

# Raw data now?

## Het ontstaan van productformules en -concepten

**W**e zijn inmiddels enkele jaren verder sinds de bedenker van het World Wide Web, Tim Berners-Lee in een inspirerende TED talk<sup>1</sup> een oproep deed voor 'raw data now!'. De beschikbaarheid van data in onbewerkte vorm is cruciaal voor de verdere ontwikkeling van het internet. Maar hoe eenvoudig is dat? En waar staan we nu?

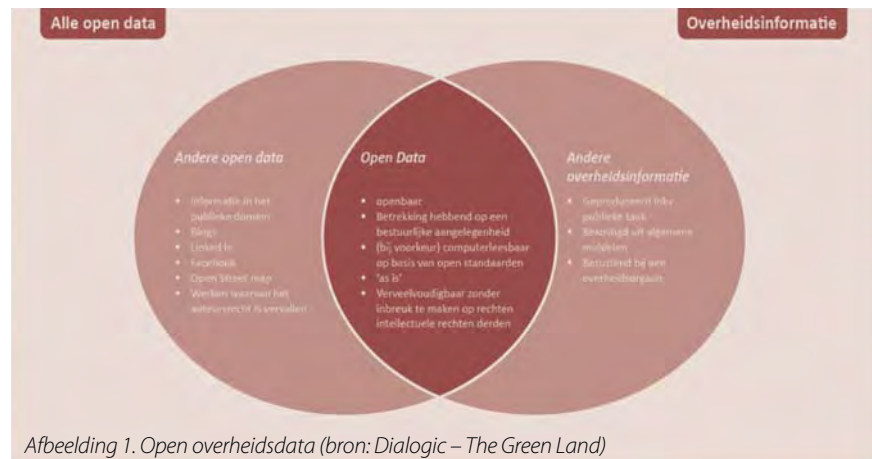
De afgelopen jaren is open data een belangrijk thema geworden. De vergaande informatisering van de maatschappij biedt dan ook veel kansen om met data tot nieuwe samenwerking en dienstverlening te komen. En samenwerking begint met het laten zien wat je in huis hebt. Daarbij ligt de focus sterk op de data van de overheid. Daarvoor, is de beredenering, is immers al door de samenleving betaald.

Niet alle data van de overheid is in potentie open data. En natuurlijk is niet alle data van de overheid. Dit wordt gevisualiseerd in het onderstaande plaatje.

### Open data op de beleidsagenda

Binnen de overheid is al langer aandacht voor het beschikbaar stellen van data aan de samenleving. In 2007 werd de Europese kaderrichtlijn Inspire van kracht, gericht op het realiseren van een Europese geoinformatie infrastructuur. In Nederland wordt de data beschikbaar gesteld via het Nationaal Georegister<sup>2</sup>. Daarnaast liepen er al langer discussies over bijvoorbeeld het Nationaal Wegenbestand en de postcode-database. Voor beide registers is in 2011 (door een rechtbank) bepaald dat de data zonder restricties vrijgegeven mag worden.

Het thema komt vol in de belangstelling als in mei 2011 de Digitale Agenda.nl<sup>3</sup> aan de Tweede Kamer wordt gestuurd. Deze agenda zet de ambities neer wat betreft inzet van ICT voor groei en welvaart en welke randvoorwaarden hiervoor nodig zijn. Open data is hierin een belangrijk gegeven. Ondernemers kunnen de data immers verwerken tot nieuwe



Afbeelding 1. Open overheidsdata (bron: Dialogic – The Green Land)

toepassingen en diensten. De precieze baten zijn niet voor de Nederlandse situatie berekend. Een Deense studie verwacht echter een indirecte waarde van 80 miljoen euro. Voor de Nederlandse economie, die 2,5 keer groter is, zou dit een waarde van commercieel hergebruik van 200 miljoen opleveren. De opbrengsten van commercieel hergebruik voor de hele Europese economie worden geschat op 28 miljard euro<sup>4</sup>.

