



ALTERRA

WAGENINGEN UR



Gegevens delen via Linked Open Data

Case study over het delen van grondwatergegevens bij de bouw van de A2 tunnel in Maastricht

Alterra-rapport 2306
ISSN 1566-7197

E. Verhelst, J. Houtkamp en Y. van Randen

Gegevens delen via Linked Open Data

Dit onderzoek is uitgevoerd in opdracht van Rijkswaterstaat
Projectcode [5239279]

Gegevens delen via Linked Open Data

Case study over het delen van grondwatergegevens bij de bouw van de A2 tunnel in Maastricht

E. Verhelst, J. Houtkamp en Y. van Randen

Alterra-rapport 2306

Alterra, onderdeel van Wageningen UR
Wageningen, 2012

Referaat

E. Verhelst, J. Houtkamp en Y. van Randen, 2011. *Gegevens delen via Linked Open Data; case study over het delen van grondwatergegevens bij de bouw van de A2 tunnel in Maastricht*. Wageningen, Alterra, Alterra-rapport 2306. 32 blz.; 11 fig.; 18 ref.

In dit rapport is een case study beschreven waarin wordt gekeken naar het delen van gegevens bij de bouw van een groot infrastructuur project (Maastricht A2). De manier van gegevensdelen die is uitgewerkt wordt Linked Open Data genoemd. Bij Linked Open Data zijn de brongegevens toegankelijk en aan elkaar verbonden met web links. Omdat de gegevens toegankelijk zijn via de bron kan elke partij zijn eigen overzichten maken en applicaties en visualisaties bouwen. De gekoppelde data vormen een wereldwijde netwerkstructuur waaraan makkelijk nieuwe gegevens kunnen worden toegevoegd. Zo wordt de samenhang tussen gegevens beter zichtbaar.

Trefwoorden: Open Data, Linked Data, gegevensuitwisseling, datamodel

Foto omslag: <http://www.flickr.com/photos/25878086@N06/2731710806/>

ISSN 1566-7197

Dit rapport is gratis te downloaden van www.alterra.wur.nl (ga naar 'Alterra-rapporten'). Alterra Wageningen UR verstrekt geen gedrukte exemplaren van rapporten. Gedrukte exemplaren zijn verkrijgbaar via een externe leverancier. Kijk hiervoor op www.rapportbestellen.nl.

© 2012 Alterra (instituut binnen de rechtspersoon Stichting Dienst Landbouwkundig Onderzoek)
Postbus 47; 6700 AA Wageningen; info.alterra@wur.nl

- Overname, verveelvoudiging of openbaarmaking van deze uitgave is toegestaan mits met duidelijke bronvermelding.
- Overname, verveelvoudiging of openbaarmaking is niet toegestaan voor commerciële doeleinden en/of geldelijk gewin.
- Overname, verveelvoudiging of openbaarmaking is niet toegestaan voor die gedeelten van deze uitgave waarvan duidelijk is dat de auteursrechten liggen bij derden en/of zijn voorbehouden.



Naamsvermelding-Niet-commercieel-Gelijk delen 3.0 Nederland
Meer informatie over deze Creative Commons licentie op:
<http://creativecommons.org/licenses/by-nc-sa/3.0/nl/>

Alterra aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Alterra-rapport 2306

Wageningen, april 2012

Inhoud

Samenvatting	7
Verklarende woordenlijst	9
1 Inleiding	11
1.1 Leeswijzer	12
2 Open Data	13
2.1 De vijf sterren van Open Data	13
3 Linked Open Data	15
4 Case study en demonstrator	21
4.1 Bepaling case study onderwerp	21
4.2 Case study beschrijving	21
4.3 Demonstrator	22
4.3.1 RDF data laag	22
4.3.2 HTML pagina laag	23
4.3.3 User interface laag	24
5 Conclusies en discussie	27
5.1 Meerwaarde van Open Data	27
5.2 Meerwaarde van Linked Open Data	28
5.3 Stappen naar implementatie	28
Referenties	31

Samenvatting

Op 15 september 2011 opende minister Donner het Nederlandse Open Data portaal data.overheid.nl. Dit moment markeerde een belangrijke stap in de uitvoering van een deel van de Digitale Agenda. Deze agenda beschrijft acties die gericht zijn op het zo open mogelijk delen van overheidsgegevens om bedrijvigheid en innovatie te stimuleren.

Om te onderzoeken hoe dit beleid kan uitwerken op de gegevensuitwisseling tussen Rijkswaterstaat (RWS) en andere partijen is een case study gestart. De case study behandelt het delen van gegevens over het grondwaterpeil bij de bouw van de A2 tunnel bij Maastricht via het principe van Linked Open Data (LOD). Om de case study te illustreren is een demonstrator gebouwd, waarbij drie partijen (RWS, aannemer en burger) gebruik maken van dezelfde gelinkte Open Data gegevens.

Open Data biedt veel meer mogelijkheden dan de huidige manieren van gegevensuitwisseling. Dit komt omdat niet alleen de data maar ook de datastructuur open is. Zo kan elke partij zijn eigen overzichten maken, analyses uitvoeren, koppelingen met andere bestanden maken en applicaties en visualisaties bouwen. De verwachting is dat dit onbekende nieuwe inzichten over de gegevens oplevert. Open Data kan de dataleverancier ook bevrijden van de verwachting om een gebruikersinterface op de gepubliceerde gegevens te bieden. Afnemers van ruwe Open Data kunnen dit voortaan zelf.

Linked Open Data is een specifieke vorm van Open Data. LOD is opgeslagen via een netwerk datamodel op het internet. Een netwerk datamodel kent geen begin en geen eind en is daardoor flexibel genoeg om gegevensstructuren toe te voegen of te wijzigen. Omdat steeds meer partijen hun gegevens via LOD publiceren ontwikkelt het internet zich tot een grote netwerk database, waaraan iedereen zijn deel kan bijdragen. Als RWS data gaat publiceren in het LOD-formaat maken de gegevens van RWS onderdeel uit van dit internet data netwerk. Dit heeft als gevolg dat de RWS-gegevens kunnen worden gebruikt door organisaties die tot nu toe niet zijn bereikt. De gegevens kunnen ook binnen ongebruikelijke en onverwachte contexten worden benut. Het omgekeerde is ook het geval: RWS krijgt toegang tot gegevens die voor de bedrijfsvoering eerder niet beschikbaar waren. Dit bevordert innovatie.

Wij danken de volgende mensen voor hun waardevolle bijdrage in dit onderzoek: Wim de Haas, Roland Drijver, Alphons Engelen van Rijkswaterstaat, Detlev van Loenhout van Ballast Nedam en Huibert-Jan Lekkerkerk van het Informatiehuis Water.

Verklarende woordenlijst

Open Standaarden	Afspraken in de vorm van specificaties die betrekking hebben op het standaardiseren van de gegevensuitwisseling tussen partijen. Dit is een technische aanduiding.
Open Data	Gegevens die beschikbaar zijn in een zo herbruikbaar mogelijk gegevensformaat en met zo min mogelijk juridische beperkingen. Dit is een technische en juridische aanduiding.
Open Source	Een open source product is een product waarvan de bron waarmee het is gemaakt zonder juridische beperkingen beschikbaar is. Dit is een juridische aanduiding.
Linked Open Data	Gedeelte van het internet dat het netwerk van data bevat.
Semantisch Web	Gedeelte van het internet dat het datamodel van het netwerk van data bevat.
API	Application Programming Interface. Door de leverancier van een specifieke dienst vrij aangeboden code om gegevens van die dienst te kunnen opvragen.
Web service	Mechanisme van servicegerichte elektronische gegevensuitwisseling waarin een aanbieder (service provider), vragende partij (service requestor) en een tussenpartij (service broker) een rol spelen. Hoe de uitwisseling plaatsvindt is vastgelegd in de service-beschrijving.

1 Inleiding

In 2003 werd in de Europese Directive on the Re-Use of Public Sector Information (EC, 2003) (PSI Directive) vastgelegd dat hergebruik van publieke informatie moet worden bevorderd. Dit was een belangrijke richtlijn die de basis heeft gelegd voor de gegevensparagraaf van de Europese Digitale Agenda (EC, 2010). Nederland volgde met een eigen Digitale Agenda.

In de Nederlandse Digitale Agenda is een paragraaf opgenomen over 'Groeï en Innovatie met 'Open Data' als grondstof' (Min. van EL&I, 2011). Een citaat hieruit illustreert het doel van de openstelling van overheidsgegevens: 'Binnen de overheid zijn veel data beschikbaar die relevant kunnen zijn voor ondernemers om te worden hergebruikt. Verschillende studies laten zien dat de economische waarde van de innovatie en bedrijvigheid die ontstaat op basis van Open Data groot is (Broek, Huijboom et al., 2011). Met Open Data als grondstof kunnen nieuwe toepassingen en diensten ontwikkeld en vermarkt worden. Overheidsdata zijn vooralsnog echter vaak niet 'open' omdat ze bijvoorbeeld niet vindbaar zijn of niet door computers kunnen worden gelezen. Hierdoor kunnen ondernemers er niet direct mee aan de slag. Het kabinet wil ondernemerschap stimuleren door zoveel mogelijk van haar eigen data via open standaarden beschikbaar te stellen voor doorontwikkeling door bedrijven. Het spreekt vanzelf dat veiligheidsgerelateerde, bedrijfsvertrouwelijke en tot personen te herleiden data daar buiten valt.'

