

Verslag van de Kennisarena kwaliteit van ruimtelijke data op 31 oktober 2022

Aanwezige genodigden:

- Douwe Blanksma, Ruimteschepper
- Gerard Nienhuis, Provincie Overijssel: GIS-adviseur, met onderwerpen als data governance, data management en data-driven toepassingen.
- Richard Witmer, Kadaster: verantwoordelijk voor de basisregistraties Topografie, 3D en Adressen en gebouwen.
- Martin Peersmann, Ministerie van Binnenlandse Zaken: programmamanager Basisregistratie Ondergrond.
- Damir Brnobic, Ministerie van Binnenlandse Zaken: directie ruimte en leefomgeving, BAG en adviseur basisregistraties met tevens kwaliteit binnen zijn taken.
- Andries Knotters, Rijkswaterstaat: Centrale informatievoorziening RWS, werkzaam in het ecologische werkveld.
- Jasper Snippe, Rijkswaterstaat: sector Grond- weg en waterbouw; BRO coördinator RWS, voor data koppelingen, verplichtingen en hergebruik wat betekent dat voor de kwaliteit.
- Henk Janssen, NEO: Gemeenschappelijk Landbouwbeleid (GLB) voor not-for profit organisatie (overheid) waar datakwaliteit een belangrijke rol speelt.
- Marien de Bakker, HAS green academy Den Bosch: data wat kan je ermee doen en wat betekent dat voor de kwaliteit, ook vanuit een beleidsperspectief.

Aanwezige leden van het Expertisecentrum Kwaliteit van Ruimtelijke Data (EKRD):

- Wies Vullings, teamleider team Applied Spatial Research en onderzoeker ruimtelijke informatie
- Maarten Storm, onderzoeker kwaliteit van ruimtelijke data
- Gerard Hazeu, onderzoeker GIS en remote sensing
- Jandirk Bulens, onderzoeker ruimtelijke informatie
- Martin Knotters, onderzoeker en practicumdocent statistiek
- Victor Mensing, informatie manager WOT Natuur & Milieu

Programma:

12.30	13.00	Inloop
13.00	14.30	Genodigden aan het woord, WUR-medewerkers luisteren vooral
14.30	14.45	Pauze
14.45	16.00	Genodigden en EKRD-ers samen in gesprek
16.00	17.00	Borrel
17.00	17.15	Afronding en vertrek

Waarom een Kennisarena:

Waarom organiseren we deze middag? Een aantal jaar geleden zagen we dat binnen onze organisatie verschillende onderzoekers met het onderwerp ruimtelijke datakwaliteit bezig waren en door elkaar op te zoeken wilden we gebruik maken van elkaars sterke kanten en van elkaar leren. We hebben toen het Expertisecentrum Kwaliteit van Ruimtelijke Data opgezet om gezamenlijk aan projecten te werken en de buitenwereld te laten weten dat we expertise hebben op dit gebied. We hebben met een aantal van jullie erg leuke en interessante projecten uitgevoerd waarin het *fitness for use principe* in kwaliteit van ruimtelijke data centraal stond. We hebben in en door de projecten die we gedaan hebben een methode ontwikkeld om fitness for use principe praktisch toepasbaar te maken. Centraal staat hierin dat we specifiek in de toepassing van die data de relevante kwaliteitselementen benoemen en waar nodig een minimumkwaliteitswaarde gezamenlijk vaststellen.

Nu staan we op een punt dat we een stap verder willen zetten, maar om de richting te bepalen hebben we input nodig. Aan de ene kant zijn we benieuwd naar of degene waar we een project mee uitgevoerd hebben wat hebben gehad aan het project en het resultaat. Heeft het jullie verder geholpen? Of niet en zo nee, waarom dan niet, wat miste er?

En aan de andere kant zijn we in bredere zin erg geïnteresseerd in waar organisaties/mensen tegen aan lopen met betrekking tot kwaliteit van data. Datagestuurde werken is een kreet die je veel hoort en dan denken wij direct:

- Welke rol speelt de kwaliteit van data hierin?
- Wanneer durf je als organisatie op de data te vertrouwen?
- Hoe check je of de data waar je belangrijke beslissingen op baseert goed genoeg is?
- Hoe bepaal je wat goed genoeg is?
- Is het nuttig en waardevol om de kwaliteit van data te kennen? Doet u iets met die kennis?

