is beschreven hoe je met webstandaarden de semantiek in geodata kunt vastleggen zodanig dat zij ook door zoekmachines begrepen wordt. Verder is uitgezocht hoe je geodata in RDF kunt beschrijven zodat het onderdeel wordt van het semantisch web en heeft het testbed praktische tips opgeleverd. Bijvoorbeeld over het reduceren van het volume van geodata.

De Lessons Learned vormen ook input voor een internationale best practice voor 'Spatial Data on the Web', die de standaardorganisaties voor respectievelijk webstandaarden en geostandaarden (W3C en het OGC) gezamenlijk aan het schrijven zijn. Daarnaast worden de Lessons Learned ook in de praktijk gebruikt, zoals bij recente Kadaster-datapublicaties.

Experiences on the PLDN 2013 - 2017

My first contact with the Platform Linked Data Netherland, PLDN, was an email by Erwin Folmer getting in touch with a request to re-use parts of the Linked Open Data - The Essentials publication in April 2013 that I could answer positively, as the whole book was published under an open license (CC-By 3.0 Austria). So parts of the LOD book were reused for a LOD publication of PLDN published at the PLDN conference in July 2013.

In the course of the conversation with Erwin he told me more about the PLDN: about ideas, objectives and activities and he furthermore invited me to an PLDN event in January 2014 at the VU in Amsterdam to present the LOD Pilot Austria as well as to participate in a discussion and hands-on session on Linked Data Tools as being a team member of the LOD2 project.

This event was a starting point to learn to know this very vital Dutch Linked Data community and for staying in touch for the following years! Meeting Erwin in person and discussing about our enthusiasm about the potentials of Linked Data, discussing with Frank van Harmelen (VU) the potential of LOD as a national infrastructure and - beside others - learning to know Lieke Verhelst (Linked Data Factory) and Jan Voskuil (Taxonic), who both became partners of our product PoolParty Semantic Suite.

This started a series of talks I gave at PLDN meetings. After Amsterdam I could speak in June 2014 with Christian Dirschl about using PoolParty Semantic Suite

at Wolters Kluwer and in September 2015 together with Eelco Kruizinger about Linked Data based services for CTCN.

And at any time at such meetings, in calls and other conversation with the PLDN, I was excited about so much enthusiasm and hands-on mentality to experiment and apply Linked (Open) Data principles and technologies by an audience from industry, academia and public administration.

We also started quickly a cooperation between PLDN and SEMANTiCS conference series. Erwin joined the organising committee and PLDN promoted the European Linked Data Contest 2015 across the Netherlands. Both prizes went to the Netherlands: NXP Semiconductors and Open PHACTS were awarded at SEMANTiCS2015 in Vienna - what an acknowledgement for PLDN.

And finally SEMANTiCS comes to the Netherlands. Taking place 11-14.9. 2017 in Amsterdam, co-organised by the VU, TNO, the Kadaster and in co-operation with PLDN. I really hope to meet the PLDN community there again and that this can be a starting point for linking data activities between the Netherlands and Austria for the next 4 years.

Martin Kaltenböck

Managing Partner & CFO at
Semantic Web Company (www.semantic-web.at)

Kadaster dataplatform: Linked Data & API's

Het Kadaster ontwikkelt een dataplatform waarmee bestaande datasets geautomatiseerd omgezet worden naar Linked Data en van waaruit de data beschikbaar gesteld wordt via meerdere kanalen. Het Kadaster bouwt het dataplatform omdat hiermee datasets op een laagdrempelige manier aan zoveel mogelijk partijen beschikbaar gesteld kunnen worden. Door het delen van gegevens en functies wordt innovatie gestimuleerd. Het ondersteunt op die manier ook het beleid van de Nederlandse overheid ten aanzien van open data.