### Een jaar van verkenning

De Digitale Implementatie Agenda.nl<sup>5</sup> uit eind 2011 beschrijft de activiteiten om de Digitale Agenda tot een succes te maken. In het programma wordt aangegeven dat de komende jaren zoveel mogelijk overheidsdata openbaar wordt gemaakt, op basis van behoefte van ondernemers. Voor het (toen zo geheten) Ministerie van EL&I werd zelfs bepaald dat alle datasets uiterlijk in 2015 beschikbaar komen volgens het principe 'open, tenzij'. Als eerste voorbeelden gelden de pilot rond de open data rond exportvoorwaarden in de zaaizadensector en de inrichting van een Nationale Satellietbeelden-databank (zie het volgende artikel van Frans Lips).

Het jaar 2012 staat verder in het teken van het wegnemen van belemmeringen rond het hergebruik van overheidsdata. Dit gebeurt onder andere door een wijziging van (hoofdstuk Va van) de Wet openbaarheid van bestuur (Wob). In een

brief aan de Tweede Kamer (mei 2011) wordt immers gemeld dat in de huidige wet het mogelijk is om voorwaarden te stellen en tarieven in rekening te brengen voor informatie die is verzameld en geproduceerd in het kader van de uitvoering van overheidstaken. Dit is een knelpunt voor open data. Een wetswijziging zou de knelpunten moeten wegnemen.

Verder is een handreiking<sup>6</sup> in ontwikkeling waarin het principe 'open, tenzij' nader wordt uitgewerkt. Dat dit nodig is blijkt uit de vele vragen vanuit meerdere overheidsinstanties over de toepassing en consequenties van het open data beleid. Vooral thema's als privacy, de kwaliteit van open data en de financiering van het beschikbaar stellen van data zorgen in de dagelijkse praktijk voor discussie.

### Privacy

Bij het beschikbaar stellen van data als open data speelt privacy een belangrijke rol. Als er geen privacyregelgeving van toepassing is op de data zijn er weinig beperkingen. Maar zodra dat wel het geval is, gelden er strenge eisen. De Wet bescherming persoonsgegevens (Wbp) geeft regels over de verwerking van persoonsgegevens. Persoonsgegevens zijn gegevens die een natuurlijke persoon identificeren of kunnen identificeren, met de naam als duidelijkste voorbeeld. Maar ook gegevens die een persoon indirect

1 Zie [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)

2 Zie <http://www.nationaalgeoregister.nl>

3 Kamerbrief 'Digitale Agenda.nl', TK 2010 – 2011, 29 515, nr. 331

4 Zie Study Vickery, <http://eurogi.org/downloads/file/139-vickery-study-potential-market-value-of-psi-re-use-in-europe.html>

5 <http://www.rijksoverheid.nl/onderwerpen/ict/documenten-en-publicaties/rapporten/2011/12/13/digitale-implementatie-agenda-nl.html>

6 Zie <https://data.overheid.nl/handreiking>

identificeren kunnen een persoonsgegeven zijn, omdat ze door combinatie met andere gegevens toch met een bepaalde persoon in verband worden gebracht.

De wet geeft een brede interpretatie van het begrip persoonsgegeven waardoor zelfs geanonimiseerde en geaggregeerde gegevens als zodanig kunnen worden aangemerkt. In het rapport<sup>7</sup> 'Aansprakelijkheid en open data' wordt ook aangegeven dat door technische ontwikkelingen het steeds makkelijker is om grote gegevensbestanden te analyseren en combineren. Dat maakt het mogelijk om persoonsgegevens te destilleren, zelfs als ze geanonimiseerd of geaggregeerd zijn.

De regimes van de Wob en de Wbp zijn, waar het de effecten aangaat, grotendeels tegengesteld: voor Wob geldt 'openbaar tenzij' voor Wbp geldt dat 'verwerking (van persoonsgegevens) niet mag, tenzij'. De Wob biedt wel ruimte voor het maken van een afweging over het (algemene) belang van verstrekking versus het eerbiedigen van de persoonlijke levenssfeer van betrokken personen.

In de praktijk blijft het voor overheden een lastig proces. Het is immers niet altijd duidelijk wanneer er nu sprake is van persoonsgegevens en de afweging om het alsnog beschikbaar te stellen in het algemeen belang zorgt voor veel discussie. Bij veel momenteel als open data beschikbare (geo-)datasets speelt privacy dan ook vrijwel geen rol of zijn wettelijke bepalingen van kracht die bepalen dat de specifieke informatie beschikbaar moet zijn (zoals Inspire).