Er zijn weinig groepjes/fora die zich met dit thema bezighouden en waar we dit soort vragen kunnen stellen, dus dachten we dat we dat zelf dan maar moeten organiseren. Het resultaat is de kennisarena.

Wat is een Kennisarena?

De kennisarena bestond uit een tweetal rondes. In de eerste ronde zaten alle genodigden van buiten WUR in een binnenste ring en de medewerkers van het EKRD zaten in een buitenring, zonder zich in het gesprek te mengen. Na een pauze begon de tweede ronde waarin juist de mensen van EKRD aan het woord kwamen om vragen te stellen aan de genodigden.

Na een korte introductie waarin iedereen aangaf wie hij was en waar hij werkte en wat de link met datakwaliteit was, begon de eerste ronde.

Verslag van de Kennisarena: eerste ronde

Al snel kwam aan de orde waarom datakwaliteit maar beperkt aandacht krijgt in het geo-werkveld. Als we kijken naar de financiële wereld, dan is daar juist wel heel veel focus op datakwaliteit. Jasper Snippe stelt: "Hoe hoger de kwaliteit, hoe breder te gebruiken".

Martin Peersmann zou graag antwoord zien op de vraag hoe data van verschillende kwaliteit met elkaar te combineren is, zoals nu bijvoorbeeld in digital twins waar kwaliteit intern verschillend kan zijn omdat verschillende modelleringen gecombineerd worden. Afhankelijk van het doel wordt ook

een andere kwaliteit gevraagd. Er wordt nog weinig hierin gedaan terwijl Peter Burrough in het verleden van ruimtelijke datakwaliteit al een onderzoeksonderwerp heeft gemaakt en dan met name over 'rekenen aan onzekerheid'.

Gerard Nienhuis zou graag zien dat op geautomatiseerde wijze van elk data- en metadatabestand er een kwaliteitsbeschrijving gegeven kan worden op een objectieve en erkende manier. Bij de provincie zijn meer dan 1500 datasets waarvan de metadata veel verschillen laten zien. Bij iedere data zit metadata, anders mag het niet beschikbaar komen in het systeem. Ook met contactpersonen erbij. Die spreken we erop aan als kwaliteit van (meta)data niet in orde is. Niet meer dan 2000 tekens in de samenvatting!

Martin Peersmann geeft aan dat het met AI / neurale netwerken mogelijk is om zaken in data als 'verdacht' aan te merken. Damir Brnobic meldt in dit kader een 'rodevlekkenkaart' om inzicht te krijgen in gebieden / objecten waar iets aan de hand kan zijn met de kwaliteit.

Martin Peersmann geeft aan dat digital twins een ultieme test zijn van de kwaliteit. Als de data niet goed genoeg is kan er niet gerekend worden.

Jasper geeft aan dat het temporele aspect ook van belang is, wanneer is data nog 'geldig'? Tijdstempels of houdbaarheidsstempels zijn noodzakelijk om data bruikbaar te maken.

Marien geeft aan dat het belangrijk naar de praktijk, het gebruik. Kan bijv een gedeputeerde beter uit de voeten als hij goed geïnformeerd is over de kwaliteit en bestaat er eigenlijk een generieke kwaliteit?

Richard geeft aan dat voor basisregistraties geldt dat er naar de grootste gemene deler wordt gekeken en de kwaliteit zo generiek mogelijk wordt besproken. Er is wel een verschil tussen basisregistraties en andere data. Dat wordt beaamd, maar op zich maakt dat voor de bruikbaarheid geen verschil, want dat wordt bepaald door de toepassing.

Martin Peersmann: In de BRO wordt gewerkt met handboeken en protocollen. Uit het onderzoek met WUR over datakwaliteit van sonderingen kwam bijvoorbeeld naar voren dat outliers er niet direct uit gehaald zouden moeten worden, maar juist ook iets zouden kunnen duiden. Het is dan belangrijk om een 360graden scan te doen en andere bronnen te gebruiken om na te gaan wat daar aan de hand is.