Momenteel zet het Kadaster daarbij in op het ontsluiten van data via de REST API's en SPARQL-koppelvlakken. Het dataplatform wordt ontwikkeld om zowel bestaande datasets van het Kadaster te ontsluiten, alsook datasets van anderen. Hierbij ligt de focus niet alleen op het ontsluiten van individuele datasets, maar ook op het ontsluiten van links tussen datasets zodat eenvoudig data uit meerdere datasets kan worden gecombineerd. Op dit moment worden de Basisregistraties Adressen en Gebouwen (BAG), de Digitale Kadastrale Kaart (onderdeel van de Basisregistratie Kadaster) en de TOP10NL (onderdeel van de Basisregistratie Topografie) aangeboden. Daarnaast worden onder andere het Nationaal Wegenbestand - Wegen (bronhouder: Rijkswaterstaat) en de Vergunningen Eindhoven (bronhouder:

gemeente Eindhoven) aangeboden. Een volledig overzicht van de beschikbare datasets is te vinden op de website van het PDOK-dataplatform.

Voor het Kadaster zijn deze nieuwe koppelvlakken aanvullend ten opzichte van de koppelvlakken die momenteel in gebruik zijn. Met deze nieuwe koppelvlakken wil het Kadaster zijn datasets voor een breder publiek toegankelijk maken en het eenvoudiger maken om data direct van de bron te betrekken. Hiermee is de gebruiker verzekerd van betrouwbare en actuele data.

Naast het beschikbaar stellen van de datasets, vindt het Kadaster het van essentieel belang dat de metadata van de beschikbare datasets, van goede kwaliteit is. In deze metadata worden zaken als de kwaliteit en de herkomst van de data opgenomen en wordt verwezen naar de definities van de data.

In de kern focust het dataplatform op het beschikbaar stellen van datasets en de bijbehorende metadata, maar het Kadaster vindt ook dat gebruikers eenvoudig toegang moeten hebben tot het model, behorende bij die data. Daarom biedt het Kadaster een aantal datavisualisaties aan via het dataplatform, waarmee het relatief eenvoudig is om door het model en de bijbehorende data te browsen om zo de rijkheid van de datasets te kunnen verkennen.

Standardisation is a community effort

One of the endless discussions among W3C staff is around what we do, what we should have done in the past, and what we should do in future to support the development and maintenance of vocabularies.

Whatever your technology, data interoperability can only be achieved if different systems use the same terms, or, at least, terms that can be programmatically mapped. The W3C process is good for creating stable vocabularies like SKOS and the RDF Data Cube, but what about vocabularies that are continually evolving? The stellar example of that is schema.org. Should W3C adapt its tooling and processes to match? What's the right balance between stability and agility, integrity and practical flexibility, centralisation and distribution?

To make progress, I need a hook, that is, an example of a vocabulary that has a significant community of interest but that is in need of maintenance. Step forward DCAT, the Data Catalogue Vocabulary, and, step forward the VRE4EIC project. A Virtual Research Environment (VRE) offers visualisation and manipulation services across multiple datasets from multiple sources. The project has looked at dozens of potential data sources and it's clear that there is no single dataset description vocabulary that dominates, nor is there likely to be. That's the Web.

At the end of November, I made my sixth and final trip of 2016 to the Netherlands. This time to run the Smart Descriptions & Smarter Vocabularies (SDSVoc) workshop that discussed these issues. The final report from that event has been published around the end of January. As well as looking at some of the many dataset description vocabularies in use, and the specific need to update and improve DCAT, we discussed the idea of content negotiation by profile. That is, that you can request data encoded in JSON or RDF or XML etc. but we want clients to be able to request data in, say, RDF using DCAT and for servers to support multiple alternative responses. Datasets with spatiotemporal aspects need specific metadata fields; tooling and search are important too of course. There's a lot to think about.

Subject to the usual W3C member-driven process, I expect to be beginning a new Working Group around April-May that will pick up on these ideas. DCAT is the poster-child, but the issues are much broader. As ever, standardisation is a community effort - and you'll be welcome to join in.