### **Kwaliteit**

Veel registraties bij de overheid zijn ingericht ter ondersteuning van de uitvoering van beleid. Dat betekent dat de data bestemd is voor een specifiek doel en dat de kwaliteit van de gegevens daarvoor toereikend zal zijn. Met het beschikbaar stellen van de data verandert deze context. De vraag is dan of de gegevensdefinities bruikbaar zijn voor derden. En ook of de dataset voldoende volledig en/of actueel is. Nieuwe gebruikers zorgen immers voor meer of andere eisen.

Soms speelt ook mee dat het beschikbaar stellen van data inzicht biedt in de gegevens waarop de overheid besluitvorming baseert. Dit kan gevoeligheden opleveren. Toch biedt open data kansen om de kwaliteit van de gegevens ook een impuls te geven, in eerste instantie door de feedback die wordt verkregen. Bij intensief gebruikte

gegevens is het, vanuit het gezamenlijk belang, goed denkbaar dat er meer gebruik gemaakt gaat worden van (open) standaarden en daarmee beter en efficiënter aansluit op reeds bestaande registraties binnen en buiten de overheid. In dat kader is het interessant de ontwikkelingen rond linked open data<sup>8</sup> in de gaten te houden. Linked open data biedt een perspectief om data aan elkaar te relateren en daardoor uniform op het web te ontsluiten.

### **Financiering**

Dat open data kansen biedt, daar zijn velen het over eens. Maar voor overheden zijn de baten vaak niet direct zichtbaar. Die liggen immers aan private zijde. De voordelen aan de zijde van de overheid (betere kwaliteit, meer samenhang in registraties) laten zich moeilijk calculeren. Het blijkt daarom in de praktijk lastig om substantiële financiering te organiseren voor het beschikbaar stellen van data, zeker als daar hoogwaardige webservices voor noodzakelijk blijken te zijn. In de praktijk zijn overheden daarom pragmatisch: Vooral de data die tegemoet komt aan een daadwerkelijk behoefte, wordt beschikbaar gesteld. In deze gevallen is de business case vaak concreet aanwezig. Hierin ligt ook het belang van overheden om actief deel te nemen aan open data bijeenkomsten.

Uiteraard wordt soms ook data vrijgegeven om er zelf, als overheidsorganisatie, wat aan te hebben. Dit speelt vaak als een kwaliteitsimpuls wenselijk is en de verwachting is dat dit met voldoende feedback mogelijk is. En soms is open data een opmaat tot het (deels) afbouwen van bestaande informatiekanaalen. De besparing ligt dan in een verschuiven in het dienstenaanbod.

### **Een stap vooruit**

Aan open data wordt veel potentie toegekend als stimulans voor maatschappelijke innovatie en om te komen tot nieuwe vormen van dienstverlening. Het afgelopen jaar is er veel aandacht geweest voor de knelpunten van open data. Ook is volop het experiment gezocht in de vorm van bijeenkomsten met zogenaamde 'hackers'<sup>9</sup>. Door deze hackatons, bijeenkomsten waarbij interaction designers en app-bouwers met overheidsdata in de weer gaan, werd duidelijk dat er nog een markt te winnen valt. Als voorbeeld geldt de app Vistory, winnaar van de wedstrijd Apps voor Nederland<sup>10</sup>, waarmee op basis van je locatie getoond wordt welke historische films in de buurt zijn gemaakt. Je kunt zelfs locaties 'geotaggen' en weer koppelen aan de gefilmde beelden.

Veel overheidsorganisaties buigen zich

over een open data strategie. Het doel is daarbij de beschikbaarheid van relevante data te vergroten. Ook het Ministerie van EZ zet in op verdere beschikbaarheid van de eigen datasets. Daarnaast is er het besef dat voor veel data geldt dat het zich niet direct leent om er interessante apps of andere toepassingen mee te bouwen. Zeker voor veel geodata geldt dat het complexe informatie betreft waarbij een zekere mate van expertise nodig is om het te combineren of toe te passen. En soms is zelfs wetenschappelijke kennis nodig om data te interpreteren. Dit geldt bijvoorbeeld voor (meet)gegevens afkomstig van sensoren. Vanuit dit besef is er een start gemaakt met het Open (geo-)Data Innovatie Netwerk (ODIN). Dit netwerk heeft als doel het dataaanbod van de overheid goed zichtbaar te maken en de kennisuitwisseling daarover te stimuleren. Daarnaast is het doel de vraag naar (geo-)data te stimuleren en te vergroten. Vanuit verschillende thema's zal in estafettebijeenkomsten dit verder vorm krijgen. De eerste bijeenkomst richt zich daarbij op Agro-Food in samenwerking met ZLTO (zie ook het laatste artikel: Open Data Estafette Agro & Food).