Op 15 september opende minister Donner het Nederlandse Open Data portaal (2011). Hij sprak hierbij de intentie uit om dit initiatief verder te ontwikkelen en te komen tot een goed nationaal Open Data beleid waardoor alle overheidsinformatie simpel, snel en gratis of tegen zeer geringe kosten te vinden is. De ministeries van EL&I en I&M volgden snel met de bekendmaking dat hun ministeries dit beleid overnemen. Op de innovatie-estafette (4 oktober) kondigde minister Verhagen aan dat er een Nationale Satellietbank zou komen waarvan de beelden gratis beschikbaar zijn (Rijksoverheid, 2011). Minister Schulz maakte bekend dat op 1 januari 2012 de basisregistratie topografie gratis wordt en uiterlijk vanaf 1 januari 2015 alle gegevens van het ministerie van Infrastructuur en Milieu beschikbaar komen. Hier gaat het voornamelijk over data over water, weer en wegen (Rijksoverheid, 2011). Veel van deze gegevens zijn verzameld en worden beheerd door Rijkswaterstaat.

Het is in het licht van deze ontwikkelingen dat Rijkswaterstaat Alterra de opdracht heeft gegeven om voor beleidsmakers inzichtelijk te maken wat de toegevoegde waarde is van het gebruik van Linked Open Data in de RWS werkpraktijk. Linked Open Data (LOD) is een vorm van Open Data waarbij de data is gekoppeld via web-links (URI's). Dit wordt verder uitgelegd in hoofdstuk 3.

Er is voor gekozen om aan de hand van een case study de werking van LOD te verhelderen. De case study is ontworpen door een schets te maken van een toekomstige situatie die gebaseerd is op de bouw van de rijksweg A2 bij Maastricht. Verschillende mensen uit het werkveld zijn hiervoor geïnterviewd. In de case study zijn fictieve databronnen gebruikt. Met deze fictieve data bronnen is een demonstrator gebouwd die laat zien hoe LOD werkt en wat het effect is op het delen van informatie. Aan de hand van deze demonstrator wordt in dit rapport verder uitgewerkt wat de meerwaarde is van Open Data en Linked Open Data vergeleken met de nu gebruikelijke manier van gegevensdeling tussen partijen.

1.1 Leeswijzer

In hoofdstuk 2 wordt het begrip Open Data kort geïntroduceerd, gevolgd door een verdere uitleg over Linked Open Data in hoofdstuk 3. De beschrijving van de case study en een technische uitleg van de demonstrator zijn opgenomen in hoofdstuk 4.

Het rapport sluit af met een onderdeel conclusies en discussie (hoofdstuk 5). Hierin wordt het antwoord op de onderzoeksvraag gegeven door de Linked Open Data-oplossing te positioneren ten opzichte van huidige werkwijzen en toekomstige ontwikkelingen. Onderdeel hiervan is een beschrijving van de stappen die genomen moeten worden om te komen tot een gegevensuitwisseling op basis van Linked Open Data.

2 Open Data

De Amerikaanse en Engelse overheden zijn als eerste begonnen met het implementeren van een Open Data beleid. Dit deden ze door het lanceren van data portalen. De voornaamste reden hiervoor had te maken met beter overheidsbestuur (Lakhami, 2010; data.gov.uk, 2011). Het ging om transparantie, efficiency, het betrekken van de burger bij het bestuur en het daardoor beter kunnen sturen op reële doelstellingen en innovatie. Belangrijk hierbij was het verminderen van toegang-barrières, zowel technisch als juridisch. De juridische barrières zijn over het algemeen intellectueel eigendom en aansprakelijkheid. De technische barrières zijn toegang en interoperabiliteit.

De Amerikaanse en Engelse overheden openden beiden een portaal¹ waarop de overheidsgegevens in een open formaat werden aangeboden. Juridische barrières werden afgedekt door een 'data policy statement' of 'terms and conditions' op te nemen. Sindsdien zijn er door publieke partijen in die landen een groot aantal toepassingen ('apps') gemaakt met deze gegevens en is er een dialoog ontstaan tussen de burger en de overheden over de kwaliteit en de bruikbaarheid van de datasets.

2.1 De vijf sterren van Open Data

Wat is Open Data? Tim Berners-Lee, de grondlegger van het World Wide Web, bedacht hiervoor een 'punten systeem', de vijf sterren van Open Data (Berners-Lee, 2006-2009).

- ★ Beschikbaar met een open licentie
- ★★ Machine leesbaar
- ★★★ Machine leesbaar zonder merkgebonden formaat
- ★★★★ Gebruik URI's zodat ernaar verwezen kan worden
- ★★★★★ Verwijs zelf ook naar andere data

Figuur 1

De vijf sterren van Open Data.

Het vijf sterren-model is eenvoudig. Je krijgt één ster als je gegevens publiceert zonder juridische beperkingen. Aan de technische openheid worden dan weinig eisen gesteld. Je krijgt er een ster bij als je de gegevens

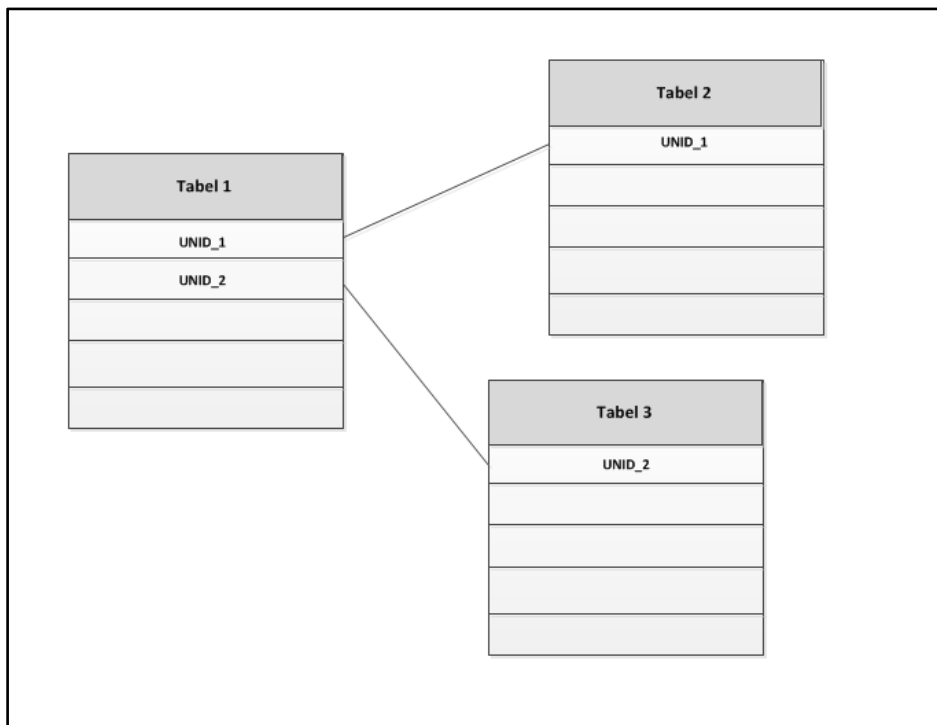
¹ Voor de US is dit: <http://data.gov> en voor de UK: <http://data.gov.uk>

publiceert in een formaat dat een machine kan lezen. Hiermee wordt bedoeld dat de machine de datastructuur kan ontdekken en gebruiken om bewerkingen op de gegevens te doen, zoals analyse, converteren, visualiseren. Voor twee sterren mag het in een merkgebonden formaat worden aangeboden (bijvoorbeeld Excel). Als het in een open uitwisselingsformaat wordt aangeboden (vb: CSV, XML) komt er een ster bij. De sterren vier en vijf zijn alleen te verdienen als het idee van Linked Open Data wordt omarmd. Deze vorm van Open Data wordt uitgelegd in het volgende hoofdstuk.

Door het idee van het toekennen van punten aan de manier van aanbieden van Open Data gaf Berners-Lee richting aan zijn toekomst visie van het web van data. Hoe meer sterren, hoe meer in de richting van zijn visie op het web van data. Dit model vindt veel gevolg door de autoriteit van Berners-Lee, op dit moment de leider van het 'architectenbureau van het internet', het World Wide Web Consortium W3C.

