

# NCG symposium - Hoezo datakwaliteit?

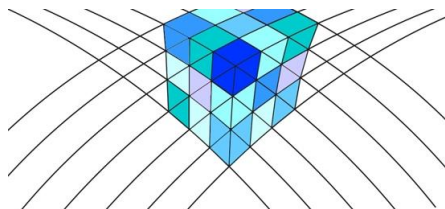
**Verslag 26 juni 2014**

Versie 1.0

## Contents

Inleiding .....	2
Kwaliteitshow .....	2
Pitches ochtend .....	3
Data kwaliteit in het onderwijs: Marien de Bakker/HAS Den Bosch.....	3
De Laan van de Leefomgeving & Hoezo Datakwaliteit?: Ton de Nijs/RIVM.....	3
Hoezo datakwaliteit? Smart Imagery Solutions.: Peter Joosten/Cyclomedia.....	3
Workshops ochtend.....	4
Wanneer is goed goed genoeg? Sytze de bruin Wageningen Universiteit.....	4
Wat doet kwaliteit in de keten? Huibert Jan Lekkerkerk/Informatiehuis Water.....	5
Wat is de kwaliteit van data uit meerdere bronnen? Renée Bekker/BIJ12.....	5
Pitches middag .....	6
Datakwaliteit onder hoge spanning: Jorrit Hansson/Enexis.....	6
Kwaliteit is mega subjectief: Henk Hoff/Open Street Maps.....	6
Linked Data: Marije louwsma/Kadaster.....	6
Workshops Middag .....	7
Wat gebeurt er met kwaliteit bij gebruik van schalen en projecties? Tjeu Lemmens/TU-Delft.....	7
Hoe beschrijf en classificeer je de kwaliteit van data ? Jandirk Bulens WUR/Alterra.....	7
Welke kwaliteit past bij bepaalde risico's bij gebruik van die data? Jaap Zevenbergen/Universiteit Twente.....	8
Slotpanel .....	8





## Inleiding

Op 26 juni 2014 organiseerde het Nederlands Centrum voor Geodesie en Geo-Informatica (het NCG) een symposium over kwaliteit van data bij Wageningen Universiteit & Researchcentrum. Zo'n 70 personen uit bedrijfsleven, onderwijs en overheid namen deel. Het symposium bestond uit een plenair interactief gedeelte, pitches vanuit aanbieders en gebruikers van data en een 6-tal workshops. Centraal stond de vraag wat de uitdagingen zijn als de kwaliteit van data in het geding is.

## Kwaliteitshow

Dagvoorzitter Karin van de Hoef-Reijnders (Directeur Werkomgeving bij de Belastingdienst) opende de show met een vraag aan het panel met Petra Loeff, Bart Beers en Tjeu Lemmens: "Wat is voor jullie de betekenis van datakwaliteit?". Petra Loeff, Programmamanager Omgevingswet bij het RIVM: "Voor de omgevingswet, bijv. de vergunning voor een dakkapel, zijn veel data nodig bij de gemeente. Al die data moeten voldoende kwaliteit bieden om de vergunningverlening goed te laten verlopen". Bart Beers, R&D Cyclomedia: "Nauwkeurige coördinaten zijn belangrijk, maar kwaliteit gaat verder". Bij het werken met beeldmateriaal in een actuele situatie is recent beeldmateriaal een voorwaarde. Tjeu Lemmens, Docent aan de TU Delft, geeft aan dat vanuit het onderwijs voor data inwinning kwaliteit een cruciale factor is die goed beschreven moet worden én compleet moet zijn positioneel, temporeel etc.. De kwaliteit van de input moet zich vertalen naar de kwaliteit van de output. Lemmens geeft aan dat hij in het onderwijs graag meer tijd aan kwaliteit zou willen besteden. Voor bijv. de toepassing van de 'kleinste-kwadratenmethode'<sup>1</sup> moet hij nu volstaan met in totaal 2 x 2 uur om dit onderwerp volledig te behandelen.

Kwaliteit in relatie tot de claimcultuur. Petra Loeff: Het is hier nog niet zo'n claimcultuur als in Amerika. Bij vergunningverlening zijn altijd wel mensen ontevreden, maar waar het om gaat is dat de beslissing stand kan houden bij de rechter. We zien ook dat goede kwaliteit zich uitbetaalt zoals bijv. bij luchtkwaliteit. Tjeu Lemmens: Vroeger was kennis over kwaliteit aanwezig bij onze rijksdiensten. Door de toename van inhuur en uitbesteding is veel van deze kennis verloren gegaan. Dit terwijl het bij hem eigenlijk gewoon 1<sup>e</sup> jaars stof is. Je ziet dat juristen tussen producent en gebruiker in gaan zitten, iets dat door de overheid dus geïntroduceerd is. Bart: geeft aan dat er aan bestekken soms eisen gesteld worden die in de praktijk niet waar te maken zijn en daarnaast heeft het bedrijfsleven ook het probleem dat het bij werving moeilijk is mensen met kennis over kwaliteit te vinden. Er is ook geen 'echte' opleiding meer op HBO/HAS waar kwaliteit voldoende onderwezen wordt.

