



Onderzoeksinstituut OTB



## Open data and beyond

Exploring existing open data projects to prepare a successful open data strategy  
Deelrapport Techniek



## Deelrapport Techniek

Dit onderzoek is uitgevoerd in opdracht van: NGInfra/ Alliander

Auteurs: *Drs. M.E. de Vries & Prof. dr. ir. P.J.M. van Oosterom*

Betrokken in de, cq. lid van de Next Generations  
programmaraad:  
Drs. Pallas Agterberg  
Dr.ir. Harry van Breen MBE  
Dr. Judith Schueler  
Prof.dr.ir. Margot Weijnen (vz.)

Betrokken bij het tot stand komen van de  
inhoud van dit rapport:  
Ing. Jan Bruinenberg  
Drs. Leen van Doorn  
Ir. Paul Juffermans  
Ing. Tanju Özel  
Dr. Ir. Marisca Zweistra  
Mr. J. van Kraaij  
Mr. Loek Muijtens

Maart 2012

Onderzoeksinstituut OTB  
Technische Universiteit Delft  
Jaffalaan 9, 2628 BX Delft  
Tel. (015) 278 30 05  
Fax (015) 278 44 22  
E-mail [mailbox@otb.tudelft.nl](mailto:mailbox@otb.tudelft.nl)  
<http://www.otb.tudelft.nl>



Dit werk is gelicenseerd onder een Creative Commons Naamsvermelding-NietCommercieel 3.0 Nederland. Bezoek <http://creativecommons.org/licenses/by-nc/3.0/nl/> om een kopie te zien van de licentie of stuur een brief naar Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

## Samenvatting

Dit deelrapport beschrijft een aantal technische aspecten van open data beleid. Ingegaan wordt op: data preparatie, leveringsmodel, web techniek, open standaarden, begrijpbaarheid, vindbaarheid en gebruikersinteractie. Per aspect (kernthema) wordt een aantal keuze-mogelijkheden voor de implementatie van open data beleid gepresenteerd. Een van de conclusies is dat er niet slechts één implementatie-strategie of één technologie uitrolt als meest optimaal. Er zijn meerdere combinaties tussen de technische aspecten mogelijk, keus voor een optie leidt niet automatisch tot een bepaalde keuze binnen een ander technisch aspect. In hoofdstuk 4 worden tenslotte een aantal punten genoemd en adviezen gegeven voor het vervolgtraject.

# Inhoudsopgave

<b>Samenvatting .....</b>	<b>2</b>
<b>1 Inleiding .....</b>	<b>3</b>
1.1 Doel en afbakening .....	3
1.2 Open data – technische aspecten.....	3
1.3 Kernthemas.....	3
1.4 Afhankelijkheden .....	4
<b>2 Kernthemas .....</b>	<b>5</b>
2.1 Data preparatie.....	5
2.1.1 Filteren .....	5
2.1.2 Aggregeren.....	5
2.1.3 Opsplitsen / anders indelen.....	6
2.1.4 Verrijken.....	6
2.2 Leveringsmodel .....	6
2.2.1 Kant en klare bestanden .....	7
2.2.2 Selectie door gebruiker via online aanvraag .....	7
2.2.3 Selectie door gebruiker via interactieve web applicatie.....	8
2.3 Web techniek .....	8
2.3.1 Als bestand op web site zetten .....	9
2.3.2 Via web service: geo-specifiek.....	9
2.3.3 Interactieve web applicatie .....	9
2.4 Open standaarden .....	10
2.4.1 Bestandsformaten .....	11
2.4.2 Web service protocollen.....	12
2.4.3 Structuur en semantiek: gegevensmodellen .....	13
2.4.4 Model transformatie .....	14
2.5 Begrijpbaarheid .....	14
2.6 Vindbaarheid .....	16
2.6.1 Eigen web site.....	16
2.6.2 (Centrale) portal, geo-register .....	16
2.7 Gebruikersparticipatie / interactie.....	18
<b>3 Internationale standaarden .....</b>	<b>19</b>
3.1 Energie sector .....	19
3.2 Geo wereld.....	19
3.2.1 Open GIS (OGC) en ISO TC211 .....	19
3.2.2 INSPIRE.....	20
3.3 W3C en linked open data .....	21
<b>4 Conclusies en vooruitblik .....</b>	<b>23</b>
4.1 Afweging kernthema's.....	23
4.2 Alliander en internationale standaarden.....	23
4.3 Vervolgstappen t.b.v. opzetten technische infrastructuur .....	24
<b>Literatuur.....</b>	<b>26</b>
<b>Bijlage A: INSPIRE - Utility Services / Networks.....</b>	<b>29</b>
<b>Bijlage B: GML encoding – voorbeeld.....</b>	<b>31</b>



# 1 Inleiding

## 1.1 Doel en afbakening

Doel van dit deelrapport is om een overzicht te geven van de technische aspecten die spelen bij een open data beleid. Op welke manier (met welke technologie) moet open data beschikbaar worden gesteld, zodat het ook echt 'open' is? Welke implementatie-keuzes zijn er, en wat is de samenhang met zowel de juridische als de organisatorische aspecten?

Mede naar aanleiding van de opmerkingen die de Programmaraad van NGInfra meegaf bij het toekennen van dit project, wordt in dit rapport ook gekeken naar internationale standaarden en ontwikkelingen die van belang zijn voor data uitwisseling in het algemeen en open data beleid in het bijzonder.

Het rapport is als volgt ingedeeld: in dit inleidend hoofdstuk worden de technische aspecten benoemd, en ingedeeld in zeven kernthemas. Tevens worden de afhankelijkheden met de juridische en organisatorische aspecten aangegeven. In hoofdstuk 2 worden de zeven kern-aspecten verder uitgewerkt. Vanwege het belang van internationale standaardisatie ontwikkelingen voor open data beleid wordt in hoofdstuk 3 een en ander gezegd over relevante standaardisatie initiatieven, zowel binnen de wereld als daarbuiten.

In het concluderende hoofdstuk 4 wordt een afweging gemaakt van prioriteiten, en wordt vooruit gekeken naar de stappen die voor het opzetten van een open data (web) architectuur het eerst nodig zijn.

## 1.2 Open data – technische aspecten

Wanneer is data in technische zin 'open'? Als criteria worden in discussies op blogs en web fora een of meer van de volgende punten genoemd: gebruik van open bestandsformaten, het aanbieden van de ruwe data, gemakkelijk te begrijpen en te vinden, goed toegankelijk (in de zin van bereikbaar, downloadbaar) en 'machine readable' (waarmee documenten, met ongestructureerde tekst, afvallen als 'open data').

Gebruiksgemak en toegankelijkheid zijn zaken die niet zo makkelijk te operationaliseren zijn. Element hierin is in ieder geval: het makkelijk en snel (zonder wachttijd) toegang hebben tot de data, als downloadbaar bestand of als output van een web service, en het makkelijk kunnen gebruiken in eigen projecten, in combinatie met andere datasets. Ook gebruikersparticipatie en continuïteit van de relatie tussen data gebruiker en data aanbieder kan hierbij een rol spelen: op de hoogte gehouden worden van updates, het kunnen discussieren op een gebruikersforum, en de mogelijkheid hebben tot het geven van feedback. Die feedback kan ook inhouden het doorgeven van geconstateerde fouten of omissies in de data, of zelfs het direkt editen van de data. Gebruikers van open data kunnen daarmee actief bijdragen aan het controleren en zo nodig verbeteren van de kwaliteit van de brondata.

## 1.3 Kernthemas

Criteria en aspecten die hierboven zijn genoemd kunnen als volgt worden gegroepeerd in een aantal kern-themas:

1. data preparatie, met als keuze: de ruwe data beschikbaar stellen, of de data (gedeeltelijk) voorbereiden;
2. leveringsmodel, met als opties: kant en klare bestanden, selectie door gebruiker via web service en directe download, of selectie via een online form;
3. web techniek: bestanden op de website of via een portal, losse web services (download, view, transformation, feedback/update services), en/of een complete interactieve web applicatie
4. open standaarden (m.b.t. bestandsformaat, gegevensmodel, web service protocol)
5. makkelijk te begrijpen: ondersteuning door documentatie, goede metadata / duidelijke inhoud / begrijpbaarheid
6. makkelijk te vinden: machine-readable metadata, portals, registers / vindbaarheid
7. relatie met gebruiker (feedback organiseren, gebruik monitoren) / gebruikersinteractie

Met name het laatste punt (gebruikersinteractie) ligt op het grensvlak van technische en organisatorische aandachtspunten (zie ook Bregt et al., 2012).

#### 1.4 Afhankelijkheden

Aan de kant van de open data aanbieder moeten een aantal implementatie-keuzes gemaakt worden: moet de data worden voorbereid of niet, hoe moet de data worden aangeboden, hoe moet de feedback van gebruikers worden gefaciliteerd, hoe te zorgen voor gebruiksgemak, begrijpbaarheid en vindbaarheid?

Bij een aantal van deze keuzes zijn er afhankelijkheden met andere aspecten, zowel intern (met andere technische aspecten) als extern (met juridische en organisatorische aspecten).

Vooraf voor de aspecten 'data preparatie' en 'leveringsmodel' plus 'web techniek' zijn er afhankelijkheden met juridische respectievelijk organisatorische aspecten.

Voor wat betreft data preparatie: de mate waarin datasets moeten worden voorbereid (gefilterd en/of geaggregeerd) is sterk afhankelijk van randvoorwaarden die voortkomen uit privacy en veiligheid.

Voor wat betreft leveringsmodel en web techniek: de keuzes die daar gemaakt kunnen worden zijn mede afhankelijk van de hoeveelheid resources (geld, mensen, skills) die beschikbaar zijn voor de daadwerkelijke implementatie van het open data beleid (en het waarborgen van de continuïteit daarvan).

Het zelfde geldt in mindere mate ook voor de aspecten 'vindbaarheid' en 'gebruikersinteractie'. Naarmate de organisatie meer waarde hecht aan de rol van open data, ook voor de 'interne' informatievoorziening en kwaliteit van gegevens, zal het belang van het faciliteren van gebruikersfeedback ook toenemen.

## 2 Kernthemas

### 2.1 Data preparatie

Een belangrijke vraag bij het verstrekken van datasets als open data is of men de data wil aanbieden zonder voorbereiding (de ruwe data, gelijk of vrijwel gelijk aan de gegevens zoals opgeslagen in de eigen databases), of dat men de data eerst bewerkt voordat ze wordt verstrekt.

Wat verstaan we onder data preparatie:

- Filteren: weglaten van gegevens (alleen bepaalde attributen, of hele records), vanwege privacy, veiligheid of om andere redenen;
- Aggregeren: in plaats van de individuele gegevens (per adres, per punt-lokatie) worden gegevens tot statistisch zinvolle eenheden geaggregeerd, bijv. per postcodegebied of per administratieve eenheid (gemeente, CBS wijk of buurt);
- Opsplitsen van heterogene datasets in meer homogene subsets, bijvoorbeeld gas- en elektriciteits leveringen in aparte datasets aanbieden, grootverbruik en kleinverbruik apart aanbieden, etc.;
- Verrijken: toevoegen van gegevens uit andere datasets, of uit eigen datasets, of uit datasets van andere organisaties.

Doel van het vooraf bewerken van de data kan zijn: anonimiseren (vanwege juridische randvoorwaarden), maar ook bijv. om duidelijkheid en doelmatigheid te vergroten (er kunnen attributen zijn die nooit gevuld zijn of juist altijd dezelfde waarde hebben). In het laatste geval is er dus ook – als spin-off – sprake van kwaliteitsverbetering van de gegevens.