Phil Archer

Data Strategist at W3C (www.w3.org)

LOD Laundromat wast Linked Data schoon



Het gebruik van Linked Data kost normaal gesproken veel tijd en vraagt de nodige inspanning. Linked Data is in de praktijk lastig vindbaar, van suboptimale kwaliteit, en niet eenvoudig opvraagbaar. Om deze problemen op te lossen is een 'wasserette' voor Linked Data gebouwd: de LOD Laundromat (1ste prijs Beste Linked Data Toepassing 2015; 3de prijs European Linked Data Awards 2015).

De LOD Laundromat zoekt op het internet naar Linked Open Data-bronnen. De meeste bronnen bevatten 'vlekken', zoals syntaxfouten, dubbele voorkomens of incorrecte waardes. De Laundromat wast deze vlekken schoon. Wat overblijft is standaardconforme Linked Data, die je zo van de 'waslijn' kunt plukken.

Om dit te realiseren maakt de LOD Laundromat gebruik van de laatste technologische ontwikkelingen op het gebied van data crawling, streaming, parsing, opslag en querying. De Laundromat publiceert alle gegevens in open en gestandaardiseerde formaten die voor lokaal gebruik kunnen worden gedownload. Daarnaast biedt zij ook webservices aan, waarmee de data direct online bevroegd kan worden.

Op dit moment bevat de LOD Laundromat honderdduizenden datasets en tientallen miljarden triples. De Laundromat wordt door duizenden ontwikkelaars op het gebied van Linked Data gebruikt. Maandelijks worden miljoenen documenten gedownload en tientallen miljoenen queries afgehandeld.

De LOD Laundromat neemt niet alleen werk uit handen bij het zoeken en vinden van Linked Data, maar helpt ontwikkelaars ook om hun data zelf beter beschikbaar te stellen. Zo worden de niet-standaardconforme eigenschappen van iedere dataset in metadata opgeschreven en is het mogelijk om schoongewassen tussenversies van de data voor eigen gebruik te downloaden.

LOD Laundromat is een ideaal raamwerk om grootschalige Linked Data-toepassingen op te bouwen. Eén zo'n toepassing is LOD Lab, een methode om grootschalige evaluaties over de gehele LOD Cloud uit te voeren (Best Paper Award ISWC 2015). Een andere toepassing is LOD Search (2de prijs European Linked Data Awards 2016). LOD Search is een grootschalige semantische zoekmachine die gebruikers in staat stelt om op eenvoudige wijze naar Linked Data-bronnen te zoeken. De verwachting is dat in de nabije toekomst meer grootschalige toepassingen op basis van de LOD Laundromat gebouwd zullen worden.

NXP Enterprise Datahub - NXP & SEMAKU

NXP biedt meer dan twintigduizend producten aan met een breed scala aan functies en technologieën. Informatie over elk van deze producten wordt in diverse interne systemen opgeslagen en beheerd. Hoe zorg je er dan voor dat informatie in alle verschillende systemen wordt bijgewerkt? NXP lost het op met Linked Data.

Informatie over producten wordt op verschillende manieren en via meerdere kanalen intern en naar de klant gecommuniceerd: bijvoorbeeld via internet, mobiel en print. Hierdoor raken data verspreid en wordt informatie gedupliceerd in meerdere applicaties en databases. Dit leidt ertoe dat mensen niet kunnen vinden wat ze zoeken of tegenstrijdige informatie vinden. De Enterprise Data Hub, een data-integratietechnologie, zorgt ervoor dat data uit verschillende databases geïntegreerd kunnen worden via Linked Dataprincipes en standaard webtechnologieën. Dankzij 'shared identifiers' kunnen gebruikers bijvoorbeeld data over de elektrische eigenschappen van een product samenvoegen, met gegevens over de levenscyclus van het product.

De NXP Enterprise Datahub haalt data op uit bron-systemen, transformeert deze naar RDF en laadt ze vervolgens in de RDF Graph Database, van waaruit ze conform Linked Data-principes worden gepubliceerd.