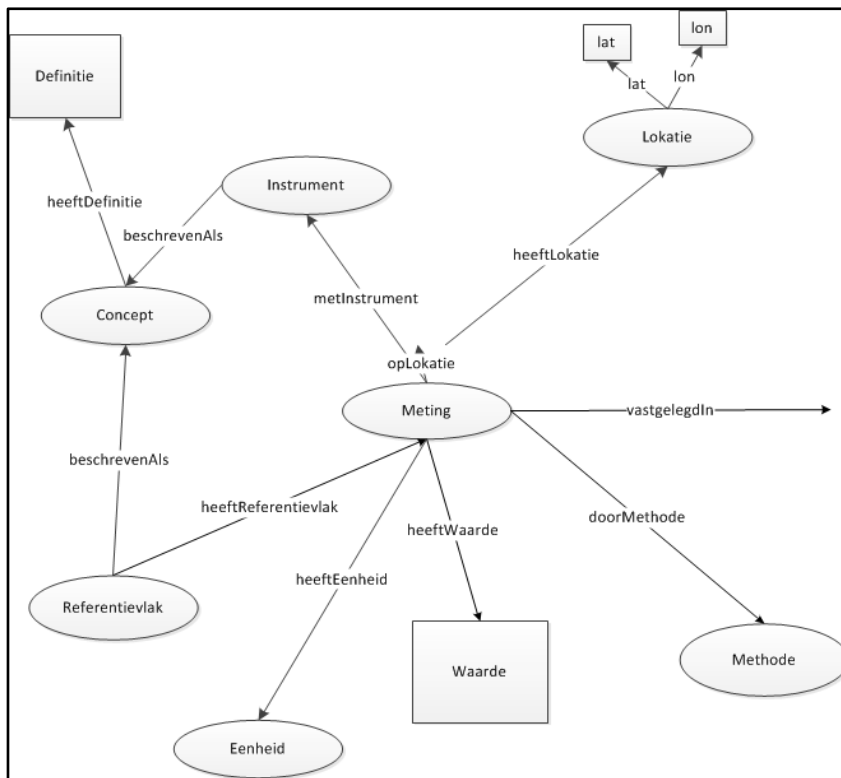
3 Linked Open Data

In navolging van het principe van het World Wide Web (WWW) bedacht Berners-Lee het web van data. Het principe van het WWW berust op het verbinden van webpagina's (documenten). De pagina's worden gelinkt via unieke adressen (Uniform Resource Identifiers - URI's), gebruik makend van het op internet gebruikelijke transportprotocol HTTP. Berners-Lee paste dit principe van het linken ook toe bij het verbinden van data. Hierdoor ontstaat een open netwerk data structuur. Het relationele datamodel, dat gebruikt wordt in gesloten omgevingen zoals bedrijfstoepassingen en intranetten, is namelijk niet geschikt voor een open omgeving. Dit komt doordat voor een wijziging of toevoeging een aanpassing in de structuur van de tabellen nodig is (figuur 2).



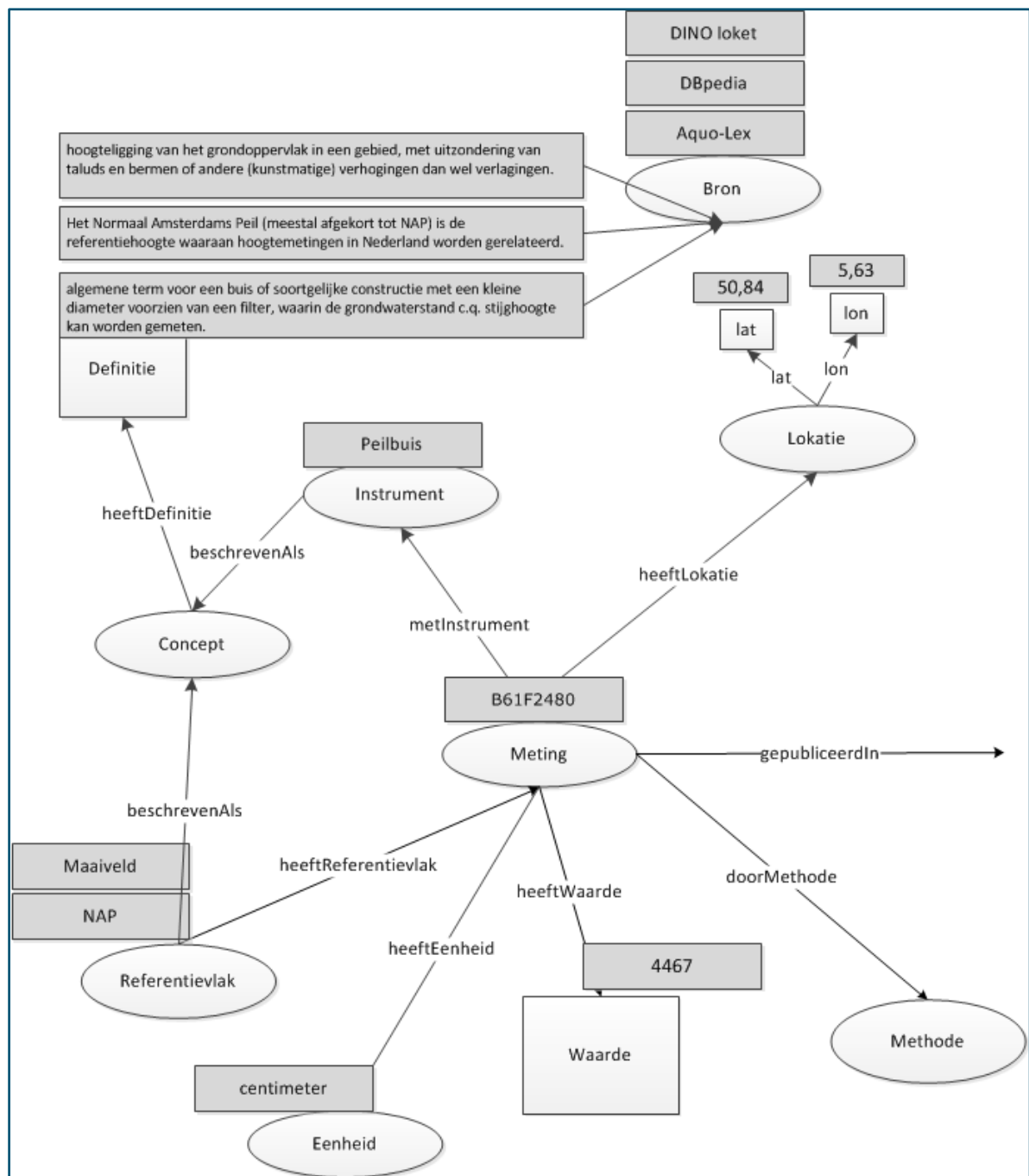
Figuur 2
Relationeel datamodel.

Om te garanderen dat organisaties wel hun eigen datamodel kunnen beheren en toch aansluiting kunnen vinden op elkaar in een open omgeving zoals het internet is een flexibeler datamodel nodig. Een netwerk datamodel heeft geen begin en geen eind. Dit is goed zichtbaar in figuur 3.



Figuur 3
 Network datamodel.

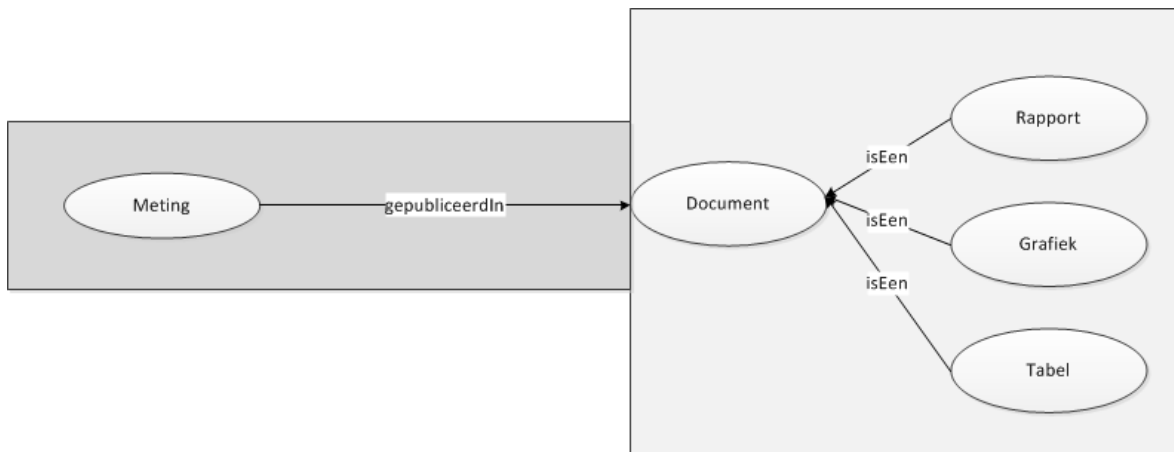
Het netwerk datamodel is zo opgezet dat nieuwe objecten (de ovalen) en eigenschappen (de pijlverbindingen) kunnen worden toegevoegd zonder dat de structuur van het bestaande model moet worden aangepast. Zo ontstaat een flexibel datamodel waaraan eenvoudig iets kan worden toegevoegd of veranderd. Dit maakt het mogelijk dat verschillende organisaties een deel van het netwerk aanbieden en dat andere organisaties daar flexibel op kunnen aansluiten. In sommige gevallen moet er een klein stukje netwerk worden aangebouwd om de koppeling mogelijk te maken. In principe kan elke partij een deel van dit hele netwerk voor zijn rekening nemen. De naam die in de vakliteratuur gegeven wordt aan het netwerk datamodel is het **semantisch web**. Het semantisch web legt de betekenis van de gegevens in een netwerk data structuur vast, net zoalsdit bij een relationeel datamodel door de kolomdefinities gebeurt. De gegevens worden vervolgens geladen in het datamodel. Dit is zichtbaar gemaakt in figuur 4. De grijze rechthoekige vlakken verbeelden de data in het netwerkmodel.