#### 2.1.1 Filteren

Filteren kan als doel hebben het waarborgen van privacy of veiligheid, maar ook de kwaliteitsverbetering die al genoemd is. Filteren kan echter zijn doel voorbij schieten als het betekent dat de data leverancier zelf bepaald welke gegevens wel of niet relevant zijn om als open data te publiceren. Daarmee wordt onvoorzien en verrassend gebruik van open data door derden, die applicaties ermee bouwen die niemand voorzien heeft, minder.

#### 2.1.2 Aggregeren

Aggregeren kan soms nodig zijn omdat anders niet aan randvoorwaarden rond privacy en veiligheid voldaan kan worden.

Indien Alliander zelf aggregeert voordat de data opengesteld wordt, dan worden – noodgedwongen - keuzes gemaakt die de latere gebruiksmogelijkheden zullen beïnvloeden, met name door de keuze van de aggregatie eenheid.

Sommige informatie is alleen in combinatie, op individueel niveau, relevant voor statistieken. Bijvoorbeeld het aantal buitengebruik gestelde aansluitingen in een wijk is interessanter als ook de reden van buitengebruikstelling wordt meegenomen in de statistieken. Maar als daardoor er minder dan 6 individuele waarnemingen overblijven (per reden buitengebruikstelling), kan dit gegeven voor die wijk niet geleverd worden.



Als Alliander zelf aggregereert, dan verdient het aanbeveling om niet alleen per postcode gebied, maar ook per CBS-wijk/buurt, of andere NUTS-regio te aggregeren, zodat gebruikers ook kunnen combineren met CBS en Europese statistieken.

Wat is een goede aggregatie-eenheid? Het 'Buurt' niveau uit de Wijk/Buurt indeling van het CBS is groter van geografische omvang dan het 6-positie postcodegebied. Dit betekent dat het minder vaak zal voorkomen dat het aantal individuele waarnemingen te klein is voor publicatie van die gegevens (als de 'minder dan 6' regel wordt toegepast). Tweede voordeel is dat op Buurt-niveau wel met allerlei statistische gegevens (van het CBS) gecombineerd kan worden, wat niet kan als geaggregeerd wordt volgens de postcodegebieden (de grenzen daarvan lopen over buurt- en wijkgrenzen heen).

<b>Aggregatie-eenheid</b>	<b>Aantal ervan (binnen NL)</b>	<b>Indeling</b>
Provincie	12	NUTS nivo 2
Gemeente	418	CBS
3-positie postcodegebied	798	
Woonplaats	2457	
Wijk	2575	CBS
4-positie postcodegebied	4035	
Buurt	11778	CBS
6-positie postcodegebied	> 20000	

### 2.1.3 Opsplitsen / anders indelen

Opsplitsen van een dataset in meerder kleinere datasets kan zinvol zijn wanneer anders een te heterogene set van gegevens blijft bestaan.

Het splitsen van leveringen van elektriciteit en terug-leveringen lijkt logisch omdat het om iets essentieel anders gaat (productie i.p.v. alleen consumptie van energie door huishoudens). Ook het opsplitsen naar type gebruik (groot- en kleinverbruik) kan zinvol zijn.

Er schuilt echter ook een nadeel in het opsplitsen. Juist daardoor kunnen snel aggregaties ontstaan van minder dan 6 waarnemingen, dit is al het geval bij de terugleveringen (slechts 1 of 2 per postcodegebied). En dat kunnen juist erg interessante statistieken zijn, ook om veranderingen in de loop van de tijd te zien.

### 2.1.4 Verrijken

Terwijl bij filteren en aggregeren er sprake is van inperken van de gegevens die worden verstrekt naar buiten, kan het ook zinvol zijn om de eigen informatie juist uit te breiden / te verrijken door al zelf te combineren met andere data (van andere organisaties). Dit kan alleen in goed overleg met die organisaties, zodat geen problemen achteraf ontstaan. Als voorbeeld: Alliander voegt zelf al de geometrie van de 4 positie postcode gebieden toe aan de dataset met aansluitingsgegevens die geaggregeerd zijn per 4 positie postcode, zodat gebruikers die de data downloaden niet zelf op zoek hoeven te gaan naar die 4 positie postcode gebieden.

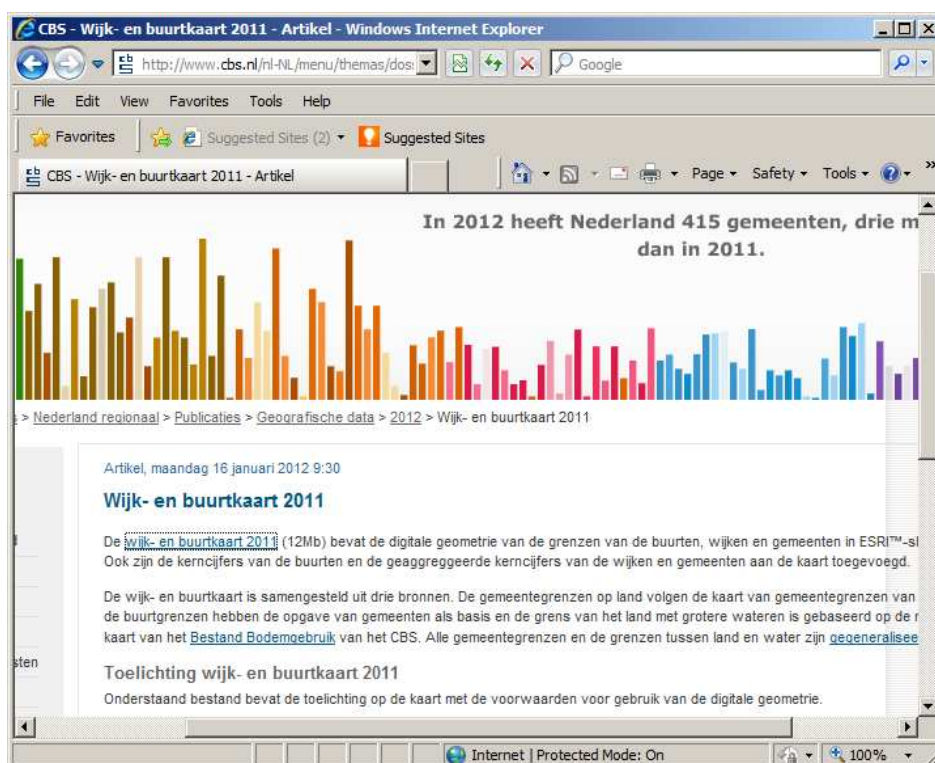
## 2.2 Leveringsmodel

De manier waarop open data technisch wordt aangeboden kan op meerdere manieren worden getypeerd. In deze paragraaf gaat het om twee vragen: kan de gebruiker zelf (beperkt of onbeperkt) kie-

zen welke data precies wordt gedownload, en, ten tweede, kan de data direkt worden gedownload of moet de gebruiker wachten, bijv. omdat het verzoek om data wordt afgehandeld door een helpdesk.

### 2.2.1 Kant en klare bestanden

Dit is een vrij gangbare manier van data aanbieden: de data staat (al dan niet in een zipfile) klaar, of op de eigen web site, of het bestand kan worden gevonden via Google, of via een web portal, een catalog service of een INSPIRE Discovery Service (zie sectie 3.2). Een gebruiker kan in deze optie geen subset selecteren, en kan ook niet eerst de data bekijken in een web viewer. Anderzijds, een voordeel is, dat een simpele download-actie volstaat, en men direkt over het bestand kan beschikken. Hier is sprake van een combinatie van 'geen selectie', 'geen wachttijd'.



**Figuur 2-1 CBS: download van bestanden, geen selectie van subsets**

Als voorbeeld: o.a. het CBS biedt een aantal datasets op deze manier aan, zie figuur 2-1. We noemen dit in het vervolg 'leveringsmodel A'.

### 2.2.2 Selectie door gebruiker via online aanvraag

Een tweede manier van data aanbieden is via een online form waarmee gebruikers een bestand of een deel van een bestand kunnen aanvragen, waarna de aanvraag bijv. binnen 24 uur leidt tot een mail aan de aanvrager met daarin de url waar de (deel)bestanden gedownload kunnen worden.

Het moeten aanvragen, met email adres en eventueel andere gegevens, kan een barriere voor geïnteresseerden. Anderzijds, hierbij heeft de data leverancier wel zicht op wat aan wie geleverd wordt, en dat kan voor monitoring doeleinden nuttig zijn (zie verder Van Loenen en Bregt, 2012).

In figuur 2.2 als voorbeeld de web pagina met selectie scherm van de Ierse Geological Survey.

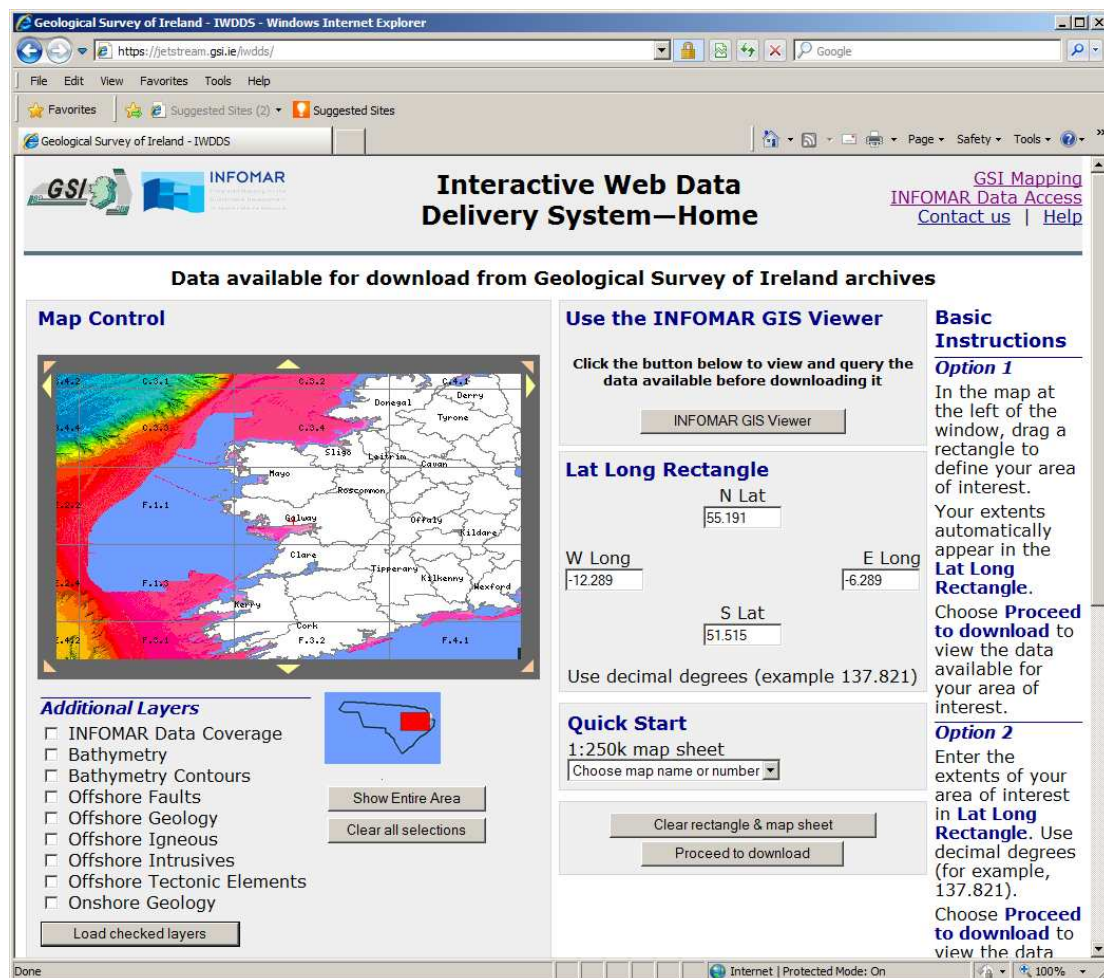


Figure 2-2 Selectie scherm; url van (deel-)bestanden wordt later gemaid

De mate waarin gebruikers kunnen selecteren via de online aanvraag kan variëren van: alleen een ruimtelijke indicatie (trekken van een rechthoek op de kaart, of kiezen van bijv. een provincie of gemeente uit een drop down list), of ook selectie op andere attributen: bijvoorbeeld alleen de gasleveringen, of alleen de storingen in een bepaalde maand, of van een bepaald type. Hier is sprake van een combinatie van '(zekere mate van) selectie door de gebruiker', 'wel wachttijd'.