De brondata bestaan uit diverse XML- en CSV-formaten en kennen verschillende updatefrequenties. In het eenvoudigste geval kunnen data automatisch worden opgehaald. In andere gevallen worden berichten verspreid via de Enterprise Service Bus (ESB) en wordt de transformatie in het kanaal geïmplementeerd. Van XML-bronnen wordt de data naar RDF (RDF/XML of Trix) getransformeerd, gebruikmakend van XSLT of XQuery. Voor CSV-bronnen wordt de CSV naar een Java-object gebracht, vervolgens naar XML, gevolgd door XSLT.

Om te voorkomen dat interne data per ongeluk publiek toegankelijk worden, zijn strikte filtermechanismen opgezet. Deze filteren de openbare subsets uit interne data. Dit gebeurt met behulp van SPARQL 1.1 Federated Query, samen met SPARQL 1.1 Update, waar de 'business rules' worden gevangen in SPARQLlanguage. Waar mogelijk heeft NXP gebruikgemaakt van bestaande standaard vocabulaires als Dublin Core, SKOS, FOAF en Schema.org. Omdat er een aantal voor de sector benodigde termen en concepten ontbraken, is ook een NXP-vocabulaire ontwikkeld met mappings (zoals `rdfs:subClassOf`) naar externe vocabulaires. Zo zijn er voldoende mogelijkheden om de data te beschrijven en kunnen externe gebruikers de betekenis van begrippen gemakkelijker nagaan.

github.com/NXPdata/NXPdata/wiki + data.nxp.com

Privacy en Linked Open Data: aan tafel!

Het is meer dan tien jaar geleden dat de term 'Linked Data' een buzzwoord werd. Als je erop zoekt, komen er vooral mooie technische termen naar boven, maar het boeiendst is het effect: in zijn meest zuivere vorm leidt Linked Data tot een onbegrensde en ongekende verbinding van online gepubliceerde data en daarmee tot contextualisering van, goed beschouwd, 'alles'. De baten hiervan zien en ervaren we iedere dag.

Dit 'linken' heeft ook andere kanten. Doordat data met elkaar verbonden zijn, wordt het niet alleen veel gemakkelijker data over bepaalde personen te vinden, ook de populatie van identificeerbare personen neemt exponentieel toe. En als je met behulp van elektronische informatie personen identificeert of kan identificeren, dan moet je je aan bepaalde wettelijke regels houden. Anders gezegd: Linked Data heeft privacybeschermingsverplichtingen als bijvangst.

Tot voor kort was de wereld van privacybeschermingsregels vooral het domein van juristen en dan ook nog eens van een erg kleine groep. Punt daarbij is dat de regelgeving over verwerking van persoonsgegevens niet alleen abstract en ingewikkeld is, maar ook nog eens zeer casuïstisch: uiteindelijk gaat het om de concrete omstandigheden van het geval.

Dit is geen goed nieuws voor Linked Data, die veelal bestaan uit of gebaseerd zijn op publieke domeindata, overheidsgegevens in het bijzonder. Immers, dit maakt houders van deze data - denk daarbij vooral aan grote verzamelingen geogerelateerde registraties - enigszins

kopschuw: de kans dat hun data tot identificatie leidt, wordt groter. Omdat zij als 'verantwoordelijke' gelden, is dat niet bepaald een stimulans voor het onverminderd doorzetten van het ingezette opendatabeleid.

Kortom, het is hoogste tijd dat zij die erover gaan, en dit kunnen oplossen, dit ook gaan doen. Dat begint met onderkennen dat er een spanning zit tussen Linked Data en privacybescherming en dat het niet doorhakken van knopen, leidt tot onderbenutting van potentieel van deze data. Vervolgens vereist dit het afwegen van belangen - het belang van privacybescherming versus het sociaaleconomisch belang bij het benutten van het potentieel - het maken van principiële keuzes - wat is toegestaan en wat niet? - en het verzorgen van de toepassing en naleving daarvan.