### **Inventarisatie Ministerie van EZ**

In 2012 is er binnen het Ministerie van EZ (toen nog EL&I geheten) door Dialogic/The Green Land een inventarisatie uitgevoerd naar datasets die interessant kunnen zijn als open data. De inventarisatie kan al aanvulling gezien op de al lopende activiteiten. Zo is bijvoorbeeld in 2012 reeds het Agrarisch Areaal Nederland (AAN), de geografische afbakening van landbouwgrond, beschikbaar. Ook is, in overleg met de sector, besloten informatie over gewaspercenten beschikbaar te stellen. Daar wordt momenteel aan gewerkt.

De inventarisatie is uitgevoerd bij alle aan EZ gerelateerde overheidsorganisaties: Agentschap Telecom, de Consumentenautoriteit, het Centraal Planbureau (CPB), PIANOo, het Staatstoezicht op de Mijnen, de Dienst Landelijk Gebied, de Dienst Regelingen, de Nederlandse Voedsel- en Warenautoriteit (NVWA), Agentschap NL, het Centraal Bureau voor de Statistiek (CBS), de Nederlandse Mededingingsautoriteit (NMa), de Onafhankelijke Post en Telecommunicatie Autoriteit (OPTA) en Staatsbosbeheer. Uiteraard beschikken niet al deze organisaties op voor het agro-domein interessante gegevens.

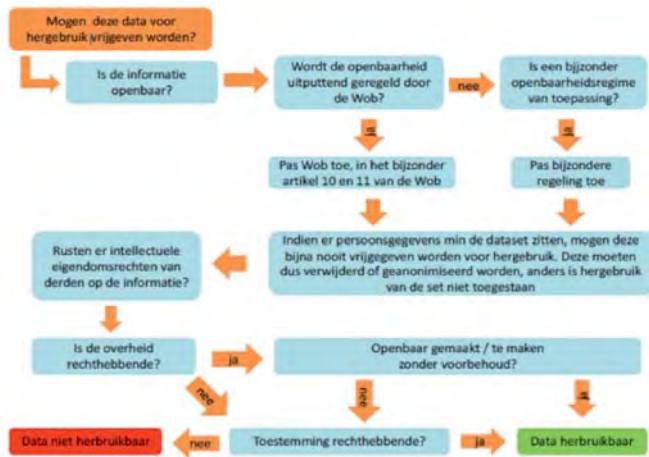
Bij het uitvoeren van het onderzoek is op basis van interviews gezocht naar de 'top 50' van meest kansrijke en representatieve datasets. Het belangrijkste

7 Zie [http://www.forumstandaardisatie.nl/fileadmin/os/documenten/Rapportage\\_open\\_data\\_en\\_aansprakelijkheid\\_def.pdf](http://www.forumstandaardisatie.nl/fileadmin/os/documenten/Rapportage_open_data_en_aansprakelijkheid_def.pdf)

8 Zie <http://vimeo.com/36752317>

9 Zie bv. <http://www.hackdeoverheid.nl/>

10 Zie <http://nationaleappprijs.nl/>



Afbeelding 2. Overzicht juridische hergebruikstoets (bron: Dialogic – The Green Land)

criterium was daarbij de economische hergebruikswaarde. Op deze data is een juridische test uitgevoerd om vast te stellen of de data ook echt openbaar en vrij van rechten van derden zijn. Daarbij is gebruikgemaakt van het onderstaande schema.

Bij de beoordeling zijn de volgende vragen gesteld:

- Welk openbaarheidsregime is van toepassing op de informatie?
- Zijn er wettelijke gronden die de informatie niet openbaar maken?
- In het bijzonder: zitten er persoonsgegevens in de dataset?
- Spelen er rechten van derden?