Figuur 4

Gegevens opgenomen in een netwerk datamodel.

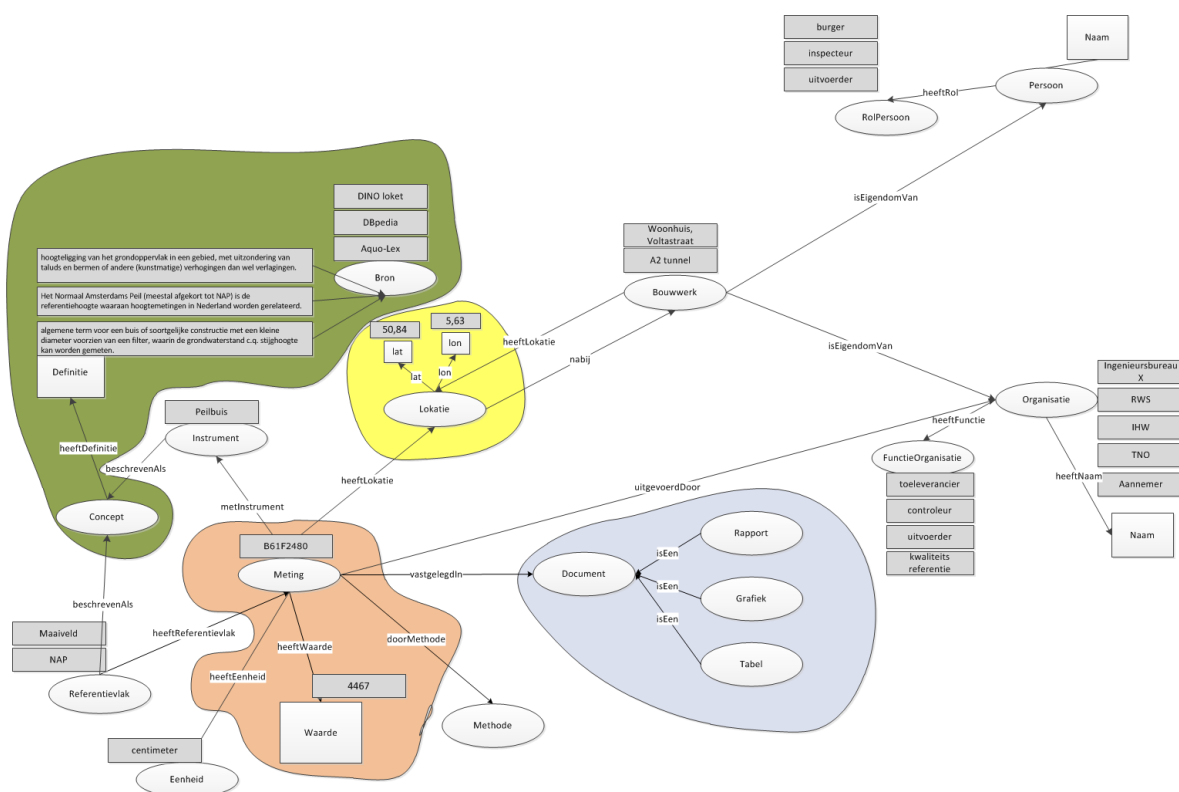
Twee afzonderlijke netwerken kunnen worden gekoppeld door een eigenschap die de twee netwerken verbindt. In figuur 4 is de eigenschap 'gepubliceerdIn' een uitgaande relatie. Als er een waarde refererend naar een externe bron wordt ingevuld, ontstaat een koppeling tussen de twee bronnen. Zo ontstaat Linked Open Data. In figuur 5 is dit principe afgebeeld.



Figuur 5

Koppelen van twee aparte databronnen (de twee grijze vlakken) via een uitgaande relatie ('gepubliceerdIn').

Het koppelen van meerdere bronnen aan elkaar levert dan een soort vlekkenpatroon op. Alle vlekken maken het web van data (figuur 6)



Figuur 6

Gelinkte bronnen als vlekken in het web van data.

Zoals eerder aangegeven wordt de data aangeduid via een URI en gekoppeld door een HTTP link. In het geval van figuur 5 is deze koppeling gecodeerd door drie elementen. Dit wordt een *triplet* genoemd:

<http://vocab.tno-metingen.nl#B61F2480-07-10-2013>

<http://vocab.tno-metingen.nl#gepubliceerdIn>

<http://vocab.tno.documenten.nl#peilbuisMetingGrafiek>

Te zien is dat het laatste deel van de triplet uit een andere bron komt (vocab.tno.documenten.nl) dan de andere twee delen (vocab.tno-metingen.nl).

De Linked Open Data triplets worden in een XML-formaat vastgelegd (Resource Description Framework, RDF). In een productieomgeving worden ze opgeslagen in een repository (triple store). Er is een querytaal voor dit soort gegevens ontwikkeld, genaamd SPARQL (W3C, 2008).

4 Case study en demonstrator

In dit hoofdstuk wordt beschreven hoe de keuze van het onderwerp van de case study tot stand is gekomen en hoe de demonstrator is opgebouwd.

4.1 Bepaling case study onderwerp

Uit de interviews met specialisten van RWS en het Informatiehuis Water is gebleken dat de bouw aan de A2 in Maastricht een goede en illustratieve casus is. In deze omgeving komen namelijk een aantal elementen samen: boven- en ondergrond en het natte en droge RWS-werkveld.

In de casus hebben we een aantal partijen opgenomen die profijt hebben van het via een open bron delen van gegevens. Ten eerste Rijkswaterstaat, de beheerder van deze weg. Ten tweede de aannemer, die de werkzaamheden uitvoert. Ten derde de burger, die graag wil weten wat er allemaal verandert en wat dat betekent voor zijn leefomgeving. Als deze drie partijen werken met dezelfde gegevens is er minder kans op miscommunicatie en misverstand over feiten bij de bouw van de A2 tunnel. Via het principe van Open Data kunnen partijen eigen overzichten maken van gegevens die zij belangrijk vinden.

4.2 Case study beschrijving

De casus spitst zich toe op het thema 'grondwater' in de regio Maastricht. Voor de beschrijving van de casus is gebruik gemaakt van de website van het A2 infrastructuur project, het Specialistisch Rapport Water (Avenue2, 2009), het Specialistisch Rapport Bouwtechniek en Bouwtijd (Avenue2, 2009) en documentatie (TNO, 2011), gegevens van het TNO DINO loket² en de Aquo standaard³ van het Informatiehuis Water. De casus is verder ondersteund door informatie uit interviews met mensen van RWS en het Informatiehuis Water.

De setting van de casus is als volgt. Tijdens de werkzaamheden aan de A2 in Maastricht wordt een tweelaagse tunnel gegraven waarin het bestemmingsverkeer en het doorgaande verkeer de stad ondergronds gaan doorkruisen. Tijdens de bouw is het belangrijk dat het grondwater de werkzaamheden niet belemmert. Het monitoren van de actuele grondwaterstand is noodzakelijk om te weten of en waar er zich mogelijk problemen kunnen gaan voordoen. De stand van het grondwater en de samenstelling van de bodem beïnvloeden in een wisselwerking de draagkracht van de ondergrond. Op plaatsen waar men er al zeker van is dat het grondwater de bouw gaat belemmeren wordt bemalen en worden tijdelijke maatregelen, zoals bijvoorbeeld een compartimenteringwand, geplaatst (Avenue2, 2009).

Voor burgers die in de buurt van de bouw van de tunnel wonen is het spannend wat het effect van de bouw is op hun woonsituatie. Vooral in de buurt van de Voltastraat is de grondsoort (kalksteen) zwak. De bouwwerkzaamheden rond de tunnel in combinatie met de grondsoort op deze plek en de stand van het grondwater kunnen voor problemen zorgen.