Kijken we de (grotendeels nog lege) tafel rond, dan zouden daar zeker een aantal ministeries moeten aanschuiven: BZK, als beleidsmaker op het gebied van open data; I&M, als beleidsverantwoordelijke en houder van grote verzamelingen (geo-)data; EZ, als hoeder van de economische groei; J&V, als verantwoordelijke voor de privacyregelgeving en de Autoriteit Persoonsgegevens als de waakhond voor de naleving. Een goede kok, een lekker menu, een fles goede wijn en een datum waarop ze allemaal kunnen. Dat lijkt mij het meest simpele recept. Hoe moeilijk kan het zijn?

Marc de Vries

Adviseur bij Geonovum (www.geonovum.nl)

Hackathon met Linked Open Erfgoed Data

Hoe komen we tot nieuwe en aansprekende voorbeelden om de kracht van Linked Data voor de erfgoedsector te verbeelden? Al een tijdje werken veel culturele instellingen, musea en bibliotheken aan het beschikbaar stellen van hun collecties als Linked Data. Maar dé toepassing die laat zien wat je daar als gebruiker mee kunt, moet nog worden gemaakt.

Ideeën zijn er genoeg, maar allemaal hebben we hetzelfde probleem: de tijd ontbreekt om er eens even rustig aan te kunnen werken. 'We zouden ons eigenlijk eens 24 uur moeten opsluiten', roepen we soms. Maar wacht eens: waarom doen we dat dan niet gewoon? En dan gelijk met een groot aantal anderen. En dan niet op kantoor, maar op een plek waar we echt niet gestoord kunnen worden en geen afleiding hebben.

Voor we het wisten, hadden we de ideale plek gevonden voor de Hack-a-LOD 2016: het voormalige gevangeniscomplex De Vrije Wolf in Utrecht. Op die locatie kregen dertig hackers op vrijdag 11 november de gelegenheid en de tijd om te werken aan innovatieve toepassingen. Na 24 uur 'gevangenschap' werden zij vrijgelaten en mochten ze hun applicaties presenteren aan de circa honderdtwintig aanwezigen die in de voormalige kerkzaal van de gevangenis benieuwd waren naar de resultaten. Die waren verrassend.

Programmeurs van Picturae bouwden de game 'Pak je monumentje!', die gebruikmaakt van de SPARQL-toegang tot DBPedia en de API van Rijksmonumenten. Ontwikkelaars van het Kadaster maakten een 3D-weergave van gebouwen op een kaart, die gekoppeld was aan de monumentendataset van RCE. En ook de winnaars van de publieksprijs, Marieke van Erp en Richard Zijdeman van Team CLARIAH, gebruikten de adressen van gebouwen en zochten automatisch naar informatie over die adressen in de advertenties van de digitale kranten in Delpher.nl. Op die manier werd in hun project 'If Buildings could talk' het levensverhaal van een gebouw verteld. Met de winnende applicatie van de juryprijs kun je ook in gesprek gaan met ons nationaal erfgoed, maar dan net even anders. Hay Kranen bouwde de Erfgoed Bot, een chatbot waarmee je in Facebook persoonlijke informatie ontvangt van schilderijen en monumenten uit Wikidata. Als prijs wordt de Erfgoed Bot door de KB verder ontwikkeld.

De Hack-a-LOD 2016 was georganiseerd door de Koninklijke Bibliotheek (KB) en het Netwerk Digital Erfgoed (NDE) en bestond uit twee tracks. Voor het open track mochten deelnemers eigen ideeën voor applicaties uitwerken. Tijdens de gesloten track werkten medewerkers van erfgoedinstellingen en softwarebedrijven samen aan plannen om erfgoedcollecties met elkaar te verbinden.