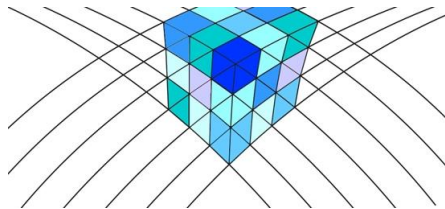
Uit de zaal vraagt Hans Versluis hoe je de gebruiker over kwaliteit informeert als je weet dat hij geen deskundige is. Bart Beers: Veel toepassingen (Apps) laten zien waar het fout (kan) gaan. Meldingen met die strekking worden in het algemeen weggeklikt omdat men dit 'gevaar' niet wil zien. Het effect is niet duidelijk, daar moet de aandacht naar uitgaan. Petra Loeff: breidt de vraag uit, Internet geeft makkelijker toegang tot alles wat rondom kwaliteit speelt. Een voorbeeld hoe dat fout kan gaan is de baarmoederkankervaccinatie campagne. Ontwikkelingen als Citizen Science op basis van input van de burger versus geavanceerde systemen kun je niet stoppen. De vraag is hoe gaan we hiermee om?

Uit de zaal onderschrijft Kyra van Onselen de noodzaak van kwaliteit en dat hulp daarbij geboden is, maar hoe krijg je de kennis daarover weer op peil? Hoe kan het onderwijs daarbij helpen? Tjeu Lemmens: Dit moet je kanaliseren, bijv. binnen GIMA waarin de universiteiten samenwerken in een geo-informatiekunde opleiding. Maar ook toepassing van de bottom-up benadering: je kunt experts van buiten niet vroeg genoeg de behoefte aan onderwijs in datakwaliteit in de opleiding laten benadrukken. Dan moet daarvoor natuurlijk wel geld beschikbaar zijn.

Arnold Bregt vraagt of er voorbeelden bekend zijn in Nederland waar kwaliteit issues rampzalige gevolgen hebben gehad voor het beleid? Tjeu Lemmens: Geen voorbeeld uit Nederland, maar wel uit

---

<sup>1</sup> De kleinste-kwadratenmethode is een rekenmethode om bij een gegeven verzameling puntenparen in het vlak uit een verzameling curven de "best passende" te bepalen (Wikipedia).



Italië: Bij verbreding van autowegen gebruikten kadaster en uitvoerder verschillende projecties, waardoor veel te smalle wegen werden gemaakt. Eenzelfde voorbeeld bij een landingsbaan van het vliegveld in Rome. Verschil in gebruik van WGS en de Italiaanse projectiestandaard leidde tot verschillen tot zeker 100 meter.”

Hiddo Velsink: Op de HAS in Utrecht wordt nog wel Geodesie onderwezen, maar dit is geïntegreerd in andere opleidingen. Ook is daar in maart een minisymposium geweest over omgaan met onzekerheid in de geodesie. Een belangrijke conclusie was dat, zeker als het gaat om deformatiemetingen (hoogtemetingen), er een grote behoefte bestaat aan verdergaande normering. Is dit een algemene tendens? Bart Beers: “Normen zijn belangrijk en geven ook richting aan goede beschrijvingen van kwaliteit. Kijk bijv. naar de standaarden van het OGC. Het nadeel van normen kan zijn dat innovatie belemmerd wordt. Er zijn nu voldoende normen beschikbaar, in ieder geval als het gaat over ruimtelijke kwaliteit.”

## Pitches ochtend

### Data kwaliteit in het onderwijs: Marien de Bakker/HAS Den Bosch

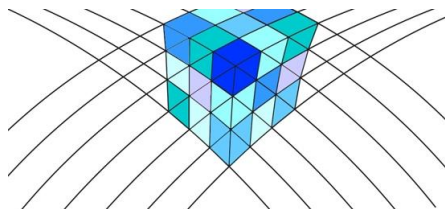
Het is lastig om concepten binnen het informatieproces voor het voetlicht te brengen. Datakwaliteit als begrip is ongrijpbaar. Kwaliteit betekent “fitness for use” Ervaring van de studenten is dat de vraag naar datakwaliteit vaak ondergeschikt is aan de vraag naar data. De opdrachtgever “neemt aan” dat het goed is, staat onder druk door tijdgebrek en daardoor gaat het nu eigenlijk meer om “tags” dan om officiële metadata. Men moet zich bewust zijn van het feit dat kwaliteit onzekerheid geeft. Dat heeft een relatie met de waarde van het onderzoek en dat geeft ook weer onzekerheid! Docenten zijn veelal onbekend met de begrippen, onderwijsbreed inbouwen is daarom een uitdaging. Concluderend: het onderwerp meer van academisch naar praktisch bewegen om ‘fitness for use’ beter zichtbaar te krijgen.