² <http://www.dinoloket.nl/>

³ <http://www.aquo.nl/>

In de demonstrator worden schermen van drie partijen getoond. Ten eerste Rijkswaterstaat, de beheerder van deze weg. Ten tweede de uitvoerende aannemer. Ten derde de burger, die graag wil weten welk effect de bouw heeft op zijn leefomgeving. De schermen illustreren dat deze drie partijen voor een deel naar dezelfde gegevens kijken. Hun schermen verschillen door de taak of de interesse die ze hebben.

De keuze van de gegevens die in de demonstrator worden gebruikt is gebaseerd op de beschrijving van het A2 project in de projectdocumentatie (Avenue2, 2009; Avenue2, 2009) en op de inhoud van de interviews.

4.3 Demonstrator

Om de casus illustratief te ondersteunen is een demonstrator gemaakt. Deze demonstrator moet inzichtelijk maken hoe open gegevens bijdragen aan een verbetering van de informatievoorziening in een project als A2 Maastricht. Door gebruik te maken van Open Data kijken alle partijen als het ware door dezelfde bril. Bovendien hebben zij door de aanwezigheid van Open Data de mogelijkheid om gegevens te combineren en te tonen op een voor hen specifieke nuttige manier. Omdat de gegevens gelinked zijn is het mogelijk om de context van de gegevens op te vragen. Hiermee wordt bedoeld dat het hele datamodel kan worden bekeken en dat de relaties tussen de begrippen in het datamodel kunnen worden vastgesteld.

De demonstrator bestaat uit drie lagen: een gegevenslaag, een HTML pagina laag en een user interface laag. Voor deze gelaagdheid is gekozen om inzichtelijk te maken hoe Linked Open Data werkt en hoe het er uitziet voor de gebruiker en data beheerder.

4.3.1 RDF data laag

Deze laag bevat de werkelijke Linked Open Data, de RDF-bestanden. Deze zijn normaal in een repository opgenomen, omdat zo betere performance en beveiliging kan worden gerealiseerd. Voor demonstratiedoeleinden in deze case study is er voor gekozen om met losse RDF-bestanden te werken.

De demonstrator is gebouwd met c.q. gebaseerd op gegevens van bestaande bronnen. Het prototype van de objectencatalogus van het Informatiehuis Water (IHW) zoals die in het project 'objectencatalogus' is opgeleverd was als enige al in het Linked Open Data formaat beschikbaar (Verhelst, Bulens et al., 2010). De andere bestaande bronnen zijn geraadpleegd, en delen ervan zijn met een eigen ontwikkeld datamodel als Linked Open Data bestand gebouwd. Dit zijn:

- TNO DINO voor details over de meting van het grondwaterpeil ter plekke.
- Meetgegevens van de NV Waterleiding Maatschappij Limburg (via DINO).

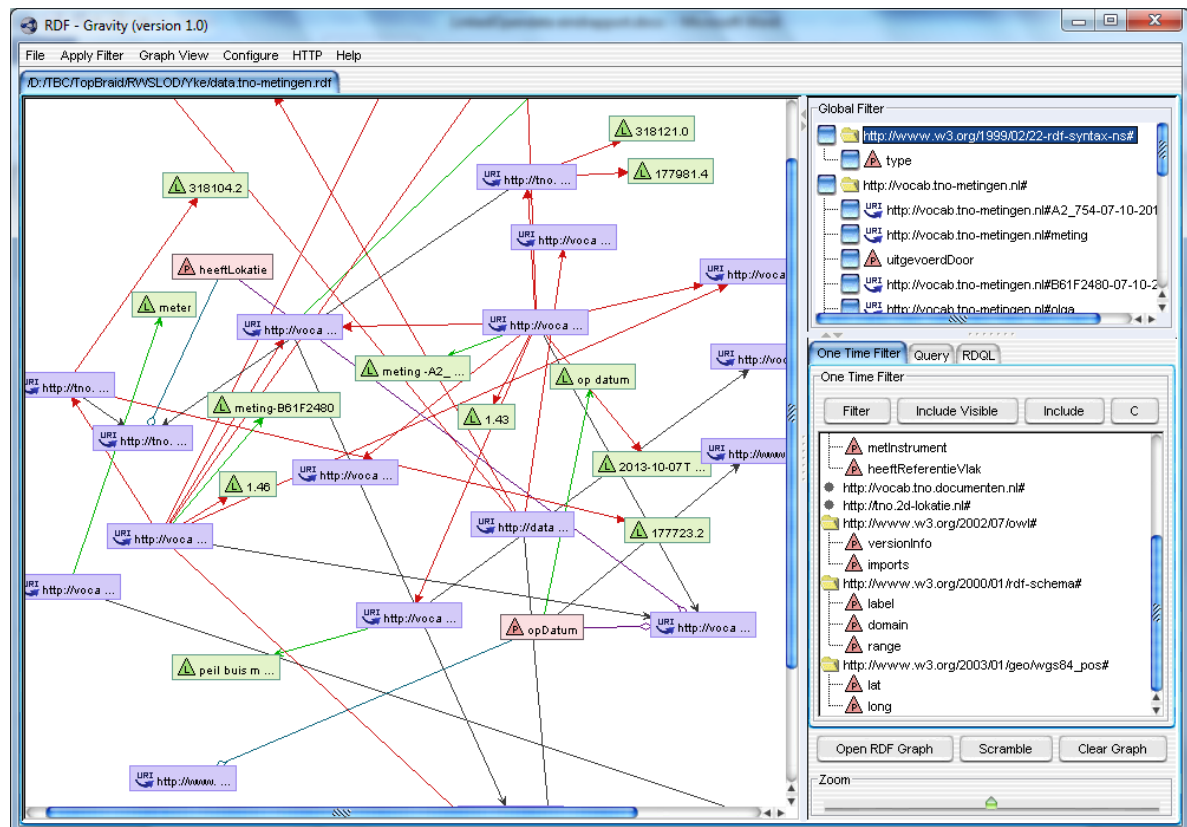
Van het bestaan van andere bronnen hebben we kennis genomen, maar we hadden geen aanvullende detailgegevens voor handen. We hebben de inhoud van deze bronnen in onze demonstrator verwerkt op basis van de indruk die we van deze bronnen hebben gekregen.

- OLGA
- DONAR

Er is gebruik gemaakt van twee afbeeldingen van het project A2 die via RWS zijn verkregen van de aannemer. Het gaat hier om een bouwschets van de A2-tunnel en een bodemprofiel. Deze afbeeldingen zijn gebruikt in de schermen van de demonstrator.

De demonstrator heeft twee soorten RDF-bestanden: data bestanden en vocabulaire bestanden. De typing verraaft het verschil: in de data bestanden zit de Linked Data, de vocabulaire bestanden vormen het datamodel van het data netwerk.

RDF bestanden kunnen worden geopend in een ontology editor, zoals bijvoorbeeld Protégé⁴ of Topbraid Composer⁵. Een goede web viewer is RDF Gravity⁶ van de universiteit van Salzburg. Deze viewer toont de RDF-bestanden als een netwerk.



Figuur 7

Eén van de RDF-bestanden (data.tno-metingen.nl) gezien via de RDF Gravity viewer.

4.3.2 HTML pagina laag

Om de werking van de links in de RDF-gegevens te tonen hebben we een extra laag aangebracht met als enige doel het tonen en doorklikbaar maken van de links in de RDF-data. Deze laag bestaat uit HTML pagina's die de RDF-data leesbaar maken via een browser. De gegevens zijn in een tabel weergegeven met links de eigenschappen en rechts de waarden. Waar er een link tussen de gegevens is, is deze doorklikbaar via een HTTP link.