"RWS deed controle op BRO. De ene gemeente levert continu gegevens, andere niet? Hoe komt dat?" Dat moet van invloed zijn op de kwaliteit van de dataset ter plaatse... Richard geeft aan dat er twee perspectieven zijn: Top down zoals bij basisregistraties en bottom up: als gebruiker moet kiezen welke data geschikt zijn dan moet je kunnen vertrouwen op de kwaliteitsinformatie die beschikbaar is. Dat wordt nog lastiger als de kwaliteit 'gemixt' is en delen een ander kwaliteitsniveau hebben. Ook geeft hij aan dat we als Nederland internationaal erg voorop lopen wat betreft de organisatie en kwaliteit van data. Ook binnen Nederland kunnen we van elkaar leren. Is het bijvoorbeeld mogelijk dat de provincies 'good practices' van RWS overneemt?

Douwe stelt de vraag of het niet zo is dat iets alleen relevant is als er ook geld voor is?

Vanuit de kring wordt aangegeven dat dingen in ieder geval ook gebeuren als ze belangrijk gevonden worden, het hangt ook af van hoe het wordt gepresenteerd, hoe je het opschrijft. Sluit ook aan op wat eerder gezegd is. Dossiers kennen een grote dynamiek in de vraag naar data. Die dynamiek komt ook terug in de protocollen en voorschriften die soms zelfs jaarlijks worden ge-update. Ook daarvoor geldt dat er per sector weer verschillende inzichten bestaan.

“Kwaliteit is voor een moment/toepassing, niet voor de eeuwigheid”, met de kennis van *dat* moment. Met nieuwere inzichten kan de kwaliteit toch lager zijn dan vooraf ingeschat... Jasper Snippe geeft aan dat RWS al heel lang databases opbouwt. In eerste instantie vooral zelf, maar tegenwoordig wordt veel werk uitgevoerd door derden. Er bestaan al 20 jaar voorschriften, die nu in contracten met derden worden gebruikt en waarin is vastgelegd hoe de data opgeleverd moet worden aan RWS.

Douwe vraagt of er wel genoeg gedaan wordt aan data?

Bij RWS wordt data gezien als een asset, maar dat leeft nog niet zo bij de bestuurders. Hoe krijgen we kwaliteit top of mind? Door het wettelijk vast te leggen. Maak data een issue bij grote projecten, door bijv een data manager als functie te benoemen. Martin voegt toe dat nu bodem en water sturend zijn bij inrichtingsvraagstukken, omdat de faalkosten niet meer te betalen zijn.

Er wordt gesteld dat bij Nederlandse bestuurders er in het algemeen nog onvoldoende bewustzijn is m.b.t. kwaliteit van data. De nieuwe generatie die nu opgeleid wordt, is veel meer bewust hiervan.

Richard geeft aan dat het zorgen voor datakwaliteit ingewikkeld is omdat het preventie is, dat maakt het moeilijk om het belang onder de aandacht te brengen.

Douwe vraagt wanneer het goed genoeg is? Hoe gaat dat bijv bij levering van data door marktpartijen?

Henk Janssen geeft aan dat een datalevering een koppelpunt is tussen twee partijen: de klant en de producent die beiden een eigen verantwoordelijkheid hebben. Als producent natuurlijk verantwoordelijk voor kwaliteit, maar wel met een onzekerheidsmarge (vastgelegd in bijsluiters). Als producent lever je data, maar weet je vaak niet wat de klant ermee doet, dat is buiten het gezichtsveld. Voor de door de producent bepaalde kwaliteit wordt vaak gekeken naar de actualiteit, nauwkeurigheid en volledigheid, maar niet op bruikbaarheid. Bruikbaarheid wordt bepaald door de vragende, ontvangende partij. Of de klant de kwaliteit echt meet, daar is geen zicht op.

80/20% regel: 80% van data is met goede kwaliteit te leveren, 20% is moeilijk of duur met goede kwaliteit te leveren.

Jasper Snippe stelt dat hoe hoger de kwaliteit, hoe meer hergebruik / hoe hoger de hergebruikswaarde. Hergebruik wordt bij RWS niet bijgehouden, maar daar wordt nu een begin mee gemaakt. Er wordt nu bijgehouden hoe vaak bestanden worden gedownload. In contracten voor aanleveren van data aan RWS kun je al vastleggen dat de data geschikt moet zijn voor hergebruik. Beleid moet dit oppakken en moet ook leiden tot dat meer bewust en bekwaam wordt omgegaan met informatie. Nu is nog steeds veel overtuigingskracht nodig om het beleid /de politiek goed te informeren. Data die veel wordt gebruikt moet beschikbaar komen in meerdere smaken. Data die niet meer wordt gebruikt, daar hoeft je geen energie meer in te steken als leverancier.