Open PHACTS

Voor farmaceutisch onderzoek zijn veel en diverse informatiebronnen beschikbaar. Met onder andere chemische, farmacologische, (pre-)klinische en genetische data, die zowel publiek toegankelijk als intern kunnen zijn. Vanwege de complexiteit van biologie, farmacologie en toxicologie is het belangrijk dat onderzoekers alle databronnen in diverse biomedische categorieën simultaan kunnen analyseren en doorzoeken. Vóór het Open PHACTS-project was dit niet mogelijk of extreem tijdrovend, maar dankzij semantische webstandaarden en Linked Data, is dat nu anders.

Open PHACTS biedt een semantisch platform waarin Linked Data en semantische webstandaarden worden toegepast om farmaceutisch onderzoek te ondersteunen. Dankzij de aandacht die deze publiek-private samenwerking heeft besteed aan ontologieën en woordenlijsten, gegevenslicenties en auteursrecht, is het nu mogelijk om een grote hoeveelheid biomedische open data te doorzoeken. Er bestaan letterlijk duizenden publieke biomedische databronnen, dus een zorgvuldige selectie was belangrijk om de meest nuttige daarvan in het systeem op te nemen. Deze selectie werd gemaakt aan de hand van relevante farmaceutische vragen die door een brede groep van industriële en academische onderzoekers werden geprioriteerd. Op dit moment bevat het systeem elf verschillende databronnen, meer dan vier miljard triples en de inhoud wordt continu uitgebreid

met nieuwe data. Een van de laatste toegevoegde databronnen is SureChEMBL, waarin specifieke moleculen uit meer dan veertien miljoen patenten automatisch geëxtraheerd worden. Het Open PHACTS-project heeft hier met automatische text mining nog eens miljoenen semantische links tussen moleculen en genen en ziektes gedefinieerd.

Naast de data zelf zijn er ook tientallen applicaties ontwikkeld die gebruikmaken van de publiek toegankelijke Open PHACTS API. Een goed voorbeeld hiervan is een set met computationele workflows in de KNIME-software die een veel voorkomende vraag in farmaceutisch onderzoek beantwoorden: aan welke eiwitten binden de moleculen die in mijn fenotypisch experiment actief zijn? Met deze workflows kunnen onderzoekers binnen minuten de nuttige informatie uit alle beschikbare bronnen halen, iets wat daarvoor gemakkelijk dagen kon duren.

Open PHACTS heeft aangetoond dat het mogelijk is om diverse en grote biomedische databronnen met semantische technologie te converteren tot Linked Data. Het is van grote invloed geweest op de vele huidige activiteiten om biomedische data FAIR te maken (FAIR: Findable, Accessible, Interoperable, Reusable). Big data is belangrijk in farmaceutisch onderzoek en om de maximale waarde eruit te halen, zijn Linked Data-concepten essentieel.

Doen we het goed of doen we het goed?

Bij een jubileum hoort het om even stil te staan bij de stand der dingen: wat hebben we gerealiseerd, wat is goed, wat is slecht. Vandaar een aantal 'persoonlijke' op ervaring gebaseerde bedenkingen en kreten om de discussie te stofferen.

Semantiek: moeten we niet van weten?

Iedereen heeft de mond vol van semantiek en modellen, maar in realiteit zie ik daar niet zoveel (goeds) gebeuren. Enkele symptomen. Bestaande standaard vocabularies/ontologieën worden in hun totaliteit zelden of nooit gebruikt. Vocabularies staan maandenlang gepubliceerd terwijl ze niet parseren. Bestaande vocabularia bevatten fouten die traag worden ontdekt. En er is de mismatch tussen de traditionele manier van modelleren (UML,...), waarbij properties en relaties bij de klasse horen en de wereld van description logics, waaruit OWL/RDFS voortkomen.
Conclusie: Linked Data te moeilijk voor modelleerders?

Dereference-able URI's: te moeilijk?

Een van de kenmerken van Linked Data is dat als je de identifier van een ding op het web http-gewijs opvraagt, je een document terugkrijgt dat het ding beschrijft. Met andere woorden dat je subjectpagina's aanbiedt, zeker als het gaat om referentielijsten die in allerlei applicaties en domeinen herbruikt kunnen worden. We wachten al jaren op deze lijsten met dereference-able URI's...en we blijven maar wachten...
Conclusie: Linked Data te moeilijk voor data publishers?