### De Laan van de Leefomgeving & Hoezo Datakwaliteit?: Ton de Nijs/RIVM

Het gaat om datakwaliteit, “Hoezo” moeten we zo snel mogelijk weglaten. Met de herziening van de Omgevingswet zien we een magistrale reorganisatie van de overheid voor het ‘gemak’ van de burger. Het gaat om data beschikbaar, bruikbaar en bestendig aan te bieden. Bruikbaar? Het voorbeeld van roet in de lucht op basis van verkoop van het aantal liters benzine laat zien dat Luxemburg er slecht uitspringt terwijl dat een onjuiste conclusie is; Bestendig? Uitvoering van het Nationaal Samenwerkingsprogramma Luchtkwaliteit (NSL) laat zien dat het aantal beroepsprocedures gedecimeerd is. De kwaliteitscriteria Gouvernance en beheer zijn belangrijk. Naast de 6 vragen die dit symposium aan de orde stelt is ook een extra vraag zeer relevant: Wat is er nodig in de Gouvernance om te waarborgen dat de kwaliteit van de data goed is? → Kan de slager zijn eigen vlees keuren? Kijk naar de processen rond luchtkwaliteit, maar ook bij natuur. Het filmpje van de Laan van de Leefomgeving staat nu op Youtube: <https://www.youtube.com/watch?v=OnyKLqkDniI> .

### Hoezo datakwaliteit? Smart Imagery Solutions.: Peter Joosten/Cyclomedia

De 360° foto’s van Cyclomedia worden voor metingen gebruikt en zijn metrisch nauwkeurig geografisch gerefereerd. Op deze wijze kunnen datasets zoals verkeersborden worden ingewonnen. Vanwege de hoge eisen die aan de nauwkeurigheid worden gesteld, moet de apparatuur aan hoge eisen voldoen en worden meerdere camera’s voor de opnames gebruikt. Kwaliteitscontroles nemen in het dataverwerkingsproces dan ook een belangrijke plaats in. Na dataprocessing worden vreemde resultaten en inconsistenties in de kwaliteitscontroles als indicatie gebruikt om deze data nader te beoordelen. De kwaliteit van het resultaat is sterk afhankelijk van meerdere factoren die gerelateerd zijn aan omgeving, atmosfeer, apparatuur en verwerkingssoftware (vaak ook van derden). Om die reden wordt aan ieder beeld kwaliteitsinformatie toegevoegd om daarover te informeren. De betrouwbaarheid wordt zo goed mogelijk gevisualiseerd: een groen (stoplichtje) is een indicatie van de betrouwbaarheid.



## Workshops ochtend

### Wanneer is goed goed genoeg? Sytze de bruin Wageningen Universiteit

Gestart werd met de kennismaking met de professionele gebruiker in de zaal. Verschillende deelnemers van EZ, Deltares, Alterra gaven aan dat ze professioneel met data bezig zijn. Kwaliteit speelt een belangrijke rol, hetzij met subsidies voor arealen grond, kustwaterstromingen en ontzanding of via crowd sourcing met een gemixte kwaliteit van waarnemingen.

Vervolgens is aan de hand van een aantal sprekende voorbeelden de vraag verbeeld wanneer kwaliteit goed genoeg zou moeten zijn. Bij precisielandbouw gaat het bij een fout al snel om hele rijen gewassen die sneuvelen bij bewerking. Bij het modelleren van onzekerheid in de geometrie gaat het om het doorrekenen van wat de gevolgen daarvan zijn. Voorbeelden zoals; zijn er inkomsten misgelopen, wat zijn de gevolgen als er calamiteiten optreden zoals de trekker in de sloot, worden er boetes opgelegd bij fouten in oppervlakten arealen voor subsidies, etc. Het voorbeeld van de loodinnname via de grond door kinderen in het Geuldal, geeft treffend weer dat de twee bronnen van onzekerheid, verontreiniging enerzijds en grondconsumptie anderzijds onevenredig bijdragen in de gevolgen. Een lage grondinname is al bepalend voor het effect. Veel investeren in verbeterde data over de loodverontreiniging is voor deze toepassing dan ook niet zinvol. Bij Citizen Science weegt de lagere kwaliteit van de data veelal ruim op tegen dure professionele data vanwege de grotere aantallen metingen, waardoor meetfouten lager zijn (bij onafhankelijkheid van metingen en geen bias), een betere gebiedsdekking geven, en afhankelijk van het doel goed passend kunnen zijn (bijv. bij early warning systemen).

**Samengevat is datakwaliteit geen intrinsieke eigenschap en sterk afhankelijk van het gebruik. Gevolgen bepaal je voor het gebruik, waarbij foutenmodellen, foutenvoortplanting en de verwachtingswaarde van informatie een belangrijk instrumentarium zijn.**

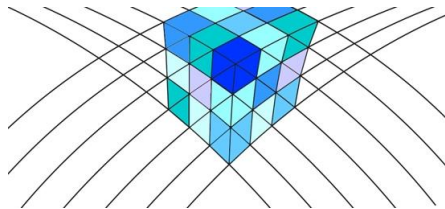
Discussie volgde rond een aantal stellingen:

- Datakwaliteit heeft imago probleem (te academisch, niet sexy, kun je niet mee scoren)
- De gebruiker kent de eigen behoefte niet
- De producent kent wensen/eisen gebruiker niet
- Rapportage van kwaliteit is te complex