Martin Peersmann: Bij de BRO zie je dat het gebruik van de data enorm is toegenomen ten op zichten van voor de BRO (bijv het gebruik van het bodembestand van WENR). Er wordt dan ook veel aan marketing gedaan binnen de BRO.

Jasper geeft aan dat er een generatiekloof is. Komende generaties zullen veel digitaal ingesteld zijn. Over twintig jaar, als de jonge mensen van nu bestuursfuncties hebben, zijn de problemen van nu geen issue meer. Hoe hoger de kwaliteit, hoe hoger de hergebruikswaarde. Hoe wordt die hoge kwaliteit herkend?

Gerard Nienhuis geeft aan dat kwaliteit op alle niveaus gerespecteerd moet worden. Het gebruik van dashboards helpt om de omvang van hergebruik inzichtelijk te maken en daarmee dus de kwaliteit inzichtelijk te maken voor alle niveaus. Het helpt dan om mensen te overtuigen van het belang van datakwaliteit en om bijvoorbeeld hier meer mensen voor in te zetten. Een dashboard helpt om het belangrijk te maken. Je moet wel uitkijken dat het niet doorslaat en dat je hierdoor niet teveel aandacht op specifieke gevallen vestigt. Het gaat meestal niet om dat ene geval, maar juist over de grote aantallen en de trends die je daarin ziet.

Douwe vraagt hoe integraal kwaliteitsbesef naar boven te halen is?

Martin Peersmann geeft aan dat je met kwaliteit goed moet kunnen zien dat het klopt. Grote ongelukken, bijvoorbeeld met het graven, hebben wel geleid tot meer aandacht voor kwaliteit, zoals het opzetten van het KLIC systeem, maar het aantal schadegevallen neemt niet meer af. Inzicht in falen blijkt dan in de praktijk niet te helpen. Locatienauwkeurigheid wordt nog te veel gefocust op x-y-nauwkeurigheid. Men vergeet te vaak de Z-nauwkeurigheid die zo belangrijk is bij kabels en leidingen. Eerder werd gezegd dat we het als Nederland al zo goed doen, maar realiseren we ons wel dat Nederland maar voor 1/3 deel op land ligt en dat 2/3 deel van ons grondgebied in zee ligt? Het deel in de zee is veel minder goed vastgelegd, maar is wel erg van belang en wordt ook steeds belangrijker geacht. Als voorbeeld van de relatief lage kwaliteit/gebruikswaarde geldt dat we bijvoorbeeld veel verschillende kustlijndefinities hebben, wat zich vertaalt in mogelijke fouten bij ruimtelijke projecten (zoals bijvoorbeeld de Brouwersdam).

Jasper: Nederland is natuurlijk wel maar een klein stukje in het geheel en binnen de EU wordt heel verschillend met kwaliteit omgegaan. Daarin moeten we ook bewust worden dat grenzen geen rol moeten spelen. Het is bijvoorbeeld gebleken dat het niet mogelijk is om door een Belgisch bedrijf geproduceerde data bij het onderzoek in de Hedwige polder op te nemen in de BRO.

Tweede ronde

Vraag Jandirk (aan Damir): Hoe gaan we realiseren dat semantiek, gebruik van ontologieën en 'Controlled' vocabularies worden gebruikt?

Bij het combineren van data moeten ook de kwaliteitsnormen worden gekoppeld en meegenomen worden in de metadata. Wordt de hoogste of laagste kwaliteit leidend? Hoe doe je dat? Er zou meer aan educatie van gebruikers moeten doen.

Er is structureel te weinig geld beschikbaar voor een heel aantal bestanden, bijv bestanden bij WUR en TNO daar krijgen de instituten nog steeds hetzelfde bedrag voor als jaren geleden. Dat bedrag zou verhoogd moeten worden zodat daar ook ruimte is voor onderzoek naar kwaliteit.

Vraag van Victor: Gaat openstellen van data helpen de kwaliteit te verhogen?