### 2.2.3 Selectie door gebruiker via interactieve web applicatie

Een selectie interface zodat gebruikers subsets kunnen aangeven (temporeel, geografisch, thematisch), en alleen die data downloaden.

Zo'n selectie zou via een kaartje kunnen: door het aangeven van een zoom/clip rechthoek met de muis, of door het kiezen van een gemeente, wijk/buurt of postcode uit een keuzelijstje. Ook meer geavanceerde selecties zijn mogelijk: op jaar, op type energie (electriciteit, gas, teruglevering electriciteit), of type verstoring of schade etc.

Wanneer sprake is van een selectie mogelijkheid spreken we in dit rapport van 'leveringsmodel B'.

## 2.3 Web techniek

Dit aspect overlapt voor een deel met het aspect 'leveringsmodel', maar bij 'web techniek' gaat het in de eerste plaats om de vraag wat de onderliggende technologie kan zijn bij het aanbieden van open data, terwijl het bij het aspect 'leveringsmodel' gaat om de vraag hoeveel invloed een gebruiker heeft

op de inhoud en andere kenmerken (bijv. bestandsformaat of gegevensmodel) van de te downloaden data.

We onderscheiden de volgende vier manieren op open data beschikbaar te stellen: via het downloaden van bestanden, via het aanroepen van web services, door het opzetten van een interactieve web applicatie, en door het aanbieden van de data als linked open data.

Verschil tussen een web service en web applicatie is dat een web service een bepaalde 'losse' data output heeft, en vanuit verschillende web of niet-web client omgevingen kan worden aangeroepen. Een web applicatie daarentegen is een combinatie van een client omgeving die communiceert met een of meer (web) services. De organisatie die de open data levert, doet dat dus via een eigen web client.

### 2.3.1 Als bestand op web site zetten

Data gebruikers kunnen de bestanden met een simpele muisklik downloaden, de data aanbieder hoeft alleen te zorgen voor het periodiek verversen van het bestand op de web site.

### 2.3.2 Via web service: geo-specifiek

Voor geografische informatie (= informatie die op een of andere manier aan een plek op aarde gerelateerd kan worden) zijn er in principe twee manieren om data op te slaan en uit te wisselen: als raster of als vector data. De meeste kaarten en kaartjes op het Web, zoals bij Google Maps, zijn raster plaatjes. Voor visualisatie is dit vaak ideaal: eenvoudige web clients kunnen gebruikt worden, tenminste als men raster formaten als PNG of JPEG gebruikt als output van de web service. Internet browsers kunnen deze formaten immers visualiseren zonder dat een plug-in of een specifieke tool nodig is. Ook satelliet images vallen in deze categorie.

Om het verschil tussen raster en vector output van web services aan te geven, worden ook de web services voor geo-informatie in twee categorieën ingedeeld: map services (ook wel image services) en feature services (ook wel data services; en in INSPIRE verband: Download Services, zie sectie 3.2).

Als men open data wil aanbieden waar een gebruiker verder mee kan werken (analyseren, combineren met andere data, verrijken), zijn data services bijna altijd de enige mogelijkheid ('bijna altijd', want satellietbeelden kunnen op zich wel als input dienen voor allerlei soorten analyse).

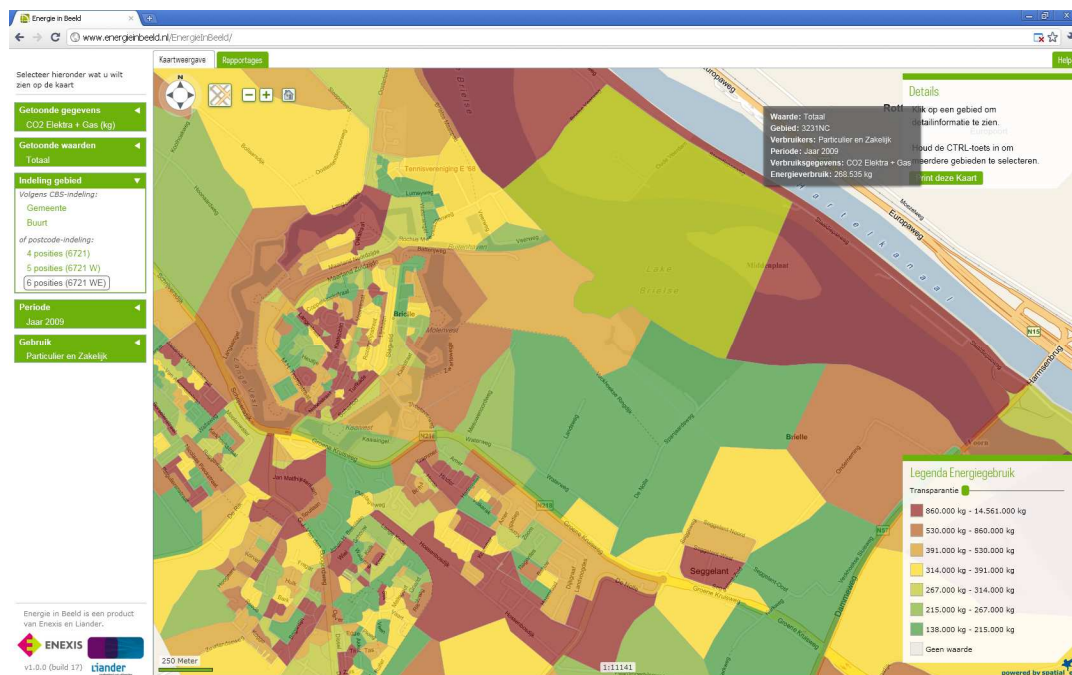
Hieronder een voorbeeld van de aanroep van zo'n geodata web service, in dit geval een web service van Rijkswaterstaat met als output GML (zie verder sectie 2.4):

[http://geoservices.rijkswaterstaat.nl/noordzee\\_wingebieden?SERVICE=WFS&version=1.1.0&request=GetFeature&maxfeatures=50&typename=rws:AAA200,rws:AAA300,rws:AAA100&srsName=EPSG:28992&output=GML3](http://geoservices.rijkswaterstaat.nl/noordzee_wingebieden?SERVICE=WFS&version=1.1.0&request=GetFeature&maxfeatures=50&typename=rws:AAA200,rws:AAA300,rws:AAA100&srsName=EPSG:28992&output=GML3)

### 2.3.3 Interactieve web applicatie

De data leverancier kan ook besluiten om zelf een web applicatie aan te bieden (server en client kant). De mate waarin de gebruiker zelf kan kiezen wat hij of zij wil zien en wil downloaden kan variëren, er zijn meerdere gradaties in functionaliteit en interactiviteit die wordt geboden.

Een goed voorbeeld van een web applicatie waar een gebruiker online de data kan bekijken, en tevens de data (als bestand) kan downloaden is de Energie in Beeld applicatie, zie figuur 2-3.



**Figure 2-3 Energie in Beeld: web applicatie met data download mogelijkheid**

Wat aan bijv. Energie in Beeld zou kunnen worden toegevoegd is de mogelijkheid om dynamisch andere kaartlagen, met data van andere open data aanbieders, toe te voegen aan de kaartjes. Op die manier wordt de toegevoegde waarde van de web applicatie nog vergroot ('mashups').

## 2.4 Open standaarden

Het gebruik van open bestandsformaten bij verstrekking van open data geldt als een van de criteria voor het 'open' zijn van de data (zie Van Loenen, 2012). Naast een open bestandsformaat zijn er twee andere onderdelen waarop standaardisatie essentieel is: standaarden op het gebied van web services, en het hebben van 'standaard' (de-facto of de-jure) sector-brede informatiemodellen, die bovendien aansluiten bij gegevensmodellen van andere sectoren, waarmee een afhankelijkheid bestaat.

Een 'open' standaard wordt in openheid voorgesteld, bediscussieerd, vastgesteld en indien nodig gewijzigd, en is niet van bijv. een bepaalde software producent, maar van een standaardisatie-gremium, zoals W3C of ISO.

Het belang van open standaarden in de context van open data heeft te maken met het voor iedereen toegankelijk en bruikbaar moeten zijn van open data: vermeden moet worden dat gebruikers gebonden zijn aan bepaalde (niet-gratis) tools om de 'open' data te kunnen bekijken en gebruiken.

Overigens is het ook o.a. om die reden dat in de EU en in Nederland overheden aandringen op het gebruik van open standaarden en open source software: om te voorkomen dat men aan de gebruikerskant aan bepaalde software en dus bepaalde software leveranciers vastzit, met alle gevolgen voor monopolisme van dien.

Een belangrijke standaardisatie organisatie en discussieplatform in de geo-wereld is het Open Geospatial Consortium (OGC), begonnen in 1994 als OpenGIS Consortium. Men houdt zich in OGC context met allerlei ruimtelijke informatie en geo-informatiesystemen bezig, niet alleen met GIS, maar ook met CAD data en systemen, mobiele toepassingen (routeplanners), sensor gegevens (van bijvoorbeeld

verkeersdrukte of luchtvervuiling), remote sensing beelden, luchtfoto's, 3D modellen van steden, en sinds kort ook met smart grid ontwikkelingen en standaarden die daarvoor nodig zijn.

Binnen het OGC Consortium wordt gewerkt aan verschillende typen standaarden, waarvan er voor open data met name de volgende twee relevant zijn:

- data encoding standaarden (fysieke formaat en het generieke data model waarop het is gebaseerd) (sectie 2.4.1);
- web service interface standaarden (die de rol hebben van 'logische' web protocollen), voor het publiceren en processen van geo-informatie via het web (sectie 2.4.2).

OGC standaarden zijn soms een implementatie van abstractere ISO standaarden, en zijn in een aantal gevallen zelf ook weer naar ISO standaarden omgezet (van de ISO TC211 serie: TC211 is de Technical Committee voor geo-informatie). Op hun beurt zijn deze ISO TC211 standaarden weer de basis voor veel van de INSPIRE standaarden en regelgeving.

### 2.4.1 Bestandsformaten

De toegankelijkheid van open data hangt direkt samen met het fysieke bestandsformaat waarin de data beschikbaar wordt gesteld. Dit dient in ieder geval een gangbaar formaat te zijn, bij voorkeur gebaseerd op een open standaard, om de redenen die hierboven al genoemd zijn: men is niet gebonden aan een bepaald software produkt om de data te kunnen inlezen en gebruiken.

Voor geo-informatie bestaan een groot aantal niet-open formaten; voor vector data is zeer gangbaar de ESRI shapefile. Een open standaard is OGC's GML (Geography Markup Language), met name ontwikkeld voor Web toepassingen. Niet specifiek voor geo-informatie is er RDF (Resource Description Framework).