⁴ <http://protege.stanford.edu/>

⁵ http://topquadrant.com/products/TB_Suite.html

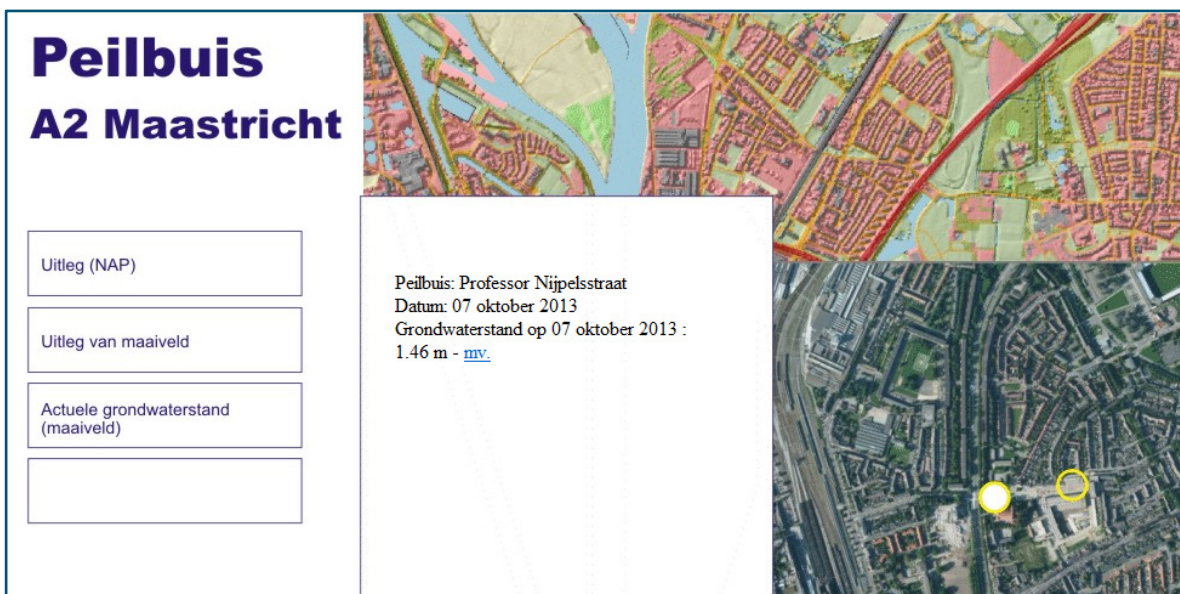
⁶ <http://semweb.salzburgresearch.at/apps/rdf-gravity/index.html>

Om deze verbondenheid in de data te zien is een goed startpunt het document met de naam StartRDFLinks.htm. Dit opent een gegevensset van de grondwatermeting, waarin doorgeklikt kan worden naar andere begrippen, bijvoorbeeld “locatie” en “heeftReferentieVlak”. Via het betreffende referentievlak (maaiveld) kan weer worden doorgeklikt (via “beschrevenAls”) naar de beschrijving van het begrip 'maaiveld' in de objectencatalogus.

4.3.3 User interface laag

Tijdens het onderzoek is geprobeerd een indruk te krijgen van de werkelijke schermen waar de verschillende partijen mee werken in een project als A2. Wij wilden graag de schermen laten zien waarin voor de drie partijen de grondwatergegevens een rol speelden. Dit is helaas niet gelukt omdat de schermen die wij voor de demonstrator voor ogen hadden in werkelijkheid niet bestonden. Men gaf aan dat schermen waarin de actuele grondwaterstand zou worden gecombineerd met de eigenschappen van het bodemprofiel ter plekke wel wenselijk zou zijn. Vervolgens hebben we gezocht naar schermen die een impressie zouden kunnen geven van hoe de partijen gegevens van grondwater zouden zien in hun eigen werksituatie. Ook dit is maar gedeeltelijk gelukt, de betreffende schermen bleken niet toegankelijk. Daarom is gekozen voor het maken van de schermen volgens een geheel vrije interpretatie op basis van informatie die we via de interviews hebben gekregen.

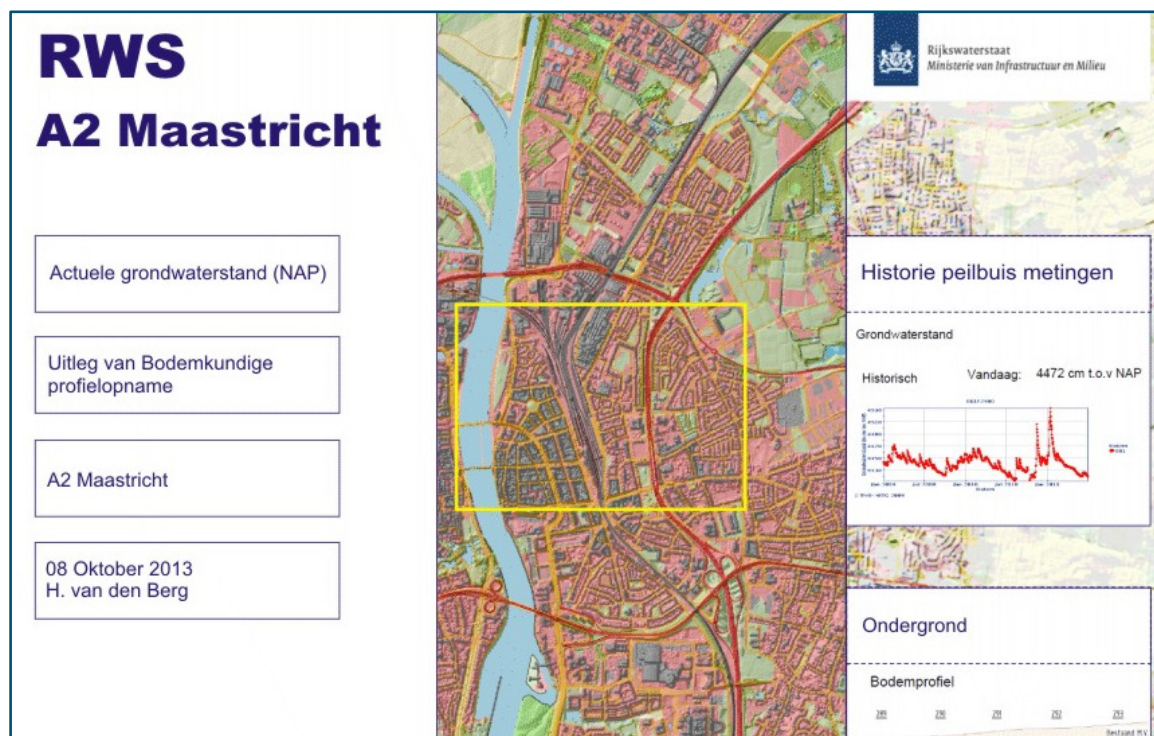
Er is voor elke partij (burger, aannemer en RWS) een scherm gemaakt. In elk scherm zijn een aantal links opgenomen die verwijzen naar bron data, die in een aantal gevallen voor alle partijen gelijk was (de actuele grondwatermeting). De schermen verwijzen ook naar andere bronnen op het internet, zoals DBpedia en documentatie van het DINO-loket.



Figuur 8
Scherf Burger.

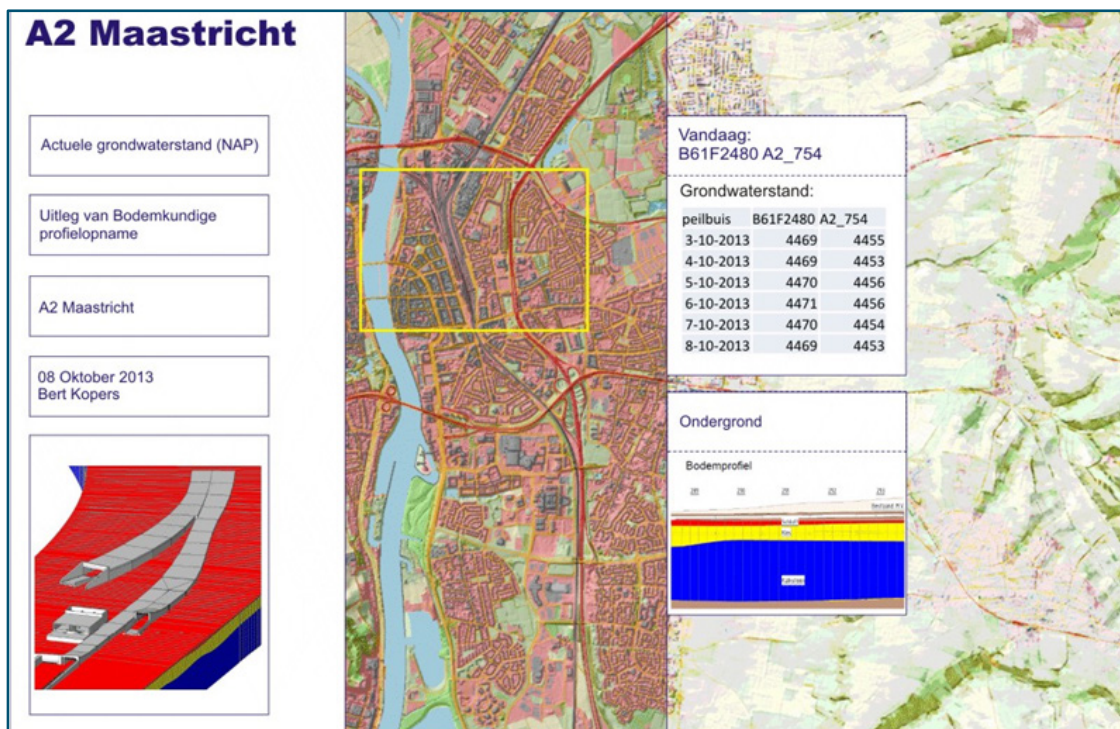
De burger is geïnteresseerd in veranderingen in de grondwaterstand door mogelijke schade aan zijn woonsituatie door verzakkingen of lekkages. Het scherm van de burger toont de grondwaterstand op een

meetpunt met als referentievlak het maaiveld. Bovendien staat in het scherm van de burger een link naar informatie over begrippen die bij deze meting een rol spelen.