Wel als gebruik enorm toeneemt en daarmee bruikbaarheid beter wordt gedefinieerd. Alleen kun je niet elke dataset openstellen om zo betere kwaliteit te verkrijgen i.v.m. gevoeligheid van de dataset (zoals waterleidingnet). Open data heeft als voordeel dat de kwaliteit omhoog gaat (terugmeldplicht), maar als nadeel dat er door misbruik een groter risico is. De reactie van mensen op opendata is meestal: fijn dat het gratis is, jammer dat er niet meer is. Het openstellen van data helpt wel, maar zou beter moeten begeleid worden. Kwaliteitsinformatie kan helpen bij het maken van een keuze voor gebruik (kwaliteit als keuzecriteria).

Vraag van Gerard Hazeu: wat is datakwaliteit? Wat is daarbij de rol van metadata?

Damir geeft aan dat het ministerie van BZK met bredere kwaliteitsmetingen correct gebruik wil stimuleren. Data wordt soms incorrect gebruikt. De overheid denkt na over het instellen van een organisatie die hier op gaat letten.

Het is belangrijk om goed te communiceren over kwaliteit, zeker als data uit verschillende sectoren wordt gecombineerd.

Vraag Wies: Jasper jij gaf aan dat over 20 jaar de bestuurders veel digitaal ingesteld zullen zijn en dat dat een aantal problemen wel zal oplossen? Wat zullen de eigenschappen dan zijn van de bestuurders?

Jasper geeft aan dat beleid nu steeds meer op data gebaseerd wordt, meer data gedreven is. Dit proces zal in de tijd voor aanpassingen zorgen, zodat data steeds meer gemeengoed worden als nieuwe generaties zich aandienen. Data-gedreven betekent nog wel dat je kennis moet hebben van die data. Bijvoorbeeld, als je zegt dat de BGT een tapijt met weeffouten is. Als we over 20 jaar terugkijken hoe onze werkelijkheid er vroeger uitzag, moet je ook de kennis hebben dat de werkelijkheid er anders uit zag en waarom dat inzicht dan anders is. Alleen datagedreven werken geeft daarmee vertekende uitkomsten.

Vraag Maarten: Kan AI (artificial Intelligence) een rol spelen in kwaliteitscontrole van data?

Damir: Het ministerie van BZK werkt risico gericht. AI kan helpen bij een eerste scan, zodat daarna mbv steekproeven specifiek gecheckt kan worden. Henk Janssen: Soms is andersom denken ook handig met AI. Bij NEO kijken ze vaak mbv van AI wat nog hetzelfde is, dat is nl makkelijker dan te kijken waar iets is veranderd. AI moet vooral gebruikt worden als 'indicator' van de kwaliteit, maar er kan niet blindelings op vertrouwd worden. AI kent ook zijn beperkingen.

Vragen Martin Knotters: Kun je beslissen op basis van data als je de kwaliteit van die data niet kent en zonder kennis van statistiek, m.a.w. kennis van beslissen onder onzekerheid? En als je kunt kiezen, kies je dan voor verbetering van kwaliteit of voor kwantificeren van kwaliteit? Het kan namelijk dat het efficiënter is om bij beslissingen rekening te houden met onzekerheid, dan om te investeren in verdere onzekerheidsreductie: EVIU (Expected Value of Including Uncertainty) kan groter zijn dan EVPI (Expected Value of Perfect Information).

Je kunt geen beslissingen nemen als je datakwaliteit niet goed kent. Statistische informatie over nauwkeurigheid/onzekerheid is echter niet altijd simpel te doorgronden zonder voldoende kennis. Als onzekerheden worden berekend geven die wel een bandbreedte aan waarmee je kunt werken. Het valt of staat met een juist verwachtingsmanagement. Deze is afhankelijk van de kennis van de gebruiker.

Martin Peersmann geeft aan dat gebruikers van de bodemkaart en het WDM in de BRO meer kennis zouden moeten hebben van de toepassingsmogelijkheden die de stochastische realisaties bieden. Nu gebruiken zij veelal alleen de beste schatting (P50), terwijl er veel meer informatie is die je ook kunt gebruiken.