In de discussie wordt aangegeven dat de opzet van de kwaliteitsbepaling belangrijk is. Hoe doe je dit Goed? Niet te sterk focussen op standaardafwijkingen, maar de vraag beantwoorden wat de aanvaardbare kwaliteit is. Een kwaliteitsstempel kan daarbij helpen. Geconstateerd wordt dat weinig met kwaliteitsinformatie wordt gedaan en ook dat pas achteraf discussie wordt gevoerd over kwaliteit wanneer iets fout gegaan is. Ervaring is ook dat wat de gebruiker op internet vindt als waar wordt aangenomen, de disclaimers en melding worden snel weggeklikt. Er zijn ook andere kwaliteitsaspecten dan alleen geometrie, zoals compleetheid, logische consistentie, temporele eigenschappen, waarbij het gebruik bepalend is. Ook kwalitatieve criteria zijn belangrijk, bijv. hoe kwantificeer je veiligheid. De grootste uitdaging ligt in kennisopbouw om de gebruikswaarde van dat goed in te schatten.

Aan de kant van het gebruik wordt onderkend dat communicatie een cruciale rol speelt, voorbeeld is communicatie over het weer, weeralarm, bodemdaling. Zeker bij onderwerpen die gevoelig liggen, of mogelijk paniek kunnen veroorzaken. Als onderzoeker/professional sta je dan machteloos als het over kwaliteit gaat. Zeker als het een juridisch spel wordt, er hoge kosten gemoeid zijn, stochastische berekeningen waarmee gemanipuleerd wordt, etc.

**Samenvattend is de vraag wanneer is goed goed genoeg lastig te beantwoorden. De zaal geeft blijk van de worsteling in welke richting dit gezocht moet worden, met daar bovenop het imago probleem dat datakwaliteit nu nog heeft. Eensgezindheid is er wel over de cruciale rol die communicatie over datakwaliteit speelt. Daar moet goed op ingezet worden.**



## Wat doet kwaliteit in de keten? Huibert Jan Lekkerkerk/Informatiehuis Water

Lekkerkerk werkt bij het Informatiehuis Water, een samenwerkingsverband. De rode draad in zijn presentatie over kwaliteit in ketens is de volgende: Men koppelt eigen informatie aan geometrie en dat hoeft niet altijd heel nauwkeurig. Men moet i.h.a. zijn eigen gebied kunnen vinden/herkennen. Voor beleidsbeslissingen is een dergelijke benadering echter te onnauwkeurig. Zo'n beleidsbeslissing is veelal gebaseerd op een groot aantal stappen in een ketenproces. In ieder stap wordt data gebruikt en toegevoegd en kan er in de kwaliteit iets mis gaan en als dat niet in de beschouwing meegenomen wordt kan er een advies komen dat niet heel betrouwbaar is. Diverse gebruikers in de keten kunnen ook weer leverancier zijn in diezelfde keten (loops). Lastig in dit soort processen is dat je een deel van de gebruikers kent en welke kwaliteit ze eisen en daar op kunt anticiperen, maar een groot deel ook niet. Belangrijk hierbij is om *fitness for use* en *conformance to requirements* in de beschouwing te betrekken. Ook de vraag hoe ver je moet gaan met het stellen van eisen, alleen kwaliteit data, of ook eisen stellen aan de wijze waarop de leverancier de data inwint. Binnen een keten werken niet alle partijen op dezelfde wijze samen en dan is het uitkijken als data gecombineerd wordt. Belangrijk is dan dat je de gehele keten moet kunnen overzien en dat ieder onderdeel in de keten moet weten wat de impact van de data is aan de input en output kant van de andere onderdelen. **De eerste stap in het proces is dus om de keten te overzien en te bepalen en welke kwaliteit nodig is om een betrouwbaar eindresultaat te krijgen.**

## Wat is de kwaliteit van data uit meerdere bronnen? Renée Bekker/BIJ12

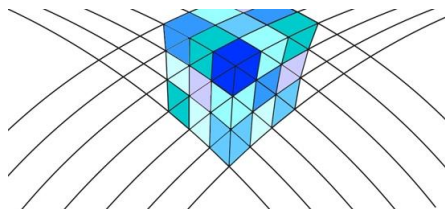
Renée Bekker stelt zich voor als Manager Kwaliteit en Ontwikkeling van de Nationale Databank Flora en Fauna (NDFF) bij BIJ12. In het eerste deel van de workshop vertelt zij hoe ze in de afgelopen jaren gebouwd heeft aan het optimaliseren van de kwaliteit van de NDFF. Zij vertelt daarbij heel beeldend **hoe de kleine modderkruiper de inrichting van Nederland beïnvloedt en hoe de 56 'Kansenkaarten' stand houden bij rechter.**

Wat speelt een rol in het kwaliteitsproces?

- Interne kwaliteitsborging in de hele keten: "U en ik herkennen madeliefjes altijd, anders is dat voor de bijv. de kleine modderkruiper". Er zijn dus 200 vrijwilligers die elke bijzondere inwinning valideren.
- Afhankelijk van de vraag die je stelt, is een heel hoge kwaliteit niet altijd nodig. Dat betreft vooral de meer zeldzame soorten die op de rode lijsten voorkomen en waarover gerapporteerd wordt aan de EU.
- Betrouwbaarheid is altijd van belang. Elke kaart die op Internet gepubliceerd wordt, wordt immers als waarheid gezien.
- De slager keurt niet het eigen vlees: er is een onafhankelijke commissie die de protocollen valideert.
- Hoe hoger de kwaliteit van de gegevens, hoe minder kans op onverwachte bouwstops.