Niet bedoeld voor mensen?

Als er dan al subjectpagina's staan, geven die zelden een 'wow' gevoel. Ze hebben de vorm van een 'fiche', die dan vol staat met URI's.
Zie: <http://vocab.getty.edu/aat/300198841>.
Conclusie: Heeft Linked Data geen toegevoegde waarde voor de 'gewone' webgebruiker?

Te vreemd voor de doorsnee developer?

We hebben in onze wereld een zeer krachtige query taal (SPARQL), maar voor de gemiddelde developer, die een REST API verwacht, blijkt het gebruik ervan toch iets erg exotisch.
Conclusie: er is een mismatch met de verwachtingen van de developer.

De rest van de wereld: 'do they care'?

Data is het nieuwe goud en de wereld van analytics en machine-learning gaan hard. Van de vele tools om OSEMN (Mason & Wiggins, 2010) met data om te gaan hebben wij er slechts twee gevonden die iets kunnen doen met Linked Data.
Conclusie: We moeten beter doen. Op naar de volgende vijf jaar!

Paul Hermans

Oprichter van ProXML (www.proxml.be)

P.S. De vele positieve ervaringen met het gebruik van semantische webtechnologie voor het oplossen van data integratie behoeften pasten niet meer in de mij toegemeten ruimte ©

CLARIAH

Disciplines in the Arts and Humanities, such as philosophy, history, literature, language and media studies would potentially benefit a lot from each other's techniques and sources, but currently such cross-fertilization is hampered by the lack of an infrastructure to share domain specific knowledge. CLARIAH is an NWO roadmap facility, that aims to enhance interdisciplinary research via shared use of methods and sources.

CLARIAH consists of five working packages. WP1 is responsible for dissemination of the results. Three working packages focus on a specific Arts and Humanities discipline and type of data. WP3 focuses on linguistics and derives algorithms to extract information from textual sources, for example to enhance corpora. Its results are compliant with the CLARIN.EU infrastructure. WP4 works with tabular (structured) data as used in social and economic history. It transposes key datasets to Linked Open Data in order to provide contextual information social and economic processes, regions and periods. WP5 represents media studies and uses audio and video sources. Currently, the tools in this domain are intertwined with a specific dataset. WP5 is building a media suite that allows for the generic use of those tools. Results of both WP4 and WP5 are Dutch contributions to DARIAH.EU. Finally, WP2 is responsible for the overarching infrastructure providing access to

the tools and data from the domain specific packages. Moreover, WP2 is responsible for generically shared concepts, such as persons and places.

Linked (Open) Data is crucial to the CLARIAH project as it provides the common 'language' through which all of the packages communicate. Where in the past domain specific rules on how to code variables or call concepts hindered exchange of information, the explication of variables and values in ontologies and vocabularies now allows for easy exchange, while domain specific terms can be preserved.

Another major advantage to the use of Linked Data in the CLARIAH project is that Linked Data is available to a large part of the world population (basically anyone with internet), allowing for cooperation on a global scale. For example, a sub-project of WP2 locates and transposes Dutch locations to Linked Data using terms for the start and end-date of municipalities from the Pleiades vocabulary, originally used for Ancient societies. Around it an international community has evolved in the Pelagios project which aims to connect geographic information across the globe. This way Linked Data cuts across time, space and discipline and is not limited to social networks for awareness. Linked Data thus allows CLARIAH to expand beyond the national borders and to foster interdisciplinary use of tools and data.

Linked Partners

Bij de ontwikkeling van Linked Data in Nederland hebben in de afgelopen jaren een groot aantal partijen een voorname rol gespeeld, waaronder...



...en al die honderden mensen die, in welke vorm dan ook, (vrijwillig) hebben bijgedragen aan de ontwikkeling en het succes van Linked Data in Nederland.