Derde ronde: wrap up

Aan het eind van de kennisarena vroeg Douwe alle deelnemers wat zij uit deze kennisarena gehaald hadden:

- **Henk Janssen** gaf aan dat hij intensiever met klanten wil praten over het acceptatieniveau: wanneer is de geleverde data voor de klant goed?
- **Andries Knotters** ziet een mogelijkheid in kostenbesparing door minder in te winnen, maar waarbij juist wel datgene wordt ingewonnen waar vraag naar is.
- **Richard Witmer** ziet een uitdaging in hoe om te gaan met de kwaliteit van BRT-Next. Dit wordt een basisregistratie die voor een groot deel gegenereerd wordt op een combinatie van andere basisregistraties. Kun je de kwaliteit van een product bepalen op basis van de kwaliteitsgegevens van de gebruikte bronnen? Dit is een veranderingsproces die hij ook graag met EKRK verder zou willen uitwerken.
- **Gerard Nienhuis** geeft aan verwachtingsmanagement belangrijk te vinden. Daarbij is communicatie erg belangrijk. Dit kan o.a. middels een dashboard waarop op geautomatiseerde wijze kwaliteitsinformatie zichtbaar is. Gerard ziet graag een methode om op een generieke manier de kwaliteit van data te bepalen en vast te leggen. Hij denkt daar graag met anderen over na.
- **Jasper Snippe** wil meer (kwaliteits)informatie aanreiken over data, door bijvoorbeeld een demo te geven en klantgebruik (intern) van BRO-gebruik meer voor het voetlicht brengen in de regio. Met Marien wil hij graag in gesprek over stageplekken voor studenten.
- **Damir Brnobic** geeft aan dat wat hij gehoord heeft goed aansluit bij de praktijk van BZK. Het is nu nog veel 'trial and error', dus des te beter gefundeerd des te beter. Graag zou hij samen met EKRK willen kijken hoe de praktijk aan de theorie getoetst kan worden, om te zien of ze op de goede weg zitten.
- **Marien de Bakker** geeft aan dat hij de verbinding tussen dataprovider en data gebruiker belangrijk vindt door meer regie en verbinding. Hij ziet nog niet dat gebruikers 'bewust' datagedreven gaan werken. Hij ziet mogelijkheden voor een cursus op dit gebied om studenten en werkveld te kunnen koppelen.

(*Martin Peersmann ontbreekt, hij moest eerder weg.)

Vanuit EKRK

- Voor het beter/meer communiceren over kwaliteit en kwaliteit een keuzecriteria laten zijn zou een sterrensysteem voor aantal generieke gebruiken een oplossing kunnen zijn.
- Het combineren van datasets is erg actueel en interessant om bijvoorbeeld in kader van BRT Next op te pakken.
- Voor het 20 jaar vooruit kijken is het interessant om ook 20 terug te kijken en te zien hoe de ontwikkeling de laatste 20 jaar is gegaan.
- Een generiek metadatasysteem.
- Datakwaliteit en gebruik: metadata geeft onvoldoende aan waarvoor data te gebruiken is/beperkingen van de data. En bij uitvraag van data worden er dan eisen aan datakwaliteit gesteld en ligt er daarbij dan een relatie naar aansprakelijkheid?

Samenvattend heeft de discussie een hoop waardevolle punten opgeleverd waarmee het EKRK verder kan om inderdaad die volgende stap te zetten. Dat was het uitgangspunt van de kennisarena. We willen graag de deelnemers over blijven informeren en betrekken. Wordt vervolgd!

Bijlage: antwoorden op vragen die vooraf per e-mail zijn gesteld

Gerard Nienhuis – Provincie Overijssel

OP WELKE WIJZE IS ER IN JULIE PROGRAMMA, ORGANISATIE OF DIENST VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?

Verantwoordelijkheid is verdeeld over 6 partijen, met ieder eigen rol en verantwoordelijkheid:

1) **inhoudelijk** verantwoordelijk contactpersoon. Dit is de eindverantwoordelijke persoon als het gaat over de inhoud van de datasets.

De naam en contactgegevens wordt voor alle beheerde datasets in de metadata opgenomen, conform iso19115 of dublin core. De naam van deze persoon is in het interne dataportaal altijd terug te vinden.

2) **Technisch** contactpersoon. Dit is de persoon die zorgt voor de technisch correct leveren van de dataset aan de databeheerders. Net als bij de inhoudelijk contactpersoon staat deze naam altijd in de dataset.

3) **Beheerders**: dit zijn data- en metadata-beheerders: dit is een groep mensen die verantwoordelijk is voor beheer en leveren van datasets, inclusief metadata.