#### **GML (Geography Markup Language)**

Volgens de INSPIRE richtlijnen is GML de default encoding voor geodata uitwisseling in geval van vector data.

GML heeft XML (eXtensible Markup Language) als bestandsformaat (encoding). Dit heeft een aantal voordelen, onder andere het feit dat ook tools en software van buiten de geo-ICT gebruikt kunnen worden om GML te produceren (exporteren) of in te lezen en te visualiseren. Een ander belangrijk inhoudelijke voordeel is dat er geen informatieverlies hoeft op te treden als GML gebruikt wordt als uitwisselingsformaat tussen geo-informatiesystemen. XML is namelijk een zeer flexibele manier om informatie te coderen. Het conceptuele gegevensmodel waarop GML is gebaseerd, is afgeleid van een aantal abstracte ISO standaarden, bijv. ISO 19107. Hierin worden o.a. de geometrische primitieven voor geo-informatie gedefinieerd.

GML heeft nu versie 3.2, en is tevens een ISO standaard, namelijk ISO 19136. GML was in eerste instantie bedoeld als output van de OGC Web Feature Service (zie sectie 2.4.2), maar kan net zo goed gebruikt worden voor data levering in de vorm van bestanden. In Bijlage B is een GML fragment opgenomen.

#### **RDF (Resource Description Framework)**

Buiten de geo-wereld moet er wat betreft bestandsformaten en ontwikkelingen op het web vanzelfsprekend gekeken worden naar wat er in de context van het W3C gebeurt. De laatste paar jaar staat in de belangstelling het via het web ontsluiten van informatie die als RDF is gecodeerd (Resource Description Framework). RDF is zeer geschikt om feiten over individuele objecten (of generieker: individuele 'dingen') op te slaan, volgens een eenvoudig principe: subject, predikaat, object. Predikaat is vergelijkbaar met een attribuut of een associatie in een database. Subject en vaak ook het object

hebben een identiteit, een unieke ID. Die ID kan dan weer gebruikt worden in verwijzingen. Van RDF als taal voor het 'semantisch web', was de stap naar het 'linked data web' snel gemaakt. Via de ID's, die tevens web adressen zijn (URI's) kan een bezoeker van het ene naar het ander RDF 'triple' klikken. Op die manier is er een 'Linked Open Data' web van informatie.

RDF wordt bijvoorbeeld gebruikt als formaat door GeoNames (zie <http://www.geonames.org>) en in een experimentele set-up door de Ordnance Survey (de Engelse tegenhanger van het Kadaster / Topografische Dienst in Nederland), zie <http://data.ordnancesurvey.co.uk/html>

Behalve de twee genoemde open opslagformaten (GML en RDF), kan nog genoemd worden: GeoJSON (de data wordt verpakt als JavaScript Objects) en voor verstrekking als bestand ook het aloude CSV (Comma Separated Values) formaat.

Een flexibele oplossing is om open data in meerdere bestands/output formaten aan te bieden, waarbij naast de de-jure open formaten ook gangbare (en daarmee de-facto standaard) formaten kunnen worden gebruikt. Een gangbaar bestandsformaat voor GIS data is bijvoorbeeld: het shapefile formaat van GIS software maker ESRI. Hoewel dit geen open formaat is, is het wel veel gebruikt. Omdat er ook conversie software is (open source) die shapefiles kan omzetten naar bijv. GML, zijn shapefiles een mogelijkheid.

## 2.4.2 Web service protocollen

Open data kan als bestand of set van bestanden beschikbaar worden gemaakt, maar ook als output van web services. In dat geval is niet alleen het outputformaat van belang (voldoet dat aan de 'open standaarden'-regel van open data?), maar ook de manier waarop de web service wordt aangeroepen via een url.

Ook hiervoor zijn standaarden gemaakt door OGC, gedeeltelijk overgenomen door ISO en INSPIRE.

De belangrijkste zijn:

- de Web Map Service interface standaard (WMS) / ISO 19128
- de Web Feature Service interface standaard (WFS) / ISO 19142
- de Filter Encoding standaard (FE) / ISO 19143
- de Catalog Service for the Web standaard (CSW)
- voor 3D data (in 2011/2012 nog een discussiestuk): de Web 3D Service (W3DS) interface
- Sensor Web Enablement (SWE): aantal (concept) standaarden in ontwikkeling
- Location Based Services (LBS)

Het doel van deze web service standaarden van OGC is ervoor te zorgen dat web server software en web client software van verschillende software producenten met elkaar kan communiceren (interoperabiliteit als kerndoel van OGC).

Hoe dit kan worden gerealiseerd is in de eerste plaats door de manier waarop de web service moet worden aangeroepen en bevraagd in een zogenaamde 'interface specificatie' vast te leggen. In de specificatie is vastgelegd hoe de 'requests' die naar de service gestuurd kunnen worden eruit moeten zien en - andersom – welke 'response' de service terug zal (= mag of moet) geven. En in de tweede plaats door de web service 'self describing' te laten zijn (door de zogenaamde GetCapabilities operatie die de meeste OGC web services hebben): de web service kan bevraagd worden 'wat heb je te bieden'.

Vooral de OGC Web Map Service (WMS) wordt veelvuldig toegepast: de client stuurt als request het verzoek om een kaartje bestaande uit een aantal kaartlagen (vegetatie, stedelijke gebied, snelwegen bijv.), binnen een bepaald rechthoekig gebied (bounding box). Dit type service komt overeen met de INSPIRE View Service.

Voor open data (de data 'achter' het kaartje) is van veel meer belang de OGC Web Feature Service (WFS): hier is de data zelf de output van de service, default in GML (zie sectie 2.4.1), maar de laatste tijd ook steeds meer in andere formaten, ook GeoJSON, en 'zelfs' het 'niet-open' ESRI shapefile bestandsformaat. Om een selectie aan te kunnen geven in de WFS request moet die selectie gecodeerd worden volgens de OGC Filter Encoding standaard. Deze twee protocollen (WFS en FE) werken dus in combinatie.

Deze web service standaarden zijn op twee manieren te gebruiken: allereerst als losse service, die aangeroepen kan worden vanuit een willekeurige client die de OGC standaarden ondersteunt. En tegelijkertijd door zelf een web applicatie te maken die van diezelfde services gebruikt maakt.

### 2.4.3 Structuur en semantiek: gegevensmodellen

De meeste van de genoemde OGC standaarden (WFS, GML, Filter Encoding) regelen vooral de syntactische interoperabiliteit tussen systemen, plus het geo-deel van de semantische interoperabiliteit (namelijk het werken met dezelfde definities van de geometrische en topologische basistypen). Maar voor de applicatie-specifieke onderdelen van de data inhoud, zijn de OGC/ISO standaarden op zich alleen een basis. Voor data integratie binnen een sector, zoals de energie sector, zijn sector-brede gegevensmodellen nodig. Bovendien is het nodig dat die gegevensmodellen voor een sector aansluiten bij de gegevensmodellen van andere sectoren waarmee wordt samengewerkt, zodat daarmee data uitwisseling mogelijk is. Een voorbeeld van een standaard die relevant is voor veel sectoren in Nederland is het data model van de BAG (Basisregistratie Adressen en Gebouwen): adressen komen immers in veel data verzamelingen voor. Wanneer organisaties onderling gegevens uitwisselen ligt het voor de hand zo veel mogelijk dezelfde standaard gegevensmodellen te gebruiken. Voor het verstrekken van open data geldt hetzelfde.

In Nederland zijn de laatste 10 jaar voor geo-informatie uitwisseling een reeks informatiemodellen ontwikkeld die (binnen Nederland) als standaard gelden of gaan gelden. Voor een overzicht zie:

<http://www.geonovum.nl/geostandaarden/informatiemodellen>.

Voor de energie sector is op internationaal standaardisatie gebied op dit moment onder meer van belang:

- Standaarden rondom smart grids, waaronder bijvoorbeeld het Common Information Model (CIM) (IEC 61970 en IEC 61968) en de IEC 61850 (Substation automation and communication format).
- Zowel binnen als buiten de geo-informatie standaardisatie wereld staan smart grids sterk in de belangstelling. Zo staat er voor de kwartaal-bijeenkomst van het Open Geospatial Consortium in maart 2012 (Austin, USA) een Smart Grid Location Standards ad hoc meeting op de agenda.
- Wat er in EU verband gebeurt in de INSPIRE data specificatie teams: voor 34 thema's, vaak samenvallend met sectoren / applicatie-domeinen, worden gegevensmodellen en andere eisen opgesteld voor de (geo)data die binnen Europa uitgewisseld gaat worden. Voor de energie netwerken sector is met name van belang het data thema 'Utility and governmental services', en daarbinnen het sub-thema 'Utility services / networks' (zie Bijlage A).

Voor de energie sector in Nederland is verder van belang:

- NORA, de Nederlandse Overheid Referentie Architectuur inclusief het Stelsel van basisregistraties bestaande uit:



- Administratieve gegevens (o.a. personen, voertuigen en architecturen voor gemeenten (GEMMA), provincies, waterschappen, E-informatieuitwisseling, StUF e-overheid standaard berichtenverkeer)
- Locatie gebonden gegevens (NEN1878/NEN3610: uitwisselstandaard voor de geoinformatie van de geo-basisregistraties, o.a. die van Adressen en Gebouwen (BAG))

In de context van open data beleid is het om een aantal redenen van belang om aan te sluiten bij standaardisatie van gegevensmodellen: echte data integratie is dan mogelijk, waarbij het niet meer uitmaakt waar en hoe de data is opgeslagen: data structuur, attribuutwaarden, classificaties en definities zijn dan immers bij alle data leveranciers hetzelfde. Dit is voor de gebruikers van de opengestelde data een winstpunt: men kan dan data van Alliander vergelijken met data van andere netbeheerders. Dit komt uiteindelijk ook Alliander weer ten goede. Overleg tussen Alliander en de andere netbeheerders in Nederland over welke (nationale en/of internationale) informatiemodellen men gaat gebruiken bij het openstellen van de data is daarom zeer gewenst.

#### 2.4.4 Model transformatie

Vaak is het niet mogelijk om productiesystemen van data leveranciers (meteen) aan te passen aan nieuwe gegevensmodellen, gebaseerd op die internationale of nationale sector-brede standaarden. Een manier om dan toch data te leveren conform die standaarden is om met 'model transformatie' (ook wel 'schema translation') te werken: bestaande data worden naar het nieuwe model omgezet op basis van te voren gedefinieerde 'schema mappings'.

Er zijn in principe twee 'momenten' om bestaande gegevens van een organisatie om te zetten naar een standaard gegevensmodel: vooraf, voordat de data beschikbaar wordt gesteld, of on-the-fly tijdens het viewen of downloaden door de gebruiker. In dit laatste geval gebeurt dit door een 'mediator' service, of in INSPIRE terminologie: een (model of schema) Transformation service.

### 2.5 Begrijpbaarheid

Open data moet makkelijk te begrijpen zijn: zowel de informatie-inhoud als de data structuur moeten duidelijk zijn voor gebruikers buiten de organisatie, en buiten de sector, inclusief 'gewone' burgers. Een bron voor onduidelijkheid zijn bijvoorbeeld de codes en afkortingen die vaak in databases worden gebruikt, zie figuur 2-4. Voor niet-experts zijn codes als deze niet te begrijpen, waardoor – als er geen duidelijke documentatie bijgeleverd wordt – open data een deel van zijn bruikbaarheid verliest.

Bij het openstellen van data kan men er voor kiezen om afkortingen en codes te vermijden, en waar mogelijk te werken met omschrijvingen die duidelijker zijn.

