

# Slechte data bestaan niet?

Onder de titel *Ruimtelijke data op orde & Datakwaliteit* organiseerde Ruimteschepper op 3 juli een bijeenkomst bij Wageningen Environmental Research (WENR). De kwaliteit van ruimtelijke data is natuurlijk van belang bij beleidsvorming en de onderbouwing ervan. Dat werd van verschillende kanten belicht. Behalve de sprekers mochten ook de aanwezigen hun zegje doen. Eén van de vragen aan de ruim 50 aanwezigen was of ze het eens waren met de stelling 'Slechte datakwaliteit bestaat niet'.

Wies Vullings introduceerde het onderwerp. Kwaliteit is volgens ISO de mate waarin gegevens geschikt zijn voor het doel waarvoor ze gebruikt worden. Het zou mooi zijn om een sterrenstelsel te hebben, zoals bij hotels, om aan te geven wat de kwaliteit is van data. Maar waar dat bij hotels heel goed werkt, omdat het doel eenduidig is (een nachtje slapen), is het doel bij gebruik van data steeds weer anders. Data wordt gebruikt voor analyses, visualisaties, juridische aspecten. En ieder gebruik stelt weer andere eisen. Fitness for use is dan ook dé manier om met kwaliteit van data om te gaan.

In de presentaties kwamen deze aspecten aan de orde. Wies Vullings sloot af met een enquête. De antwoorden, gemiddeld vijftig per vraag, staan in het kader 'Antwoorden'.

## Interne en externe kwaliteit

Als producent van een databestand wil je natuurlijk weten of je eigen interne kwaliteitsnormen gehaald worden. Kadaster liet daartoe de kwaliteit van de Basisregistratie Topografie onderzoeken. Spreker Maarten Storm (WENR) schetste hoe hij deze 'BRT Audit' had aangepakt. Na het opstellen van een controleprotocol werd een

## Antwoorden

1. **Waar werk je?**  
41% (Lokale) overheid, 39% GIS/GEO-Bedrijf, 16% Anders, 4% Onderwijs/Onderzoek
2. **Slechte datakwaliteit bestaat niet.**  
80% oneens, 20% eens
3. **Welk kwaliteitskenmerk vindt u het meest belangrijk?**  
42% betrouwbaarheid, 20% actualiteit, 14% volledigheid, 6% nauwkeurigheid, 6% toegankelijkheid, 12% anders
4. **Kwaliteit is eigenlijk een kostenverhaal: Risico's bepalen welke marge je aan kan houden**  
78% eens, 22% oneens
5. **Wat is volgens u de belangrijkste succesfactor voor het realiseren van benodigde datakwaliteit?**  
56% Gerichtheid op business en gebruikers, 42% Regie & samenwerking in de gegevensbeheerketen, 2% Structurele toereikende financiering
6. **Onze organisatie heeft een datakwaliteitsstrategie.**  
48% Nee, 28% Ja, 24% Wordt aan gewerkt
7. **Architectuur en procesmanagement is leuk; maar de data moet gewoon precies kloppen**  
63% oneens, 37% eens
8. **Het opstellen van data architectuur moet gedaan worden:**  
51%: Op hoofdlijnen voordat je de toepassing gaat implementeren, 37%: Voordat je de toepassing gaat implementeren, 12%: Tijdens gebruik van de toepassing, wanneer herbruikbaarheid nuttig blijkt, 0%: Terwijl je de toepassing aan het implementeren bent.
9. **Laat burgers zelf meten en data verzamelen: het is dé garantie voor goede datakwaliteit.**  
63%: De wet van de grote getallen bevordert de datakwaliteit, 33%: Massa heeft collectief meer kennis en creativiteit dan officiële instanties, 4%: Dat kan nooit wat worden



Wies Vullings, WENR.

detailuitwerking gemaakt. Dat omvat wat er wordt gecontroleerd, naar welke kwaliteitsaspecten wordt gekeken, hoe groot de steekproef moet zijn en wat als referentie wordt gebruikt. Bij de uitvoering van de datacontrole is het nodig om kennis te hebben van controle methodieken, inwincriteria en de data zelf. Met de uitkomsten van de controles wordt het resultaat berekend. Idealiter zou dat op 100% uitkomen. Maar zelfs een databestand met een interne kwaliteit van 100% is niet zomaar geschikt voor externe gebruikers. Externe kwaliteit is wat anders. Immers, als je een door anderen gemaakte dataset wil gebruiken, dan wordt de kwaliteit bepaald door de bruikbaarheid voor jouw gebruiksdoel, de 'Fitness for Use'. Daarbij gaat het erom in hoeverre de dataset aan jouw criteria voldoet.