Ontbreekt correcte metadata, dan worden datasets niet in beheer genomen. Kwaliteit van metadata wordt gecontroleerd (structureel en geautomatiseerd, inhoudelijk en technisch): resultaten worden in Tableau dashboards weergegeven.

Kwaliteit van data wordt gecontroleerd, meestal handmatig bij intake en op ad hoc basis tijdens beheerproces. Ook op regelmatige basis (jaarlijks, bij "voorjaarsschoonmaak". Soort kwaliteitscontrole van data is afhankelijk van soort dataset hoe grondig en volledig dit wordt uitgevoerd.

Er wordt al jaren gezocht naar methoden voor verder objectiveren en automatiseren van kwaliteitscontroles van datasets.

Dit blijkt lastiger dan gedacht. In Nederland zijn hier heel weinig ervaringen mee, en ook in buitenlandse literatuur is weinig te vinden over dit onderwerp.

Let op: er is wel literatuur over controle van standaard/basis datasets, maar weinig tot niets over steeds wisselende soorten datasets.

ps: beheerders zijn ook verantwoordelijk voor beschrijven, nakomen en laten doorgroeien van beheerprocessen, waarbij afspraken vastgelegd zijn in DAP (dossier met afspraken en procedures). Afstemming van verandering alleen na OK van CDO.

4) Chief Data Officer (CDO, meestal lijnmanager van medewerkers die in team data- en metadatabeheerders zitten) en betrokken unit-directeur.

Zijn eindverantwoordelijk formeel en ook in praktijk eindpunten van escalatie, als er onoplosbaar probleem in procedure van data of metadatabeheer niet zelf door inhoudelijk of technisch contactpersoon of door data/metadatabeheerders kan worden opgelost.

CDO stemt werk af met Chief Informatie Officer (moet nog verder vorm krijgen, CIO rol is net nieuw bij provincie Overijssel, voorloper CI office bestond al eerder)

5) functioneel en technisch beheerders van systemen die betrekking hebben op data: FB en TB voor Geoportaal, Tableau, Arcgis enz.

Dit zijn medewerkers van de provincie, maar soms is een TB ook lid van een externe organisatie (share service centrum, ONS).

6) Datagebruikers (data-analisten, kaartenmakers enz): dit zijn de mensen die nauw samenwerken met de inhoudelijk contactpersoon als het gaat over datakwaliteit. Problemen met kwaliteit (data of metadata), die worden gevonden, verbeter suggesties en andere issues, die worden besproken met de inhoudelijk contactpersoon en zo nodig met beheerders.

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIG HOUDEN?

zie hierboven intern. extern: erg weinig gevonden, hoop snel meer te vinden voor eigen toepassing kwaliteitsmonitoring en verbetering

- Kadaster richt zich op "eigen" bestanden (basisregistraties), CBS idem
- banken en verzekeraars hebben eigen systemen (veel te complex in te groot voor volledige toepassing bij provincie, denk ik)
- IBM: wel theoretisch kader, alleen in theorie onderzocht: vooral gericht op software
- Tmap: wel in aanleg wat over kwaliteit en testen, maar heb hier geen praktische voorbeelden van gezien
- nb: aantal softwareleveranciers bieden middelen aan, maar nog geen ervaring mee opgedaan. is nog groeiend: cursussen DAMA-DMBOK worden nu meer gevolgd dan jaar geleden, dus gecertificeerde partijen zullen ook ontstaan. Vraag is hoeveel ervaring de betrokkenen hier dan mee gaan hebben
- nb: hogeschool Utrecht heeft opleiding DAMA, nog geen ervaring mee (niet echt op kwaliteit gericht, vrees ik, meer algemeen)

HEBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN?

Databeheer doe je niet zonder metadatabeheer. Wij hebben bij de provincie Overijssel zowel databeheer als metadatabeheer gekoppeld, zodat wij "in control" zijn en blijven, wat beheer betreft. Bij veel veranderingen en veel data-sets red je dat niet meer met alleen een simpele excelsheet.

Je hebt daar wat meer geavanceerde software voor nodig, zoals dashboards (bv Tableau, zie verder).

Wat je verder nodig hebt is een ontwikkelstrategie: DAMA-DMBOK is aanrader.

Strategie moet je paar jaar echt volhouden, begint niet van 1 op andere dag.