De begrijpbaarheid van open data kan daarnaast worden vergroot door het leveren van documentatie en metadata, ook al zullen niet alle 'gewone' gebruikers die eerst bestuderen. Het beschrijven van datasets gebeurt binnen een organisatie ook, bijv. in de vorm van een gegevenswoordenboek. Bij het publiceren van open data aan niet-experts is nog meer aandacht nodig voor een goede uitleg. In ieder geval moeten de namen en betekenis van attributen worden beschreven, plus de attribuutwaarden die kunnen voorkomen (vooral in het geval van codes en afkortingen), en indien van belang ook de onderlinge samenhang tussen attributen ('als dit attribuut deze waarde heeft dan kan dat attribuut alleen maar die waarden aannemen').

	AB	AC	AE	AF	AG	AH
	CAP_TA	PHYSICAL_CAPACITY	APPLIANCE	CONNECTION_SUB_TYPE	CONNECTION_TYPE	GRID_AREA
1	87420202C	G4		GWN	NRM	1237185180C
2	87420202C	G16	School	GWN	NRM	1237185180C
3	87420201C	1x25		LDN	NRM	1236123200C
4	87420201C	3x25		LDN	NRM	1236123200C
5	87420202C	G4		GWN	NRM	1237185180C
6	87420202C	G16		GWN	NRM	1237185180C
7	87420202C	G4		GWN	NRM	1237185180C
8	87420201C	1x25		LDN	NRM	1236123200C
9	87420202C	G6		GWN	NRM	1237185180C
10	87420202C	G6		GWN	NRM	1237185180C
11	87420202C	G4		GWN	NRM	1237185180C
12	87420201C	3x25		ODN	NRM	1236123200C
13	87420201C	1x25		LDN	NRM	1236123200C
14	87420202C	G4		GWN	NRM	1237185180C
15	87420201C	3x25		LDN	NRM	1236123200C
16	87420201C	3x50		LDN	NRM	1236123200C
17	87420201C	3x50		LDN	NRM	1236123200C

**Figure 2-4 Codes en afkortingen in attributwaarden**

Als er geografische gegevens in het bestand zitten (coördinaten van punten, vlakken of lijnen) is van belang dat er ook geo-specifieke metadata is: resolutie (mate van detail), kaartschaal waarvoor de geometrie geschikt is, inwinningsmethode (hoe gemeten?, zelf gemeten of afgeleid van andere bestanden?), actualiteit (wanneer ingewonnen) en andere informatie over de betrouwbaarheid en nauwkeurigheid van de geometrie. Wat in geval van geo-data ook altijd aangegeven moet worden is het coördinaat-referentiesysteem ('spatial reference system') van het bestand. Binnen Nederland wordt voor GIS-data nog steeds overwegend het RD coördinaat-referentiesysteem gebruikt (RD = Rijksdriehoeksmeting), echter veel niet-GIS applicaties gebruiken het GPS lat/lon systeem (WGS84). In Europa zal op den duur overgeschakeld worden op een op WGS84 lijkend referentie systeem, namelijk ETRS89, en ook voor het aangeven van de hoogte, nu in Nederland d.m.v. NAP, zal er uiteindelijk een EU-brede standaard zijn.

Een bestand dat de Nederlands RD coördinaten heeft kan niet zomaar gebruikt worden in combinatie met een bestand dat de WGS84 of ETRS89 lat/lon coördinaten heeft. Om de juist coördinaat transformatie uit te voeren, moet een GIS applicatie weten in welk referentiesysteem de input data is. Om die reden is informatie over het gehanteerde coördinaat-referentiesysteem een essentieel onderdeel van de documentatie.

Andere metadata gegevens zijn: de update frequentie (wanneer kan men een nieuwe versie van de data verwachten), maar ook gebruiksvoorwaarden en bijv. een aantal toepassingen waar de data voor gebruikt kan worden.

Voor wat er tot de minimale (verplichte) set van metadata behoort zijn zowel in Europees verband (INSPIRE), als in Nederland (Geonovum) afspraken gemaakt. Voor niet-geo gegevens kan men ook kijken naar de Dublin Core metadata standaard en best practices, die overigens voor een deel ook terugkomen in de standaarden waar INSPIRE zich op baseert (namelijk ISO 19115 en ISO 19139).

Aan goede metadata wordt ook voor reguliere (niet per se open data) gegevensuitwisseling veel waarde gehecht in de geo-wereld, echter in het kader van open data is het belang van metadata en

documentatie nog groter, omdat de potentiële data afnemers een veel gevarieerdere gebruikersgroep is, van energie-experts tot wetenschappers uit verschillende hoek, en burgers.

Los van de 'officiële' metadata, zoals hierboven genoemd, is een uitleg in vrije tekst op de web site of in een begeleidend document een minimale vereiste.

## 2.6 Vindbaarheid

Open data moet makkelijk te vinden zijn, anders blijft het gebruik ervan achter bij het potentiële nut van de data. Vindbaarheid kan op een aantal manieren worden bereikt. We noemen er twee, een relatief eenvoudige (volgende sectie), en een manier waarvoor meer organisatie en budget nodig is (sectie 2.6.2).

### 2.6.1 Eigen web site

De minste investering kost het om op de eigen web site een 'open data' item of tab toe te voegen die de aandacht trekt van de toevallige bezoeker, en die bovendien via zoek-machines als Google gevonden kan worden. Als een organisatie bovendien op bijeenkomsten en fora uitdraagt dat ze open data gaat verstrekken, is de web pagina of web applicatie die toegang geeft tot de data binnen niet al te lange tijd bekend.

### 2.6.2 (Centrale) portal, geo-register

De tweede optie is: de bestanden ook via geoportals beschikbaar stellen: in Nederland is er het Nationaal Georegister (figuur 2-5), in de EU wordt in INSPIRE context een begin gemaakt met een Europees breed geoportal (figuur 2-6). Dergelijke geoportals bieden vaak meerdere zoek-mogelijkheden, en metadata beschrijvingen zijn er een integraal onderdeel van.

Het is niet per se nodig dat bestanden een geografische component hebben om in een geoportal te worden opgenomen, echter wil men de data eerst bekijken voordat men download, dan is online bekijken als kaart of kaartlaag niet mogelijk in zo'n geval. Niet-geo data van te voren bekijken en exploreren als online tabel of grafiek zou kunnen worden geïmplementeerd door bestaande geoportal software uit te breiden (als het om open source software gaat, zoals GeoNetwork, <http://geonetwork-opensource.org/>).

Als alternatief kan men in plaats van de bestanden via een geo-portal vindbaar te maken ook zelf een zogenaamde Catalog Service opzetten, conform de OpenGIS standaard daarvoor, en/of conform de INSPIRE richtlijnen voor Discovery Services.

Voor niet-geo gegevens zijn er op het moment niet zo veel vergelijkbare mogelijkheden, ook niet als men de data aanbiedt als RDF. In de Linked Data wereld blijkt vindbaarheid net zo'n issue als in de geo-wereld, ondanks de gunstige voorwaarden: RDF als 'machine understandable' metadata formaat en de mogelijkheid van verwijzingen tussen individuele data items die inherent is aan RDF als taal (zie sectie 2.4). Desondanks geldt ook hier dat men vaak afhankelijk is van web pagina's met lijstjes met RDF datasets voor het opsporen van mogelijk interessante open data in RDF formaat.

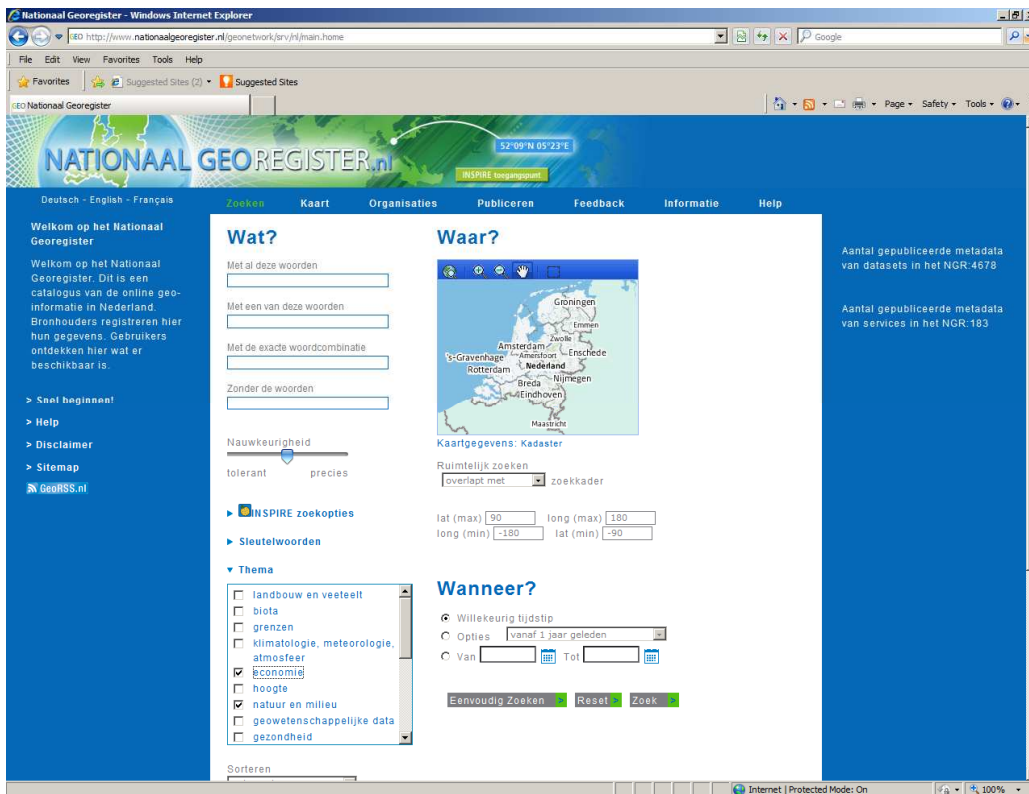


Figure 2-5 Nationaal Georegister

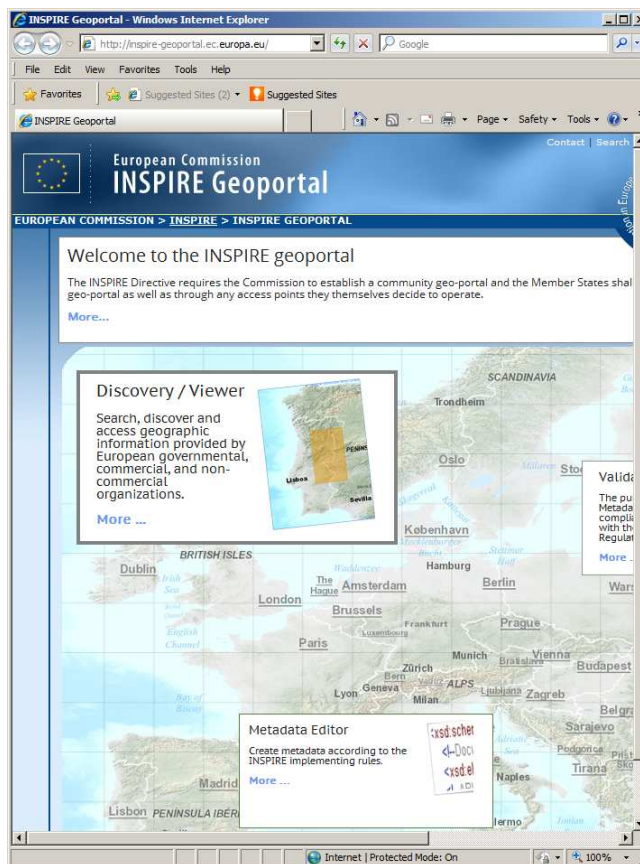


Figure 2-6 INSPIRE Geoportal

## 2.7 Gebruikersparticipatie / interactie

Niet alleen de maatschappij heeft baat bij goede open data. Ook de organisatie die de data verstrekt kan hier direct profijt van hebben, in de vorm van feedback van gebruikers als gegevens niet kloppen of niet meer actueel zijn. Er zijn verschillende vormen van feedback door gebruikers: gebruikers kunnen zelf de fouten herstellen door een edit-mogelijkheid in de web applicatie, of er is een mogelijkheid om fouten of ongerijmdheden via email te melden. Een goed voorbeeld van een dergelijke actieve gebruikersparticipatie is OpenStreetMap, wat niet verwonderlijk is gezien het karakter van dit initiatief: een succesvolle vorm van crowd-sourced data (zie figuur 2-7).