In het tweede deel van de workshop stelt Renée een vraag aan de aanwezigen: In hoeverre heb je te maken met kwaliteitseisen en hoe ga je daarmee om? Algemene tendens:

- Het opstellen van protocollen en normen en het delen van ervaringen is op veel gebieden nog in de kinderschoenen.
- Crowdsourcing: kwaliteit is vaak beter dan je denkt. Open Street Map is een goed voorbeeld: Gebruik de data in combinatie met andere kaartlagen en zo kun je knelpunten ontdekken.
- Homogene data is vaak een illusie. Communiceren over de standaardafwijking helpt de gebruiker enorm.
- Kwaliteit kost geld, dat blijft een drempel.
- FIT FOR PURPOSE is niet hetzelfde als FIT FOR ANY USE.
- Duidelijke taal gebruiken als het over datakwaliteit gaat: **gebruik geen vakjargon.**



- Metadata zouden aangevuld moeten worden met kwaliteitstags. Veel beter te visualiseren en daardoor laagdrempeliger voor de gebruiker.

## Pitches middag

### Datakwaliteit onder hoge spanning: Jorrit Hansson/Enexis

Datakwaliteit onder hoge spanning. Voor een netbeheerder voor gas en elektriciteit is dit een uitdaging. Doelstelling is de bedrijfsvoering te optimaliseren ter ondersteuning van de core business: een continue en veilige energievoorziening. Herstel van storingen door voortdurend te bewaken, sturen en coördineren. Bij werken onder hoge spanningen is automatiseren een noodzaak. Slimmere netten en op afstand schakelen. Bijv. de jaarlijkse uitvalduur voor middenspanning van 12 naar 7 minuten terugbrengen beperkt de geleden schade enorm. Dit gaat niet zonder goede data, dus meer GIS en datakwaliteit. Netlogica (visueel) moet vertaald worden naar GIS logica zodat ook de computer ermee kan rekenen. **Maak fouten direct zichtbaar, beschrijf de fout, voorkom het negeren van foutmeldingen, maak objecten met een fout duidelijk zichtbaar, een rood stoplichtje en manage het proces.**

Concluderend zijn de ervaringen van Enexis in "tegeltjestaal":

- **Het leveren van kwaliteit kost geld. Het niet leveren van kwaliteit kost kapitalen.**
- **Een berg verplaats je steen voor steen.**

### Kwaliteit is mega subjectief: Henk Hoff/Open Street Maps

OSM is een wikipedia voor geodata: Open, Vrij, Gratis; en voor iedereen. De professional reageert al snel Oh, als dat maar goed gaat! Wat gestart is als data die ook interessant zouden zijn voor de verzamelaar is in de loop van de tijd enorm gegroeid in kwaliteit. In Nederland is dat snel gegaan toen AND (Automotive Navigation Data) de data voor OSM beschikbaar stelde en in een korte slag een landsdekkend beeld is ontstaan. De groei is echter gebleven, nu wordt dat gezocht in uitbreidingen voor fietspaden, huizen etc.

Kwaliteit is erg subjectief. Bij OSM is nauwelijks discussie over kwaliteit. Als ernaar gevraagd wordt, is het eenvoudige antwoord neem een sample en controleer die. Kwaliteitseisen worden ook niet geformaliseerd. Neem rolstoelgebruikers die stellen eigen eisen. Uitgangspunt is dat de gebruikers dat zelf prima kunnen doen. Je gebruikt de crowd voor wat ze zelf belangrijk vindt.

De kwaliteit meten? Dat loopt via de gebruikers. Bijv de App bouwers, die koppelen de fouten van gebruikers zelf weer terug. Je hoeft niet bang te zijn om via editen fouten te introduceren. De praktijk wijst uit dat dat meevalt en bovendien edits kunnen eenvoudig weer worden teruggedraaid. Het is ook een bewustwordingsproces en het gaat om vertrouwen. Dat is belangrijk voor de kwaliteit. Gebruik OSM waardoor je zelf de afwijkingen ziet, beschouw het als referentiemateriaal.

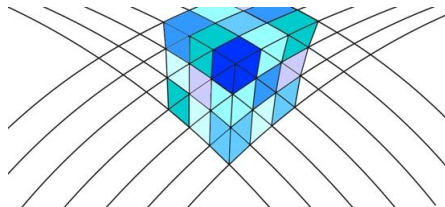
Samenvattend:

#### **Kwaliteit is:**

- **mega subjectief;**
- **leg uit;**
- **de doelgroep bepaald!**

### Linked Data: Marije louwsma/Kadaster

Vult een presentatie door uitval wegens ziekte. Ze laat een filmpje zien over Linked Data als een van de resultaten van de Pilot Linked Open Data (PilotOD) <https://www.youtube.com/watch?v=eWqRB-HFxI>