Figuur 9
Scherf Rijkswaterstaat.

Rijkswaterstaat is als beheerder van de weg gebaat bij zo min mogelijk verstoring en vertraging rond de bouw van de tunnel in algemene zin. Dit gaat verder dan de bouwput alleen. De schermen van RWS hebben vooral een inspectie doel. Veranderingen van de grondwaterstand in de tijd zullen de RWS-inspecteur interesseren.



Figuur 10
Scherf aannemer.

De aannemer wil graag de grondwaterstand zien voor het beschermen van de bouwput. De historie van de metingen interesseren hem, omdat dit een indicatie kan zijn voor problemen in de stabiliteit van de ondergrond. Hij wil ook een overzicht zien van de vorderingen van het werk.

De links op de schermen openen gegevens uit de RDF-data (zichtbaar via de HTML pagina's). Voorbeeld: actuele grondwaterstand (meting), uitleg maaiveld. In sommige gevallen openen ze links naar bronnen op het internet (b.v. uitleg NAP uit DBpedia, dit begrip staat niet in de Aquo catalogus). In het scherm van de aannemer en van RWS is een link opgenomen naar documentatie over het gebruik van datatypen in de BRO (TNO, 2011).

5 Conclusies en discussie

In deze case study hebben we een indruk gegeven van de gegevensuitwisseling via Linked Open Data tussen partijen in het RWS-werkveld. Hiermee zijn nog niet alle aspecten van dit onderwerp voldoende belicht. Daarom wordt in het afsluitende hoofdstuk aandacht besteed aan de volgende vragen:

- wat is de meerwaarde van Open Data vergeleken met de huidige manieren van gegevensuitwisseling?
- wat is de meerwaarde van het linken van Open Data?
- hoe moet de gegevensuitwisseling via Linked Open Data worden voorbereid (welke stappen zijn nodig)?

5.1 Meerwaarde van Open Data

Aangezien er in de afgelopen decennia hard is gewerkt aan integratie en standaardisering zijn de volgende vragen gerechtvaardigd: 'welke vernieuwing brengt Open Data vergeleken met de bestaande mogelijkheden van gegevensuitwisseling door webservices en API's? Kan de use case worden gebouwd met bestaande technieken en standaarden?'

Om met de eerste vraag te beginnen: met Open Data kan meer dan wat mogelijk is op basis van uitwisseling via een webservice en het real time opvragen van gegevens (bijvoorbeeld via een API). Het verschil zit vooral in de directe toegang tot de datastructuur. Webservices bieden toegang via een servicegericht vraag- en aanbod protocol. API's geven toegang tot achterliggende gegevens via een set van programmacoderegels, gespecificeerd door de leverancier van een software product. Open Data gaat over toegang tot de data (de records) inclusief de datastructuur (datamodel). Dit biedt de gebruiker de mogelijkheid om zelf query's op te stellen, de resultaten te filteren, te analyseren, te visualiseren en te combineren met andere data. De impact hiervan kan het beste geïllustreerd worden door te kijken naar wat de Open Data site van de UK (data.gov.uk) mogelijk maakt. De Engelse overheid heeft via dit portaal van de in totaal meer dan 7700 datasets, 157 sets gepubliceerd in domein 'Transport'⁷, het domein van RWS. De transport data bevat gegevens over de drukte van de wegen, het aantal ongelukken, de locatie van servicepunten et cetera. Op basis van deze data zijn door het publiek naar eigen inzicht apps gebouwd die, al dan niet tegen betaling aan de ontwikkelaar, door iedereen kunnen worden gebruikt. De gebruikers van deze data laten via social media en de feedback-mogelijkheden op het portaal aan de overheid weten wat ze met deze gegevens willen en wat de kwaliteit van de gegevens is. Zo ontstaat een terugkoppeling die effect heeft op de kwaliteit van de gegevens en de waardering van het publiek op de geleverde overheidsdienst. Bovendien is door het publiceren van de ruwe data de overheid niet meer primair de aangewezen partij om de gegevens te tonen via een website. Beide ontwikkelingen kunnen besparingen en verbeteringen opleveren.

Door de open structuur van de gegevens kunnen zij worden gecombineerd met in principe elke andere Open Data set. Dit kan tot verrassende nieuwe inzichten en innovatieve toepassingen leiden. In onze case study is aangegeven dat er meerwaarde ontstaat wanneer grondwater- en bodemgegevens in een open dataformaat worden gepubliceerd. De verschillende partijen uit de case study kunnen hiermee zelf hun eigen visualisaties en applicaties ontwikkelen, toegespitst op hun eigen behoefte. Omdat bovendien andere partijen (bijvoorbeeld waterschappen, ingenieursbureaus, burgerinitiatieven) andere analyses en combinaties van data kunnen maken

⁷ Website bekeken op 14-12-2011

ontstaat er een heterogeen en optimaal gebruik van de data. Hierdoor ontstaan nieuwe inzichten die weer nieuwe data opleveren.

Het antwoord op de tweede vraag wordt gegeven in de volgende paragraaf.

5.2 Meerwaarde van Linked Open Data

De meerwaarde van Linked Open Data vergeleken met Open Data kan worden geïllustreerd door de vraag te beantwoorden welk voordeel het gebruik van een netwerk datamodel biedt ten opzichte van een geïsoleerd (relationeel) model. Als gebruik wordt gemaakt van Linked Open Data, hetzij door LOD te serveren, hetzij door het te gebruiken, maak je deel uit van het web van data. Daarin is in principe geen beperking van de 'database' die gebruikt kan worden, alle data is gekoppeld en via links vindbaar door te queryen via SPARQL. Dit is geïllustreerd in hoofdstuk 3.

Door de eigenschap dat de gegevens gelinkt zijn, is bovendien de context van de gegevens waarmee gewerkt wordt opvraagbaar. Zie nogmaals figuur 4 op pagina 17: Gegevens opgenomen in een netwerk datamodel. Hierin is te zien dat een referentievlak een relatie heeft met een locatie via een meting. Dit heeft vooral meerwaarde in omgevingen waarbij door de context de betekenis van een begrip sterk kan veranderen. In de praktijk van RWS bijvoorbeeld heeft het begrip 'route' binnen de natte RWS een andere betekenis dan bij de droge RWS. Door het begrip 'route' op te nemen en in beide contexten te definiëren (bijvoorbeeld: 'route voor vaartuig', 'route voor voertuig', 'af te leggen weg om op een bestemming te komen') wordt het data netwerk geschikt voor meerdere toepassingen.

In de demonstrator kan de werking en de meerwaarde van LOD het beste worden gezien door de links door te klikken in de HTML-schermen. De user interface-schermen zijn hiervoor minder geschikt, doordat een willekeurige user interface in principe niet wordt bepaald door het soort data dat er onder ligt. Met andere woorden: de gebruiker merkt niets van de Linked Data. Een user interface kan gegevens uit elke bron op dezelfde manier zichtbaar maken door gebruik te maken van programmeer frameworks (zoals Jena⁸ voor RDF). Dat is niet anders voor Linked Data. Het antwoord op de vraag aan het begin van het hoofdstuk is daarom: de user interface-schermen van de use case kunnen ook worden gebouwd met bestaande technieken en standaarden. De HTML-schermen illustreren beter de meerwaarde van LOD, doordat zij de gelinkte structuur zichtbaar maken.

Bestaande voorbeelden van user interfaces die gebaseerd zijn op onderliggende Linked Data zijn de websites van de BBC (bv: <http://www.bbc.co.uk/nature>). Omdat bekend is dat hier Linked Data als bron onder ligt is de data structuur herkenbaar (de geordendheid van de gegevens, de links naar classes zoals <http://www.bbc.co.uk/nature/animals/#class-Mammal>). Over de achterliggende techniek van de BBC website van het WK voetbal in 2010 is in een blog gepubliceerd (Rayfield, 2010).

De volgende paragraaf schetst hoe RWS kan aansluiten op het web van Linked Open Data.

5.3 Stappen naar implementatie

Hoewel Linked Data niet altijd 'open' hoeft te zijn, er zijn scenario's denkbaar dat deze techniek in een corporate intranet wordt ingezet, ligt het voor de hand om als eerste te onderzoeken hoe ingewikkeld het is,

⁸ <http://jena.sourceforge.net/>

vooral organisatorisch en juridisch, om gegevens open te stellen. Hiervoor is door het project Handreiking Open Overheidsdata (Burke) een duidelijke handleiding met infographic gemaakt.