Nawoord

Trots! Dat is gevoel dat bij mij overheerst na vijf jaar in de wereld van Linked Data actief te zijn. Marcel Reuvers van Geonovum benaderde mij vijf jaar geleden op de receptie na de (succesvolle) verdediging van mijn proefschrift. “Nu heb je tijd voor wat anders”, was zijn simpele constatering. Dat was zeker waar, maar op dat moment had ik nog geen idee dat Linked Data mijn werk voor de komende jaren zou gaan bepalen.

Er is in de afgelopen vijf jaar veel veranderd. Waar in de beginjaren de focus lag op verkennen, overtuigen van nut en noodzaak, mensen kennis laten maken met Linked Data, daar ligt de focus nu op het ‘doen’. Daarbij spelen vragen als: Hoe krijgen we het werkend? Hoe kunnen we de eerste stappen zetten? Hoe kunnen we links gaan aanbrenge? Fantastisch om te zien dat er steeds meer Linked Open Data gepubliceerd wordt, er steeds meer succesvolle cases zijn en er steeds meer erkenning plaatsvindt in de vorm van prijzen. Daarin heeft het Platform Linked Data Nederland (PLDN) een belangrijke rol in gespeeld.

Wat is de kern van het Platform Linked Data Nederland? De open community. Iedereen die in deze community deelneemt (bij de sessies, online, soms in discussies, soms als presenter) draagt bij aan het succes van deze open community. Ik ben trots op al deze vrijwilligers. Een aantal springt daar bovenuit Met het risico dat ik anderen tekort doe, wil ik Pieter van Everdingen, Gerard Persoon en Richard Nagelmaeker in het bijzonder bedanken en benoemen. Zij betekenen erg veel voor PLDN!

Trots ben ik ook op Geonovum. Zij hebben het platform geïnitieerd en zijn jarenlang een warm clubhuis voor PLDN geweest. Trots ben ik ook op de partners, die eigenlijk zonder echt grote directe tegenprestatie een financiële bijdrage hebben geleverd, waardoor wij veel nadruk op open en toegankelijk konden leggen. Verder ben ik trots op al die locaties die ons ter beschikking zijn of worden gesteld en op al die sprekers die hebben meegewerkt.

Ook met mijn Kadasterpet op ben ik trots. Extreem trots dat we de eerste basisregistraties als Linked Data hebben kunnen ontsluiten en nog trotser dat we het met een ‘dataplatform’ doen wat ook voor andere datasets (ook buiten Kadaster) inzetbaar is. Dit is gelukt met dank aan het Platform Linked Data Nederland. Voor ons is dat de rijkgevlude vijver aan kennis en mensen. Als Kadaster hebben wij in de afgelopen jaren ontzettend veel gehaald bij PLDN en ons streven is om aan de community net zoveel terug te geven.

Tot slot. Trots, maar ook dankbaar ben ik dat ik een rol heb mogen spelen in deze prachtige community rondom Linked Data. Ik hoop dat ik nog lang trots kan en mag blijven!

Erwin Folmer
Geonovum, Kadaster, Universiteit Twente en trekker Platform Linked Data Nederland



Colofon

Contact PLDN

Erwin Folmer (Erwin.Folmer@kadaster.nl)
Pieter van Everdingen (info@openinc.nl)

Redactie

Platform Linked Data Nederland
Onelevel Tekstbureau, Den Haag

Vormgeving

Remwerk, Amersfoort

Illustratie voorpagina

Elco van Staveren

Drukwerk

VdR druk & print, Nijkerk

Oplage

1000 exemplaren

Maart 2017

ISBN 978-90-365-4312-5



Blijf op de hoogte!

Website

www.platformlinkeddata.nl

LinkedIn-groep LOD Nederland

www.linkedin.com/groups/4662786

Twitter

[#linkeddatanl](https://twitter.com/linkeddatanl)

Nieuwsbrief

www.pilod.nl/wiki/Nieuwsbrieven



Platform Linked
Data Nederland

PlatformLinkedData.nl