## Workshops Middag

### Wat gebeurt er met kwaliteit bij gebruik van schalen en projecties?

Tjeu Lemmens/TU-Delft

Lemmens ging in op de data inwinning, de beschrijving van de keten van inwinning tot product en de geometrische kwaliteit. In de klassieke situatie was de producent degene die de kwaliteit bepaalde voor de gebruikers. Momenteel moet er een goede relatie tussen producent en gebruiker zijn om de kwaliteit van de benodigde data te bepalen (fitness for use). Daarbij is het belangrijk dat de gebruiker inzicht moet hebben in wat kwaliteit van data is. De vraag is wat de kwaliteit van data bepaalt. Lemmens legt uit dat het begint bij het op te nemen object, hoe scherp is dat te definiëren en moet dat gedefinieerd worden. Uit deze analyse volgt de inwinningsmethode. Kwaliteit is actualiteit, tijdsaspect, thematische nauwkeurigheid, volledigheid van attributen, compleetheid. Lemmens gaf enige voorbeelden van gebruik van gecombineerde data van verschillende kwaliteit en wat de gevolgen kunnen zijn als bij het advies of conclusie geen rekening gehouden wordt met de kwaliteitsverschillen.

**Bij geen kwaliteit dan kun je op termijn op hoge maatschappelijke kosten rekenen.**

Een andere les die meegegeven werd, is dat **om kwaliteit te kunnen begrijpen, er een noodzakelijk inzicht in de inwinningstechnieken en de te doorlopen verwerkingsstappen nodig is.**

### Hoe beschrijf en classificeer je de kwaliteit van data ? Jandirk Bulens

WUR/Alterra

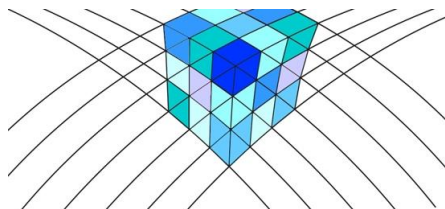
Als introductie wordt een overzicht gegeven over wat is kwaliteit, welke standaarden zijn er en kwaliteitscertificering. Er zijn taalkundige en technische definities maar het ligt voor de hand om die van ISO 8402 te gebruiken die specifiek over termen en definities gaan: *'Kwaliteit is het geheel van eigenschappen en kenmerken van een product of dienst dat van belang is voor het voldoen aan vastgestelde of vanzelfsprekende behoeften'*. Het belang geeft aan dat kwaliteit pas betekenis krijgt door de context waarin het wordt gebruikt, vaak uitgedrukt als **fit for purpose** of **fitness for use**.

Er zijn standaarden voor kwaliteit voor organisaties (ISO9001), software (ISO 9126) systeemsoftware, hardware, zelfs de fysieke locatie en gegevens. We beperken ons tot de laatste waarvoor **ISO 19157** de standaard is voor de ruimtelijke data. Deze standaard geeft aan welke kwaliteitselementen je hebt voor het beschrijven van kwaliteit vanuit de producent of de consument bekeken.

Het vaststellen van kwaliteit is nodig om te kunnen vergelijken en te kunnen selecteren van geschikte data voor het bedoelde gebruik. **Volledige kwaliteitsbeschrijvingen bevatten volgens ISO 19157 elementen in de categorieën: Completeness, Logical consistency, Positional accuracy, Thematic accuracy, Temporal quality en Usability.** Het toetsen met een kwaliteitslabel wordt nog niet veel gedaan. Bij Alterra is in het kader van softwarekwaliteit voor modellen al vrij lang (de huidige versie sinds 2007) een systeem voor kwaliteitsstatustoekenning in gebruik. De praktijk leert dat met name modellen hieraan getoetst worden, bij data is dat beperkter maar vooral bij wettelijke en ondersteunende taken Natuur (WOT-Natuur) wordt dit model gebruikt. In essentie wordt bij status A toekenning inhoudelijk/administratief getoetst, of aan documentatie eisen wordt voldoen, er getest en gevalideerd is en het beheer geregeld is.

In de workshop is 'gemindmapt' over nut en noodzaak van standaarden, toepassen van het principe Fitness for use en het hoe en waarom van accreditatie voor kwaliteitscertificering. Uit de discussie komt op hoofdlijnen naar voren dat communicatie belangrijk is naar de gebruiker (met een range van de gewone burger t/m de opdrachtgever). In de communicatie zijn standaarden te vertalen naar een bijsluiter en moet je goed de gebruiker kennen, vgl. de presentatie van het weerbericht.

Het toepassen van standaarden is maatwerk, je kan kiezen voor een beschrijving in de vorm van een standaard versie, meer algemeen toepasbaar en een maatwerkversie die toegesneden is op het specifieke gebruik op basis van de criteria die de gebruiker is gevraagd aan te geven. Evaluatie en beheer zou een vast onderdeel moeten zijn in de beheercyclus van data. Standaarden moeten ook vooral



bruikbaar zijn, kijk daarvoor naar wat al gebruikt wordt door gebruiker of community en baseer je op bestaande ervaringen en best practices die bepalen al vaak welke standaarden de facto geaccepteerd zijn. Voor accreditatie zou je een soort van prestatieladder moeten gebruiken voor een beter begrip, vergelijk dit met wat voor bijv. met het energielabel bij wasmachines wordt gedaan.