Als de data ge'open'd mag worden, dan is de volgende stap een conversie van het huidige formaat naar het open formaat. In dit geval LOD formaat, triples, RDF (of een andere triple notatie). Daarvoor is eerst het datamodel nodig. Het datamodel, als onderdeel van het semantisch web, wordt ook wel vocabulaire genoemd. Het is aan te raden eenvoudig te beginnen en niet meteen te ingewikkelde vocabulaires te maken. Als er al vocabulaires bestaan voor het gewenste doel, dan is het aan te raden deze te gebruiken. Dit scheelt niet alleen veel werk, maar 'men' verwacht ook dat de geëigende vocabulaires worden gebruikt.

Een eenvoudig vocabulaire is een thesaurus. Dit is een datamodel waarbij de begrippen worden gedefinieerd door beschrijvingen (bv: synoniem, voorkeursterm) en door eenvoudige relaties tussen de termen (bv. bredere relatie, nauwere relatie) . Voor een thesaurus datamodel bestaat een gestandaardiseerde codering. SKOS (Simple Knowledge Organisation System). SKOS is gebruikt voor de Aquo objectencatalogus.

Enkele geëigende (van 'gestandaardiseerd' kan nog niet worden gesproken) vocabulaires zijn naast SKOS zijn:

- Dublin Core (standaard termen)
- WGS84 (voor de positie van een locatie)
- FOAF ("Friend of a Friend", mensen, relaties en organisaties)
- Geonames (plaatsnamen en andere topografische kenmerken)

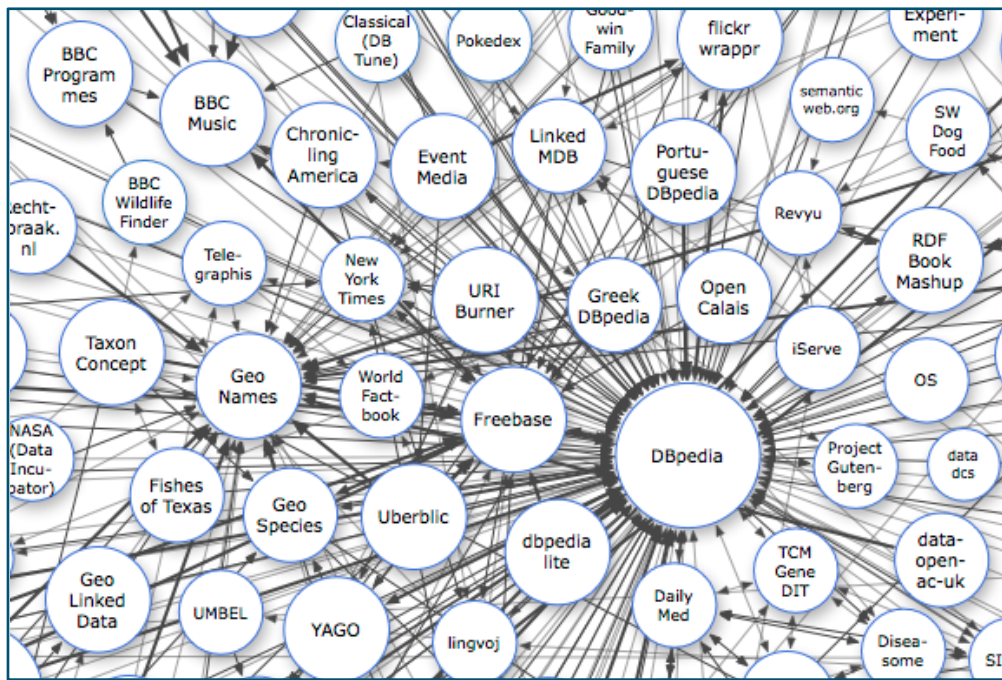
De Nederlandse overheid schrijft voor ('pas toe of leg uit') de OWMS (Overheid Web Metadata Standaard).

Een netwerk datamodel (semantisch datamodel) van RWS moet dus idealiter bestaan uit bovengenoemde vocabulaires als basis met daaraan toegevoegd specifieke RWS-begrippen en hun relaties.

De stap hierna is het converteren van de data die men wil publiceren naar een triple formaat, en dit te koppelen aan het datamodel (vergelijk figuur 4 op pagina 17). In een voor dit doeleinde geschikte editor (Protégé, Topbraid Composer) is dit hetzelfde als het invullen van een formulier. Eén en ander kan ook worden geprogrammeerd voor een batch load. Maak hiervoor gebruik van de Jena API.

Het eigen Open Linked Data dataset is nu gereed. Om het open te stellen moet het worden gepubliceerd op een server achter een voor andere partijen herkenbaar webadres. Gebruikelijk is: <http://data.rijkswaterstaat.nl>.

Het LOD publiceerwerk is gereed als het LOD-dataset wordt gekoppeld aan een ander LOD-dataset, zodat het niet meer geïsoleerd in het web van data staat. De eenvoudigste manier is het te koppelen aan DBpedia, de LOD-versie van Wikipedia. Een veel gebruikte koppel-relatie is 'same As'. Figuur 11 geeft een deel van alle gekoppelde LOD-bestanden op het internet (de zogenaamde LOD cloud) weer waarin goed is te zien dat DBpedia een centrale rol speelt.



Figuur 11
Deel van de LOD cloud (Cyganiak en Jentzsch, 2011).

Referenties

Avenue2, 2009. Specialistisch Rapport Bouwtechniek en Bouwtijd, Dialoogronde IV 22 april 2009.

Avenue2, 2009. Specialistisch Rapport Water, Dialoogronde IV 22 april 2009. de Groene Loper Vanzelfsprekend.

Berners-Lee, T., 2006-2009. 'Linked Data.' 2011, from <http://www.w3.org/DesignIssues/LinkedData.html>.

Broek, T.v.d., N. Huijboom et al. 2011. Open Overheid, TNO.

Burke, J. Z., Ton. 'Handreiking Open OverheidsData.' from <http://vrijedata.nl>.

Cygniak, R. and A. Jentzsch, 2011. 'The Linking Open Data cloud diagram.' From <http://richard.cygniak.de/2007/10/lod/>.

data.gov.uk. 2011. 'Transparency (address of UK PM Cameron).' From <http://data.gov.uk/about>.

EC, 2003. DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information, THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION.

EC, 2010. Een digitale agenda voor Europa. Brussel, EC.

Lakhami, K., R. Austin en Y. Yi, 2010. Data.gov., Harvard Business School.

Min van EL&I, 2011. Digitale Agenda.nl ICT voor innovatie en economische groei.

overheid.nl, 2011. 'Toespraak minister Donner bij de opening van het open dataportaal.' From <http://www.rijksoverheid.nl/regering/het-kabinet/bewindspersonen/piet-hein-donner/toespraken/2011/09/15/toespraak-minister-donner-bij-de-opening-van-het-open-dataportaal.html>.

Rayfield, J., 2010. BBC World Cup 2010 dynamic semantic publishing. From http://www.bbc.co.uk/blogs/bbcinternet/2010/07/bbc_world_cup_2010_dynamic_sem.html.

Rijksoverheid, 2011. 'Innovatie-estafette 2011: baaiert aan innovatieve initiatieven.' From <http://www.rijksoverheid.nl/nieuws/2011/10/04/innovatie-estafette-2011-baaiert-aan-innovatieve-initiatieven.html>.

Rijksoverheid, 2011. 'Satellietbank schatkamer voor app-bouwers in topsectoren.' From <http://www.rijksoverheid.nl/nieuws/2011/10/04/satellietbank-schatkamer-voor-app-bouwers-in-topsectoren.html>.

TNO, 2011. Basisregistratie Ondergrond (BRO), Datatypen met uitleg. Infoblad.

Verhelst, E.C.H., J.D. Bulens et al., 2010. IDSW objectencatalogus: onderzoek naar de toepasbaarheid van een semantisch datamodel als overkoepelende Aqua standaard. Alterra-rapport 2027. Wageningen.

W3C, 2008. SPARQL Query Language for RDF.



Alterra is onderdeel van de internationale kennisorganisatie Wageningen UR (University & Research centre). De missie is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen UR bundelen negen gespecialiseerde en meer toegepaste onderzoeksinstituten, Wageningen University en hogeschool Van Hall Larenstein hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 40 vestigingen (in Nederland, Brazilië en China), 6.500 medewerkers en 10.000 studenten behoort Wageningen UR wereldwijd tot de vooraanstaande kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen natuurwetenschappelijke, technologische en maatschappijwetenschappelijke disciplines vormen het hart van de Wageningen Aanpak.

Alterra Wageningen UR is het kennisinstituut voor de groene leefomgeving en bundelt een grote hoeveelheid expertise op het gebied van de groene ruimte en het duurzaam maatschappelijk gebruik ervan: kennis van water, natuur, bos, milieu, bodem, landschap, klimaat, landgebruik, recreatie etc.

Meer informatie: www.alterra.wur.nl