Onder dit aspect vallen ook andere vormen van gebruikersinteractie: er kan een wiki worden opgezet, een community blog, of een andere vorm van het communicatie worden aangeboden. Hoe meer belang de organisatie die de data openbaar maakt hecht aan het gebruik van die data, hoe meer aandacht dit punt zal krijgen.

Ook het monitoren van het gebruik van de data op andere manieren, zoals het loggen van specifieke web service requests of van download acties hoort bij dit aspect (zie ook Van Loenen en Bregt, 2012).

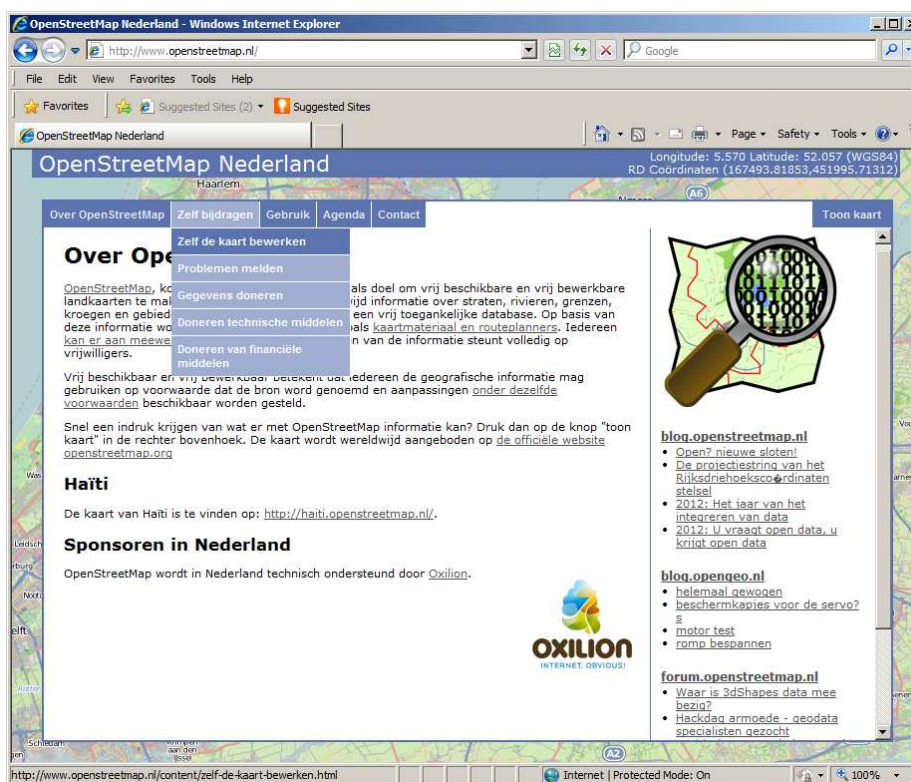


Figure 2-7 OpenStreetMap: editen en andere vormen van feedback

## 3 Internationale standaarden

Aansluiting bij internationale standaarden en ontwikkelingen is aangegeven als belangrijk punt door de Programmaraad NGInfra. In het vorige hoofdstuk, in sectie 2.4, is al een eerste indeling gemaakt in soorten standaarden, en erop gewezen dat er zowel in de energie sector als daarbuiten relevante internationale standaardisatie-initiatieven zijn. In de energie sector is er o.a. het Common Information Model (CIM), in de geo-wereld het werk van het Open Geospatial Consortium (OGC), ISO TC211 en op Europees vlak: INSPIRE (=INfrastructure for SPatial InfoRmation in Europe). In de niet-geo wereld is er de W3C als bron van nieuwe standaarden en best practices, o.a. recentelijk op het terrein van Linked Open Data. In dit hoofdstuk wordt kort ingegaan op deze ontwikkelingen ter aanvulling op met name sectie 2.4 (Open standaarden) in het vorige hoofdstuk.

### 3.1 Energie sector

Binnen de energiesector zijn op het gebied van standaarden de volgende internationale ontwikkelingen te onderkennen:

- Op Europees niveau worden sectorstandaarden in samenwerking met CEN/CENELEC/ETSI gekozen en geharmoniseerd door de Taskforce Smart Grids van de Europese Commissie. Momenteel zijn 4 Expert groepen bezig met de uitwerking. Het gros van de standaarden wordt overgenomen, alleen daar waar privacy en security een rol spelen (mede ingegeven door de Nederlandse ervaringen op dit onderwerp) worden standaarden aangepast.

Daarnaast wordt er in het kader van de INSPIRE Directive (zie sectie 3.2.2) gewerkt aan standaard gegevensmodellen voor een groot aantal sectoren. Voor de energie sector is met name van belang het data thema 'Utility and governmental services', en daarbinnen het sub-thema 'Utility services / networks'. Deze en andere INSPIRE data modellen moeten uiterlijk in 2017 in de EU lidstaten verplicht worden toegepast. Vandaar ook dat het van belang is de INSPIRE ontwikkelingen actief te volgen en waar mogelijk een actieve bijdrage te leveren aan de besluitvorming.

- Op internationaal niveau worden standaarden voor de energie sector in samenwerking met ISO, IEC en ITU gekozen en geharmoniseerd door het NIST (National Institute for Standards and Technology in de VS), in samenwerking met de FERC (Federal Energy Regulatory Commission). Het gaat hier met name om het verder uitwerken en invoeren van het Common Information Model (CIM).

### 3.2 Geo wereld

#### 3.2.1 Open GIS (OGC) en ISO TC211

De term 'Open GIS' is nauw verbonden met de naam van het OpenGIS (nu Open Geospatial) Consortium (OGC). Daarom wordt hier kort ingegaan op ontstaansgeschiedenis, doelstellingen en werkwijze van het OGC.

Naarmate het aantal GIS- en CAD-pakketten in de jaren '90 toenam, kreeg de geo-wereld er ook een probleem bij: hoe ervoor te zorgen dat de verschillende bestandsformaten niet een barrière zouden vormen voor het uitwisselen van geo-informatie tussen die verschillende pakketten. Uitwisselen van

geo-informatie gebeurde met het op floppy zetten en later ook e-mailen of downloaden van bestanden, maar het was altijd maar de vraag of de bestanden goed konden worden ingelezen in het GIS of CAD-systeem van de ontvangende partij. Tegen deze achtergrond werd in 1994 OGC opgericht door een aantal software bedrijven, overheidsinstellingen en universiteiten met als belangrijkste doel om de uitwisselbaarheid van geo-informatie binnen organisaties en tussen organisaties te verbeteren.

De term 'Open GIS' is sindsdien een zelfstandig leven gaan leiden en in de loop der jaren synoniem geworden voor: interoperabiliteit, uitwisselbaarheid, gestandaardiseerde web services, informatie integratie, open standaarden in combinatie met aansluiting op algemene ICT trends.

Twee ontwikkelingen kwamen hier bij elkaar: de toegenomen vraag naar geo-informatie voor allerlei (beleids)toepassingen en maatschappelijke vraagstukken, en de nieuwe mogelijkheden die Internet bood. Veel van de OGC standaarden hebben dan ook direct of indirect te maken met Web technologie.

OGC houdt zich naast het werken aan standaarden voor geo-data en geo-data services ook bezig met het organiseren van interoperabiliteits-testbeds. Daarmee is het veel meer een 'hands on' organisatie dan ISO. Dat neemt niet weg dat OGC samenwerkt met ISO, en een aantal belangrijke OGC specificaties is al omgezet naar ISO TC211 standaarden (de 19nnn serie). Dit geldt niet alleen voor GML, maar ook voor bijvoorbeeld de WMS en WFS specificaties voor web services (zie sectie 2.4).

Tenslotte nog iets over het woord 'open'. Het is uitdrukkelijk niet zo dat 'Open GIS' of 'open interface' gelijk staat aan 'open source'. In het geval van 'open source' software is niet alleen de gecompileerde versie maar ook de broncode van de software beschikbaar. Bij 'Open GIS' gaat het daar niet om. Het is alleen belangrijk dat de software aangeroepen kan worden door software van een andere leverancier: doordat bijv. in het geval van een web service de interface (het protocol) van die web service 'compliant' is met de standaard, en doordat de gebruikte dataformaten en gegevensmodellen geïmplementeerd zijn volgens één van de OGC of ISO standaarden. Op die manier kan interoperabiliteit in een open, gedistribueerde web architectuur worden bereikt.

Het is overigens wel zo dat bij de eerste implementeerdere van de OGC web service standaarden ook een aantal open source initiatieven zijn, zoals GeoServer (<http://geoserver.org>) en UMN MapServer (<http://mapserver.org/>).

### 3.2.2 INSPIRE

Zeer bepalend voor het uitwisselen en harmoniseren van geodata in Europa zijn sinds 2005 de standaardisatie activiteiten in het kader van INSPIRE.

Inzet van INSPIRE is dat er binnen de EU op een makkelijke en kwalitatief hoogwaardige manier geo-informatie kan worden uitgewisseld, tussen landen en binnen landen, waarbij de data zo wordt aangeboden dat er voldaan wordt aan de data specificaties (EU-brede gegevensmodellen) en de regels over netwerkdiensten, metadata, monitoring en toegang.

Daarvoor zijn een groot aantal standaarden en afspraken gemaakt en nog in de maak. Het gaat om formele wet- en regelgeving, te beginnen met de Directive 2007/2/EC van het Europese parlement (14 maart 2007).

Samengevat is het hoofddoel van INSPIRE, volgens artikel 1.1 van deze Directive:

"The purpose of this Directive is to lay down general rules aimed at the establishment of the Infrastructure for Spatial Information in the European Community (hereinafter referred to as Inspire), for the purposes of Community environmental policies and policies or activities which may have an impact on the environment." (EU 2007)

INSPIRE schrijft zogenaamde Implementing Rules (IR) voor op de volgende terreinen:

- Metadata
- Network services
- Data specifications
- Data and service sharing
- Monitoring and reporting

Relevant voor de manier waarop (open) data via het web beschikbaar gesteld kan worden is de onderverdeling die INSPIRE maakt binnen de Network services:

- Discovery Services
- View Services
- Download Services
- Transformation Services
- Invoke Spatial Data Services

Bij Download Services wordt nog een onderscheid gemaakt tussen een 'gewone' download service en een 'direct access' download service. In het laatste geval is sprake van een interactieve connectie met de data (via een web service).