## Welke kwaliteit past bij bepaalde risico's bij gebruik van die data? Jaap Zevenbergen/Universiteit Twente

Jaap Zevenbergen is Professor Land Administrations Systems bij het ITC in Enschede. Centrale vraag die hij in deze workshop stelt: "Welke kwaliteit past bij bepaalde risico's die gebruik van de data met zich meebrengt?" Vrij vertaald naar de data-aanbieder: "Wanneer ben ik aansprakelijk?"

Vragen die spelen zijn:

- Mag je oplettenheid verwachten van de gebruiker van open data?
- Had je kunnen voorzien dat iemand er iets mee zou gaan doen dat tot dit ... bijv. ongeluk zou leiden?

Kyra van Onselen van Rijkswaterstaat geeft als voorbeeld dat RWS Waterdieptekaarten beschikbaar stelt. Stel dat iemand deze gaat gebruiken voor navigatie van een zeilboot en vast komt te zitten? Is RWS dan aansprakelijk?

Jaap verwijst naar <https://data.overheid.nl/handreiking/stap2d/aansprakelijkheid> waar op het Open Dataportaal van de overheid een artikel is opgenomen: "Aansprakelijkheid voor gebrekkige Open Data". Het gaat hier om stap 2: Doe de juridische check. Wat als deze niet goed is uitgevoerd? Hoe zit het dan met aansprakelijkheid? **Doe de check dus goed!**

### Ontwikkelingen:

- Gaan snel, de maatschappij is in beweging, dit is een momentopname.
- Claimcultuur groeit: niet geschoten is altijd mis
- Reputatieschade voorkomen
- Bij goed zoeken, krijg je (deels) tegenstrijdige data
- Crowdsourcing kan zelfreinigend werken, maar altijd vertraagd
- Big data middelt en middelt en zoekt correlatiepatronen: vraagt een andere benadering van kwaliteit.

Zijn er al claims in de VS die een relatie hebben met het gebruik van open data? Antwoord: Niet voor zover bekend, maar in de VS levert de overheid veel minder en veel minder uitgewerkte open data!

Zou de wetgever niet over datakwaliteit moeten spreken? Maak de relatie van de kwaliteit van basisregistraties bijv. duidelijk in relatie tot het begrip FIT FOR PURPOSE.

Organisaties beginnen zich te realiseren dat de boete op het niet leveren van open data lager kan zijn dan schadeclaims die vergoed moeten worden bij verkeerd gebruik. Deskundigen gaan er teveel vanuit dat gebruikers weten waar je de data voor kunt gebruiken. Mensen varen steeds meer blind op hun systemen. Misschien een idee om een proefproces uit te lokken met de dieptegegevens van RWS?

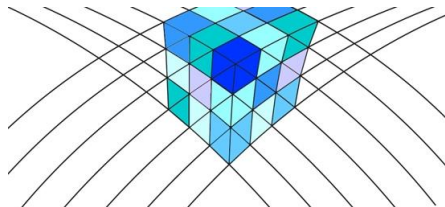
### Communicatie en interactie over kwaliteit van open data is heel wezenlijk voor een verantwoord gebruik.

De vraag die dat oproept is hoeveel moeite kun je verwachten van datapublicerende partijen? Moeten we naar een datakeurmerk toe?

## Slotpanel

Uit de workshops zijn een aantal belangrijke zaken naar voren gekomen die aan de orde gesteld werden in een panel bestaande uit Bart Beers (bedrijfsleven), Jantien Stoter (TU-Delft/wetenschap), Petra Loeff (overheid) en Frans Backhuijs (overheid/VNG, bestuurder/beslisser). Jantien opende als panellid met de





opmerking dat ze blij verrast is door het grote aantal deelnemers bij een toch zo “droog” onderwerp als datakwaliteit, waarmee aangegeven is dat het onderwerp wel degelijk in de belangstelling staat

De discussie werd geopend met een vergelijk van de presentatie van het weerbericht. **Zou het gebruik van eenvoudige communicatie en het gebruik van standaarden datakwaliteit niet beter tot uitdrukking brengen?** Petra stelt dat dat afhangt van de vertaling van het begrip standaard. Niet iedere toepassing hoeft dezelfde kwaliteit te hebben. Er zijn verschillende toetsingsinstrumenten die hun eigen eisen stellen. Een oplossing om hiermee om te gaan is wat eenvoudig kan, ook eenvoudig te houden.

Bart geeft aan dat criteria wel helder moeten zijn, het gaat om het vergelijken van data. Welke criteria wordt door de gebruiker bepaald, die selecteert en kiest op basis van fit for purpose. Standaarden moeten op data gebaseerd zijn en niet op het proces, omdat dat de innovatie hindert. Er is geen eenduidige definitie van kwaliteit. Het is afhankelijk van waar je behoefte naar uitgaat, hoe relevant is de kwaliteit voor je eigen behoefte.

Frans: Bestuurders praten i.h.a. anders over kwaliteit dan professionals. De bestuurders vanuit wat nodig is om een maatschappelijk probleem op te kunnen lossen en de professionals vanuit het vakgebied.