### Baten van meer detail

Het gebruik van de Bodemkaart van Nederland schaal 1:50.000 (Bod50NL) voor het berekenen van oogstschade door drinkwaterwinning is een voorbeeld van onvoldoende externe kwaliteit. Martin Knotters (WENR) beschreef hoe hij had aangetoond dat een bodemkartering met schaal 1:25.000 geschikter is voor het schatten van de opbrengstschade dan die met schaal 1:50.000. De verbetering is €13 waard, per hectare, per jaar. De contante waarde hiervan is berekend op €258 per hectare, terwijl de kosten van de bodemkartering, schaal 1:25.000, €30 per hectare bedroegen. De kosten-batenverhouding is dus circa 1:8. Helaas is die gedetailleerde kartering niet landsdekkend beschikbaar. Martin Knotters meent dat een landelijke 1:25.000 bodemkartering economisch aantrekkelijk is. Wetenswaardigheid: voor België is een bodemkaart met schaal 1:20.000 beschikbaar.

### Informatiearchitectuur

Een manier om de potentiële bruikbaarheid van datasets te vergroten is om ze op te bouwen volgens een informatiearchitectuur. Aad Verboom (Verboom Advies) gaf aan dat het gebruik van informatiemodellen, zoals NORA of GEMMA, daarbij helpt. Als toegevoegde waarde van data architectuur noemde hij: vindbaarheid, herbruikbaarheid, het versterken van onderdelen en het versterken van gegevensmanagement.

### Beleg de taken

Als een organisatie data over het eigen beheergebied toelevert aan een basisregistratie, zoals de BGT, dan moet de interne kwaliteit hoog zijn. Coen Wessels (NEXPRI), adviseur bij de jonge fusiegemeente Hoeksche Waard, wees erop dat daarvoor een heldere regie nodig is op de dataketen. De verschillende taken en verantwoordelijkheden in en om databeheer moeten duidelijk belegd zijn. Daarnaast zijn voldoende capaciteit en budget onmisbaar. Zonder dat is de kans kleiner dat de kwaliteitseisen worden gehaald. De spreker verwees onder meer naar DAMA-DMBOK (Data Management Body of Knowledge) als referentiekader.

### Perfectie niet nodig

Bruikbare data hoeft niet altijd in alle opzichten van hoge kwaliteit te zijn. Dat bleek uit het verhaal van Rolf Bruijn (Glasvezel buitenaf). Als een van de twee geo-analisten voor een glasvezelkabellegger beschreef hij op provocerende wijze dat hun werk ook gedaan kan worden met minder ideale datasets. Voor hen was het belangrijkste criterium dat de beschikbare datasets volledig zijn, daarmee komen ze een heel eind. In hun situatie

hebben ze te maken met veel databronnen en veel datatransportkanalen. Ze werken zonder datamodellen en architectuur. Indruk: in dit stukje bedrijfsleven wordt geroeid met de riemen die men heeft.

### Smarticiperen

Datakwaliteit is ook van groot belang in een Smart City. Hierover sprak lemke Idsingh (i4Urban). De door allerlei stedelijke bronnen (sensoren) continu geleverde (big) data moeten automatisch verwerkt en gecombineerd worden tot signalen van eventuele problemen in het stadsgebeuren. Naar zijn mening moeten we meer ruimte geven aan het bewustzijn van kwaliteit. Er moet meer geharmoniseerd en verbonden worden. In die verbinding zit volgens hem de ware kwaliteit. Het lijkt hem mooi als burgers kunnen gaan 'smarticiperen', oftewel zelf data aanleveren.

### Datakwaliteit is relatief

De boodschap van deze middag zou je kunnen zien als: datakwaliteit mag lager zijn in het geval van incidentele interactieve bewerkingen. Naarmate die bewerkingsketen langer, complexer en meer geautomatiseerd is, is meer datakwaliteit nodig. Dat is haalbaar door in de bewerkingsketen meer regie te voeren en meer standaarden en modellen te gebruiken. Samengevat: de fitness for use van data hangt af van het gebruiksdoel.

*De presentaties van deze dag zijn te vinden op [ruimteschepper.nl/kenniskring/delen-is-leren/kenniskring-presentaties](http://ruimteschepper.nl/kenniskring/delen-is-leren/kenniskring-presentaties)*

*Frans Rip, redacteur*

## Rond ALV internationale oriëntatie bij GIN?

Een Europese website geeft meer dan de 'cover' van een Duits boekje (van beroepsverband DVW, een GIN-tegenhanger).



Het onzekere handschrift is niet toevallig: '...for an authentic look, make sure to fill out your name with your non-dominant hand,' aldus is op de Engelstalige website onder het plaatje te lezen. [blogs.eu.eu/divisions/g/tag/surveying/](http://blogs.eu.eu/divisions/g/tag/surveying/)



Een andere website geeft deze 'first female president' van de Association of Ontario Land Surveyors (AOLS dus). Let op de ambtsketen met die superpenning met een driepoot, niet leeg, maar met daarop een modern meetinstrument!