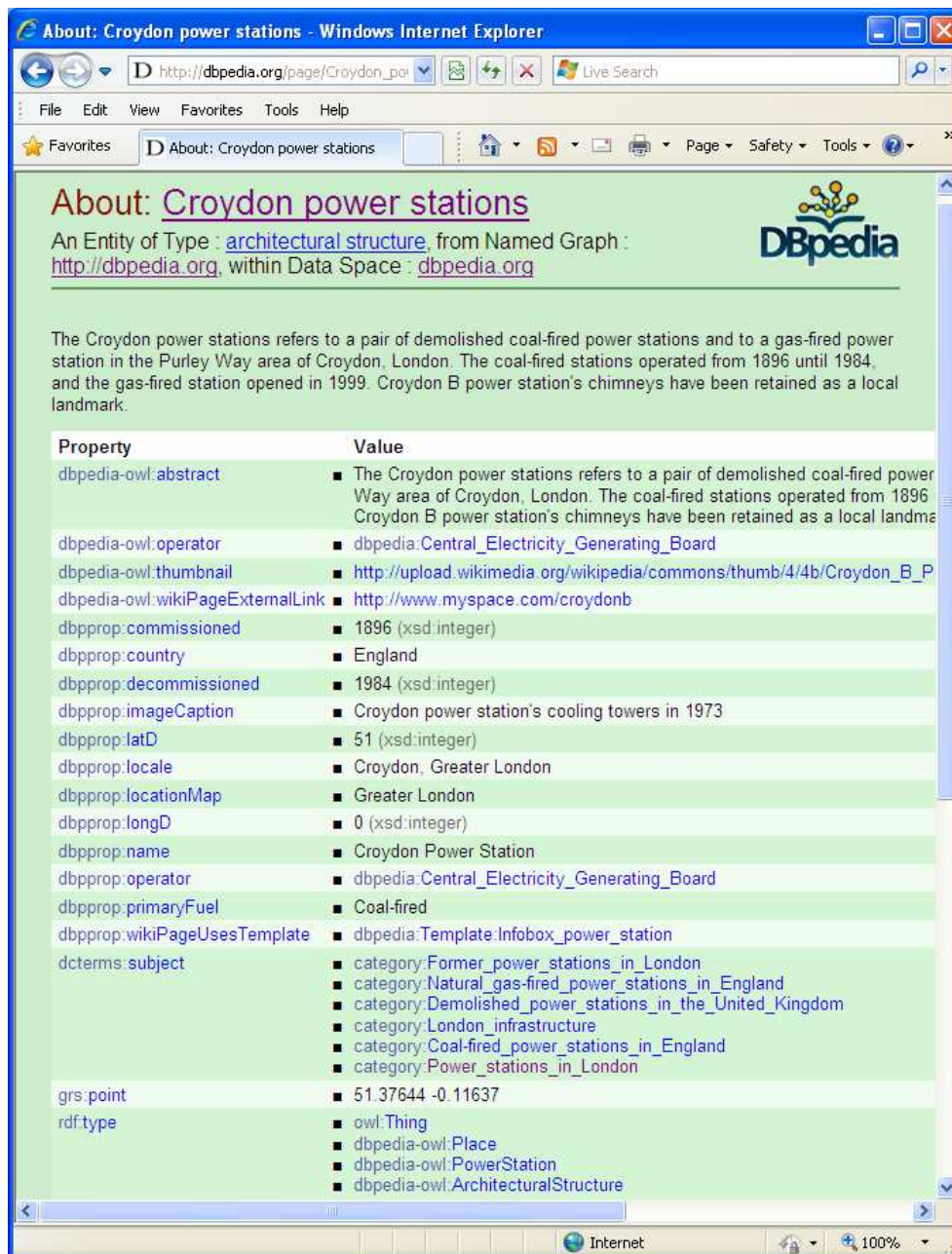
In Nederland wordt de voorbereiding en implementatie van INSPIRE gecoördineerd en ondersteund door Geonovum: <http://www.geonovum.nl>.

### 3.3 W3C en linked open data

Internet biedt de mogelijkheid om van de ene url naar de andere te navigeren. Gebaseerd op dat basiskenmerk van het Web is de laatste 5 jaar het idee van Linked Open Data ontstaan. De 'uitvinder' van het Web, Tim Berners-Lee formuleerde vier basisprincipes van linked data: gebruik URIs (Uniform Resource Identifier) als (unieke) namen voor 'dingen' waarvan je gegevens vast legt (mensen, plekken op aarde, fotos, concepten etc.); gebruik web adressen voor die URI's, zodat er echt een url is die naar een web pagina of foto of muziekfragment etc. op het web leidt; verschaf bruikbare informatie over die dingen en gebruik standaarden zoals RDF (Resource Description Framework); en als vierde principe: neem ook links op in de beschrijving van iets naar iets anders. In plaats van een Internet van web paginas die via hyperlinks verbonden zijn, ontstaat zo een Web van Data waar RDF data met andere RDF data verbonden is via links.

Een voorbeeld van dit soort toepassingen is DBPedia, waarbij informatie uit Wikipedia wordt gepubliceerd als RDF. Door de links in de RDF kan van het ene naar het andere informatie-item (met bijbehorende web pagina) worden gesprongen, zie figuur 3-1.





**Figure 3-1 Linked data in een viewer**

Linked data kan op meerdere manieren worden gebruikt en geëxploreerd: in een viewer, zoals in figuur 3-1, maar ook via een query tool, die gehele RDF databases doorzoekt. De query taal die gebruikt wordt is in dat geval (meestal): SPARQL.

## 4 Conclusies en vooruitblik

### 4.1 Afweging kernthema's

In dit rapport zijn per technisch aspect een aantal keuze-mogelijkheden voor de implementatie van open data beleid gepresenteerd. Een van de conclusies is dat er niet slechts één implementatiestrategie of één technologie uitrolt als meest optimaal. Er zijn meerdere combinaties tussen de technische aspecten mogelijk, keus voor een optie leidt niet automatisch tot een bepaalde keuze binnen een ander technisch aspect.

Een strategie die toelaat dat stap voor stap wordt geëxperimenteerd met open data kan er zo uit zien: De minste investering aan de web server kant is: leveringsmodel A: klaarzetten van bestanden, geen selectie door de gebruiker, maar ook geen wachttijd, wat met een online form waarmee de gebruiker een bepaalde subset kan aanvragen, meestal wel het geval is.

Als vervolgstap kan daarna een web service (bijv. een OGC/INSPIRE Web Feature Service) worden opgezet, die meer mogelijkheden biedt tot online viewen (exploreren) en selectie. Ook dit kan in twee stappen gebeuren: eerst alleen 'losse' web services, die dan aangeroepen worden (via url) vanuit de client software van gebruikers. Uiteindelijk, als oplossing die initieel de meeste investering vergt, maar die zich later terugverdient, kan men ervoor kiezen om zelf een interactieve web client online te zetten, die veel meer biedt dan alleen de data download: een selectie frame, documentatie, links naar andere data, eventueel transformatie services die de data naar een gewenst coördinaatsysteem en een gewenst output gegevensmodel transformeren.

Met een dergelijke client is de web techniek geavanceerd, en schuift het leveringsmodel op naar type B: gebruiker kan zelf selecties maken, en alleen die data downloaden die men wil hebben ('data on demand'). Men kan zo'n interactieve web client bovendien zo bouwen dat gebruikers de data real-time kunnen combineren met data afkomstig van services van andere data leveranciers.

Tenslotte, het gebruik van open standaarden is besproken als een belangrijk aspect: zowel het bestandsformaat, als het gegevensmodel en de typen web services die worden ingezet.

Hoewel de keuze voor een geschikt (open) bestandsformaat zeker belangrijk is, is het advies hier om een aantal formaten aan te bieden, zowel de-jure standaarden (GML, RDF, CSV) als de-facto standaarden (ESRI shapefiles, desnoods MS Excel spreadsheets).

Wat betreft gegevensmodel en web service standaarden: op de langere termijn zijn die van meer belang dan het bestandsformaat.

### 4.2 Alliander en internationale standaarden

De scope van het werkveld van een organisatie als Alliander is zowel nationaal als internationaal. Bij internationaal gaat het niet alleen om voldoen aan Europese regelgeving, maar kan het ook gaan om participatie in grensoverschrijdende projecten, bijvoorbeeld een samenwerkingsverband tussen een Nederlandse provincie en een Duitse deelstaat. Door goed te letten op wat er in INSPIRE verband gebeurt en er waar mogelijk zelf in INSPIRE werkgroepen te participeren is het geo-deel van de standaardisatie initiatieven in Europa afgedekt (data thema 'Utility and governmental services', View en Download Services, en ook de ontwikkelingen en best practices op het gebied van Transformation Services voor model transformatie en voor coördinaat transformatie).

Voor de vertaling van de INSPIRE regels naar de Nederlandse situatie is Geonovum een goede intermediair.

Daarnaast kan gekeken worden naar de ontwikkelingen in de US (het Common Information Model), als de INSPIRE data modellen te beperkt zouden zijn om de informatie-inhoud van de data van Alliander af te dekken.

Tenslotte is het op het front van de Linked Open Data goed te kijken hoe organisaties als de Ordnance Survey in de UK, en Geonovum in Nederland, omgaan met de linked data principes. Hier zijn vooral de ontwikkelingen en (de facto) standaarden van W3C van belang: RDF voor de data inhoud, SPARQL voor het zoeken, en allerlei RDF viewers voor het aanbieden van linked data.

### 4.3 Vervolgstappen t.b.v. opzetten technische infrastructuur

Belangrijk voor het opzetten van de technische infrastructuur voor de vervolgfase van dit project zijn in ieder geval de volgende vragen en aandachtspunten:

1. Hoe is de data nu opgeslagen? Is dit in Oracle of een ander object-relacioneel database management system? En zijn de voor open data in aanmerking komende datasets in dezelfde database opgeslagen, of in verschillende database systemen?
2. Heeft Alliander een web server draaien waar makkelijk een extra web service (voor de sandbox testbeds) aan toegevoegd kan worden? Of heeft men de voorkeur voor een web server die bijv. bij de TU Delft draait en die door de TU Delft wordt onderhouden? In dat geval kan de data toch bij Alliander blijven: de web service die de data beschikbaar maakt draait dan op een andere host dan waar de datasets staan. Meerdere implementaties zijn hier mogelijk, en in de sandbox testbeds kunnen de verschillende opties worden uitgeprobeerd.
3. Data preparatie: Ook als men de 'ruwe' data beschikbaar stelt (en niet eerst bijv. aggregiert) is het aan te raden een kopie of extract maken van de data die via de web service beschikbaar wordt gesteld, en niet rechtstreeks toegang tot de productiedata te geven. Bij grote databases vergt dat enige planning, en eventueel extra hardware.
4. Toegang tot de sandbox testbeds: voor de pilot zou een wachtwoord-beveiligde web site kunnen worden opgezet, zodat een aantal opties kunnen worden uitgetest (wel/niet anonimiseren bijv.) zonder dat dit problemen geeft
5. Voor het vinden van een optimale manier voor het extracten en transformeren van de datasets van Alliander die als open data naar buiten gaan, is het aan te bevelen om een testbed omgeving op te zetten specifiek voor die (on-the-fly) transformatie (zowel model /schema transformatie als coördinaat transformatie).
6. Wat te doen als in datasets wel adressen of postcodes zitten, maar niet de coördinaten? Dan de coördinaten met geo-referencing technieken toevoegen (verrijken van de datasets)?
7. Speelt 3D een rol? (electriciteits kabels en gasleidingen op een realistische manier afbeelden? Hoogbouw/laagbouw onderscheid bij woningen, volumeberekening in relatie tot energieverbruik?, etc.
8. Welke gratis of open source web clients kunnen het best gebruikt worden voor de sandbox tests? Idem de web service software (Geoserver, UMN Mapserver of deegree?) en in dit verband ook: moet met SOAP binding worden gewerkt?
9. 'Begrijpbaarheid' wordt in dit rapport aangewezen als belangrijk aspect, juist omdat het in het geval van open data gaat om een gevarieerde groep van gebruikers, van specialisten tot 'gewone' burgers. Hoe dit te realiseren (i.e. het punt van het overbrengen van de betekenis van de (ruwe) data, zodat de data goed geïnterpreteerd kan worden) verdient de aandacht in de pilots.
10. Moet de data (ook of juist met name) als linked / linkable data (in RDF) beschikbaar worden gesteld? Dit zou goed aansluiten op ontwikkelingen in W3C en semantisch web hoek.