Jantien: Wat duidelijk is, is dat je standaarden nodig hebt om de kwaliteit te kunnen meten. De gebruiker zou dit ook willen doen zonder kwaliteit echt te benoemen. Maar hoe doe je dat? Termen visualiseren?

In de workshops is eveneens geconstateerd dat er nog veel open vragen zijn en is er de worsteling kwaliteit te duiden. Datakwaliteit heeft ook een imagoprobleem, zoals ook uit de tegeltjes uit de pitch van Enexis bleek. **Wat voor soort fouten in de gegevens kosten nu eigenlijk kapitalen? Zijn daar voorbeelden van te vinden in Nederland?**

Tjeu kent geen Nederlands voorbeeld maar kijkt wel naar Zwitserland. Daar is het maken van kaarten tot kunst verheven en wordt daarom ook anders gewaardeerd. Vinden we hier het beheer van data belangrijker, het zijn daar de cartografen die een prominente rol spelen bij het aanbieden van informatie.

Petra: Harmonisatie van data en definities is erg belangrijk. Informatie is een abstractie van de werkelijkheid. Veelal is er een achtergrond waardoor deze zaken verschillend zijn. Goed voorbeeld is *het adres*. In de BAG heeft *het adres* een andere definitie dan in de basisregistratie personen. Dit is de burger niet goed uit te leggen. Ook hier geldt: wees consistent en duidelijk in je specificaties, vergeet niet dat de gebruiker vraagt en bepaalt!

#### **Hoe precies heb je de data nodig om je eigen werk zo goed mogelijk te kunnen doen?**

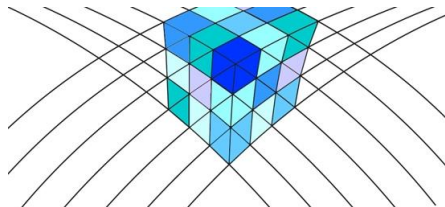
Open data is een stimulans voor innovatie en werkgelegenheid (min EZ) Probleem bij de overheid is veelal dat de maatschappelijke kosten van data ergens anders vallen dan de revenuen ervan. Het beschikbaar stellen van open data door de overheid heeft als voordeel dat innovatie met die data sneller verloopt, het nadeel kan zijn dat de innovatie voor de inwinning zal verminderen omdat de overheid de kosten van de inwinning voor de gebruikers vergoed/betaald. Dit is een politiek discussie die rijksbreed moet worden gevoerd.

#### **Wie heeft het overzicht van alle onderdelen van de keten?**

Petra: Dit speelt ook bij de Omgevingswet. Met de nieuwe opzet moet het door verdergaande integratie in de systemen te bouwen ook het overzicht komen. Er wordt geconstateerd dat er wel steeds meer duidelijkheid en begrip komt. De vindbaarheid van data wordt groter. Jantien geeft aan dat dit een proces is dat al meer dan 10 jaar geleden startte, veel is nog onbekend, maar vandaag is er het NGR, PDOK waardoor meer overzicht is gecreëerd. Geldt ook voor het opzetten van de basisregistraties en daardoor wordt het ook steeds makkelijker. Bij Geonovum wordt overzicht gecreëerd door ook de concepten van alle informatiemodellen in beeld te brengen.

#### **Er is vandaag ook geopperd om een proefproces uit te lokken. Is dat zinvol?**

Frans: de vraag is of de overheid zich realiseert wat de juridische gevolgen kunnen zijn bij open data. Dit



moet nader uitgewerkt worden. In een proces moet aangetoond worden dat de zorgplicht goed ingevuld is, dat is geen garantie dat alles klopt. Er moet gecommuniceerd worden dat niet alles 100% kan/moet zijn. Men moet leven met een zekere onzekerheid. De overheid zal de zorgplicht steeds meer gaan beleggen bij de makers van de data.

#### **En de consequenties voor datakwaliteit?**

*Petra:* De gebruiker/burger zit dubbel in de discussie over kwaliteit en openheid, enerzijds wil de overheid meer vrijheid bieden, maar zodra het met de vrijheid fout gaat de overheid dat helemaal dicht regelen, terwijl ze dat eigenlijk niet moet doen.

*Kyra:* *Als je weet dat de burger de bijsluiter niet leest kan dat toch niet anders?* Kijk naar de nieuwe Omgevingswet. Daar wordt voorzien in een "voorwasstraat" bedoeld voor een eerste screening, wat wettelijk bepaald is zal daarna nog moeten worden bepaald, bijv de bijdrage aan de luchtkwaliteit bij vergunningverlening. De valkuil van de overheid is dat ze niet moet doorschieten in de zorgen van de samenleving. Vergelijk de auto-industrie, een auto gaat een keer stuk.

Afsluiten er is een informatieparadigma: meer en betere data en modellen, waarom kan dat niet altijd gebruikt worden. Er is nog veel onduidelijk, misschien kan een proefproces wel helpen. En pas op voor de valkuil schieten we niet te veel door (door de claimcultuur?). Dit roept om een vervolg... NCG, Geo Samen, ...of?

Met de paneldiscussie kwam een eind aan deze dag over ruimtelijke datakwaliteit. Onder de borrel werd teruggekeken op een geslaagde dag over dit onderwerp.