Uit de inventarisatie blijkt dat er interessante (potentieel) open datasets op het gebied van landbouw, natuur en voedsel zijn. Zodra het rapport is afgerond, zal deze informatie ook beschikbaar zijn. Het toegankelijk maken van deze data is nog onderdeel van besluitvorming, maar de behoefte in de markt heeft daarbij wel een bepalende rol. Kortom, ook het Ministerie van EZ neemt open data serieus. Het Ministerie wil daarbij aanspreekbaar zijn voor ondernemers met goede ideeën. Maar daarbij hoort ook een goede afweging tussen de verschillende belangen die spelen in het op de juiste wijze beschikbaar stellen van data. Het verzoek voor 'raw data now!' beantwoordt de overheid door er serieus naar te kijken en verschillende initiatieven te ontplooiën.

## Casusbeschrijving

### Open data voor de Zaaizaadsector in Nederland

Nederlandse zaaizaden staan bekend om hun hoge kwaliteit en worden om die reden geëxporteerd over de hele wereld. Nederland is na de VS de nummer twee in de wereld met een jaarlijkse omzet van ongeveer 2 miljard.

De productie van zaaizaden vindt voornamelijk buiten Nederland plaats. Zo zijn bij de veredeling van tomatenzaad soms acht verschillende landen betrokken, elk met hun specifieke importeisen. Bij zowel de productie als de afzet is het daarmee van groot belang te weten wat, en onder welke voorwaarden mag worden geëxporteerd. De planning van de productie en de afzet van zaaizaden is dus zeer complex. Elke exporteur houdt daarom een eigen administratie bij als ondersteuning voor de planning van productie en afzet. In 2009 werden de Nederlandse exportbedrijven geconfronteerd met verscherpte reglementen rond de export. Met name de export van zaaizaden naar landen buiten de EU werden door deze verscherping geraakt. De verscherping maakte het plannen van de productie en de export nog complexer.

### Open data Client Export

De zaaizaadsector maakt al twee jaar intensief gebruik van CLIENT Export<sup>1</sup> voor de export buiten de EU. Dit overheidssysteem bevat o.a. informatie over de exporteisen voor alle zaaizaden naar vrijwel elk land in de wereld. Deze informatie is van belang om het (fytosanitaire) exportcertificaat op te maken en te controleren. De sector heeft, onder andere naar aanleiding van de verscherpte maatregelen, de overheid gevraagd de informatie uit het CLIENT-Export voor hergebruik beschikbaar te stellen. Hier wordt momenteel aan gewerkt in een pilot met twee bedrijven. Zodra dit is afgerond is de data ook als open data beschikbaar.

De verwachting is dat door het vrijgeven van de data voor de sector de grootste voordelen liggen in het snel verkrijgen van inzicht in optimale productielocaties. Ook is er direct inzicht in de verkoopbaar-

heid van voorraden en zijn minder complicaties bij export te verwachten (denk aan vernietiging van partijen zaaizaad en tijdverlies door exportbelemmeringen op locatie). De financiële baten zijn daarbij moeilijk precies aan te geven, maar het minimaliseren van onverkooptbare voorraden scheelt veel (waarschijnlijk tonnen) omzet. Ook wordt veel geld bespaard door optimalisatie van de planning.

Door de data van Client Export als open data beschikbaar te stellen kan Nederland zijn positie als exporteur van zaaizaden verder uitbouwen. Bovendien liggen er kansen voor uitbreiding naar andere sectoren.

### Techniek

De informatie blijkt in de praktijk, door de grote hoeveelheid data, moeilijk direct bruikbaar. Daarom is er een webservice ontwikkeld om de inhoud van alle tabellen aan te bieden. Ook het ontwerp van de database zelf is vrijgegeven. Hierdoor kan het bedrijfsleven zelf applicaties maken die konden worden ingepast in het eigen bedrijfsmanagementsysteem. Samen met de sector zijn afspraken gemaakt over de actualiteit en beschikbaarheid. De data wordt een maal per 24 uur geactualiseerd.

In het onderstaand schema wordt deze informatieketen gevisualiseerd.

Voor vragen: John van Blaricum (john.van.blaricum@vwa.nl) of Willem Jongkind (w.j.jongkind@minlnv.nl)