11. Geavanceerd zou zijn om behalve open data ook een aantal innovatieve web services aan te bieden: behalve de al genoemde INSPIRE transformation services, ook andere processing services: een geo-coding service voor algemeen gebruik, of bijv. een aggregatie-service, voor online afleiden van statistieken en het tekenen van grafieken.

## Literatuur

Bizer, C., Heath, T. and Berners-Lee, T., 2010, Linked Data - The Story So Far. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), 5, pp. 1-22.

Bregt, A. en W. Castelein, 2012, Open data and beyond; Exploring existing open data projects to prepare a successful open data strategy; Deelrapport organisatorische aspecten voor een succesvolle open data strategie. Delft: Next Generation Infrastructures.

CIMug (=Common Information Model users group), 2011, Current CIM Model, <http://cimug.ucaiug.org/CIM%20Model%20Releases/Forms/AllItems.aspx>

EDSN, 2009, Centraal Aansluitingenregister: Attributenlijst Versie:2.0.

Egyedi, T.M. (2010). Standards and Sustainable Infrastructures: Matching Compatibility Strategies with System Flexibility Objectives. In: Sherrie Bolin (Ed.), Unifier or Divider?

e-Overheid.nl, 2010, NORA (Nederlandse Overheid Referentie Architectuur) v3.0, <http://www.e-overheid.nl/onderwerpen/architectuur-en-nora/nora-30>

EPRI, 2011, Common Information Model Primer. First Edition. Electric Power Research Institute.

EU, 2007, DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

EU, 2010, COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services.

EZ, 2010, Op weg naar intelligente netten in Nederland. Discussiedocument van de Taskforce Intelligente Netten. Ministerie van Economische Zaken.

Forum Standaardisatie, 2010, Open Standaarden, <http://www.open-standaarden.nl/>

Geonovum, 2010, Nederlands profiel op ISO DIS 19142 Geographic information — Web Feature Service. 1/6/2010.

Geonovum, 2010, NORA dossier geo-informatie 1.0, 19 november 2010.

Geonovum, 2010, Raamwerk van geo-standaarden, versie 2.1, 21 mei 2010.

Howard, M., Payne, S. and Sunderland, R., 2010, Technical Guidance for the INSPIRE Schema Transformation Network Service. Version: 3.0. Status: Final.

IEC, 2012, Core IEC Standards, <http://www.iec.ch/smartgrid/standards/>

INSPIRE, 2010, D2.5: Generic Conceptual Model, Version 3.3.

INSPIRE, 2011, D2.8.III.6 Data Specification on Utility and governmental services – Draft Guidelines.

ISO/TC211, <http://www.isotc211.org/>

ISO, 2007, ISO 19136:2007: Geographic information — Geography Markup Language (GML).

Lehto, L., 2007, Real-Time Content Transformations in a Web Service-Based Delivery Architecture for Geographic Information. Helsinki University of Technology.

van Loenen, B., 2012. Open data and beyond: Exploring existing open data projects to prepare a successful open data strategy, Deelrapport inleiding open data. Delft: Next Generation Infrastructures.

van Loenen, B. A. Bregt, 2012. Open data and beyond: Exploring existing open data projects to prepare a successful open data strategy, Deelrapport Instrumenten voor de monitoring van de impact van open data (bij Alliander). Delft: Next Generation Infrastructures.

Lopez-Pellicer, F.J., et al., 2010, Exposing CSW catalogues as Linked Data. In: 13th AGILE International Conference on Geographic Information Science 2010 (Guimarães, Portugal).

NIST, 2010, NIST -identified Standards for Consideration by Regulators. Release 1.0.

OGC, 2012, OGC® Standards and Specifications, <http://www.opengeospatial.org/standards>

OGC, 2005, OpenGIS® Catalogue Services Specification 2.0 - ISO19115/ISO19119 Application Profile for CSW 2.0. Version: 0.9.3.

OGC, 2010, OpenGIS Web Feature Service 2.0 Interface Standard. Version: 2.0.0. In: OpenGIS® Implementation Standard.

OGC, 2010, OpenGIS Filter Encoding 2.0 Encoding Standard. Version: 2.0.0. In: OpenGIS® Implementation Standard.

OGC, 2010, Draft for Candidate OpenGIS® Web 3D Service Interface Standard. Version: 0.4.0. A. Schilling and T.H. Kolbe (Eds.).

Oosterom, P.J.M. van, INSPIRE activiteiten in het jaar 2008, GIST Rapport No. 52, Delft, 2009, 142 p.

Reitz, T., B. Schneiders and M.E. de Vries, Mediator Service Component Specification, Humboldt Report A5.2-D1 [2.2], 2009, 46 p.

Reitz, T., M.E. de Vries and D. Fitzner, Conceptual Schema Specification and Mapping, Humboldt Report A5.2-D3 [3.3], 2009, 23 p.

Rijksoverheid.nl, 2010, Overzicht basisregistraties, <http://www.rijksoverheid.nl/onderwerpen/basisregistraties/overzicht-basisregistraties>

Schade, S., 2010, Computer-Tractable Translation of Geospatial Data. International Journal of Spatial Data Infrastructures Research, 5.

Tennison, J., 2010, Distributed Publication and Querying,  
<http://www.jenitennison.com/blog/node/143>.

Vries, M.E. de and A. Fichtinger, Models and schemas in HUMBOLDT, Humboldt Report, 2009, 8 p.

W3C, 2004, RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004.

W3C, 2008, Best Practice Recipes for Publishing RDF Vocabularies. W3C Working Group Note.

## Bijlage A: INSPIRE - Utility Services / Networks

Hieronder een overzicht in vogelvlucht van het (draft) gegevensmodel voor het INSPIRE thema 'Utility and governmental services', sub-thema 'Utility services / networks', met op de volgende pagina een lijst van afhankelijkheden met andere data themas.

(bron: INSPIRE, 2011, D2.8.III.6 Data Specification on Utility and governmental services – Draft Guidelines, versie 2011-06-17, p.

Pipeline – oil, gas

- category of content
- segment id
- capacity, max
- average volume
- diameter
- pressure regime
- construction system
- date of construction
- responsible organization

Sewage system network

- segment id
- capacity, max
- average volume
- construction system, including e.g. material used for building the network (cast iron, cement ...)
- altitude
- date of construction
- responsible organization

Water supply system network

- segment id
- capacity, max
- average volume
- construction system, including e.g. material used for building the network (cast iron, cement ...)
- date of construction
- responsible organization

Electricity transmission lines

- segment id
- capacity, max
- average volume
- construction system
- date of construction
- responsible organization

Transmission network for different kind of data/ signals

- segment id
- category of object (feature type) e.g. antennas, base-stations, relay-antennas, cables etc



- category of data/signal, e.g. radio, data, mobile, TV
- date of construction
- responsible organization

Overlaps / links with other themes

This sub-theme might overlap with themes:

- Hydrography (A-I.8)
- Buildings (A-III.2)
- Land use (A-III.4)
- Environmental monitoring facilities (like treatment plants/pumping stations) (A-III.7)
- Production and industrial facilities (A-III.8)
- Energy resources (A-III.20)

Current sub-theme holds potential dependencies with the following themes:

- Annex I
  - o Coordinate reference systems (geo-referencing of the point)
  - o Geographical grid systems (geo-referencing of the point)
  - o Geographical names (Identification of the point and of the place where it is located)
  - o Administrative units (that contain the point)
  - o Addresses (referencing of the point)
  - o Cadastral parcels (that contain the given service and from which the service is provided)
  - o Transport networks (that provide access to/from the services)
  - o Protected sites (that may contain services or being potential receptors of these)
  
- Annex II
  - o Elevation (referencing of the point)
  
- Annex III
  - o Statistical units (that contain the point)
  - o Buildings (that contain the given service and from which the service is provided)
  - o Population distribution – demography (potential service “clients”)
  - o Utilities (that the service make use/depend on)

## Bijlage B: GML encoding – voorbeeld

Het GML voorbeeld hieronder is een fragment van de output van een Web Feature Service van Rijks-waterstaat.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<wfs:FeatureCollection xmlns:rws="http://mapserver.gis.umn.edu/mapserver"
  xmlns:gml="http://www.opengis.net/gml"
  xmlns:wfs="http://www.opengis.net/wfs"
  xmlns:ogc="http://www.opengis.net/ogc"
  numberOfFeatures="200">
  <gml:boundedBy>
    <gml:Envelope srsName="EPSG:28992">
      <gml:lowerCorner>-592.114856 538587.989251</gml:lowerCorner>
      <gml:upperCorner>219682.906443 846361.186170</gml:upperCorner>
    </gml:Envelope>
  </gml:boundedBy>
  <gml:featureMember>
    <rws:AAA200 gml:id="AAA200.92">
      <rws:geometry>
        <gml:Polygon srsName="EPSG:28992">
          <gml:exterior>
            <gml:LinearRing>
              <gml:posList srsDimension="2">116071.303344
601395.492086 116577.834889 599847.330280 118623.242696 600516.601760
118116.733727 602064.730799 116071.303344 601395.492086
              </gml:posList>
            </gml:LinearRing>
          </gml:exterior>
        </gml:Polygon>
      </rws:geometry>
      <rws:OBJECTID_1>92</rws:OBJECTID_1>
      <rws:OBJECTID>92</rws:OBJECTID>
      <rws:OPMERKING> </rws:OPMERKING>
      <rws:BESTEM_ZAN>Strandsuppletie</rws:BESTEM_ZAN>
      <rws:BLOK_NCP>L12</rws:BLOK_NCP>
      <rws:NAAMGEBIED>L12B</rws:NAAMGEBIED>
      <rws:SOORT_GEB>Zandwinning</rws:SOORT_GEB>
      <rws:STATUS>Verlaten</rws:STATUS>
      <rws:UITGEPUT>Nee</rws:UITGEPUT>
      <rws:OMSCHRIJVI>Zandwinning</rws:OMSCHRIJVI>
      <rws:AANMAAK>00:00:00 01/01/2003</rws:AANMAAK>
      <rws:MUTATIE>00:00:00 02/02/2006</rws:MUTATIE>
      <rws:IN_GEBRUIK>Nee</rws:IN_GEBRUIK>
      <rws:M__WINDIEP>2</rws:M__WINDIEP>
      <rws:HECTARE>350.465</rws:HECTARE>
    </rws:AAA200>
  </gml:featureMember>
</wfs:FeatureCollection>
```

```
        <ows:SHAPE>&lt;shape&gt;</ows:SHAPE>
    </ows:AAA200>
</gml:featureMember>

<gml:featureMember>
    <ows:AAA200 gml:id="AAA200.170">
        ...
    </ows:AAA200>
</gml:featureMember>
etc.

</wfs:FeatureCollection>
```



**Onderzoeksinstituut OTB**

Delft University of Technology

Jaffalaan 9, 2628 BX Delft, The Netherlands

Postbus 5030, 2600 GA Delft, The Netherlands

Telefoon +31 (0)15 278 30 05

Fax +31 (0)15 278 44 22

E-mail [mailbox@otb.tudelft.nl](mailto:mailbox@otb.tudelft.nl)

**[www.otb.tudelft.nl](http://www.otb.tudelft.nl)**