Als je moet starten met datagovernance en datamanagement, denk dan niet dat je alles tegelijk perfect moet hebben. Begin klein, maak jezelf als databeheerder zichtbaar, richt je direct ook op metadatabeheer en hou het praktisch. Begin klein, maar denk groot (uiteindelijk heb je hele spectrum DAMA wel nodig).

Overleg regelmatig met partijen die met vergelijkbare zaken bezig zijn

Dialog, demo en toelichting over werkwijze provincie Overijssel is altijd te regelen: mail naar g.nienhuis@overijssel.nl)

Maar kijk ook ter inspiratie bij andere organisaties zoals: andere provincies, PBL, CBS enz.

Literatuur:

zeer nuttig, maar erg veel tegelijk: DAMA-DMBOK (begin praktisch is beste advies, groei langzaam door)

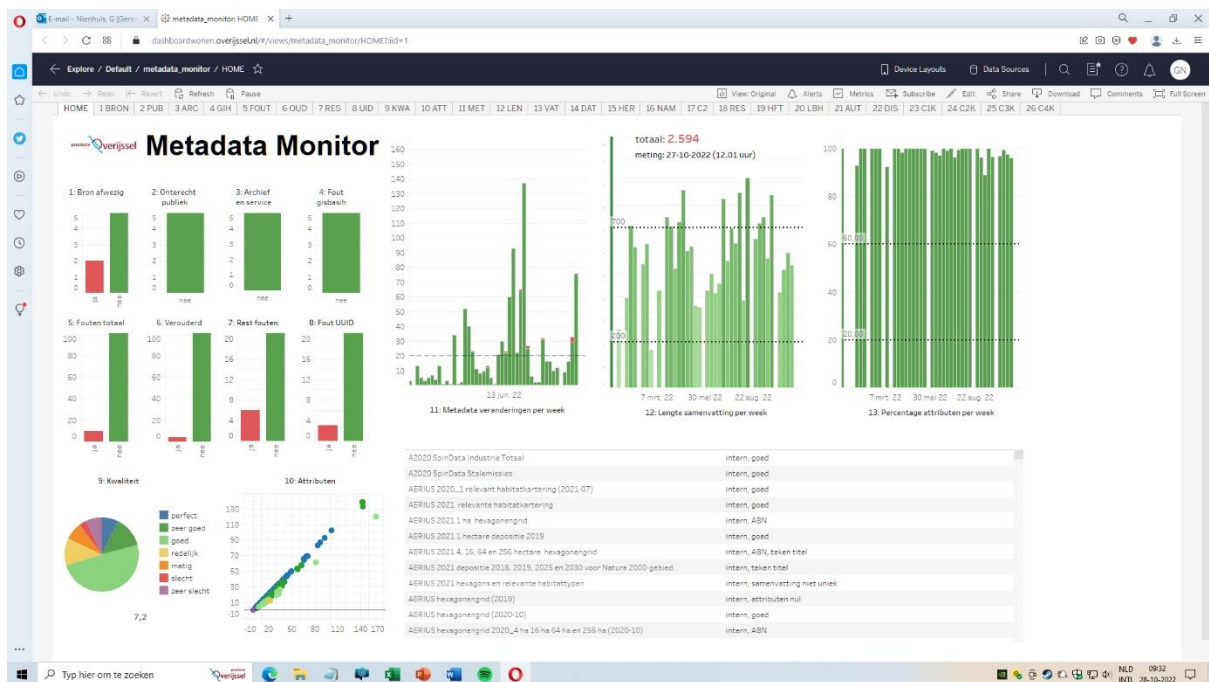
zeer inspirerend: Non Invasive Data Governance (Seiner)

leuk als achtergrondinfo, maar deel staat al in DAMA-DMBOK: Data stewardship (Plotkin)

veel te theoretisch, maar ook nuttig: Practitioner's Guide to Data Quality Improvement (Loshin)

ZO JA, Zouden jullie daar iets van toe kunnen sturen ter voorbereiding?

schermafdruk Tableau dashboard :



Dit is natuurlijk een momentopname van de kwaliteit van de data en metadata. Deze versie is van 27-10-2022, zoals je kan zien in het plaatje. Dagelijks wordt gewerkt, zodat er nieuwe data en metadata bij komt, bestaande wordt verbeterd en oude wordt gearchiveerd.

HEBben jullie een voorbeeld van een situatie waarin kwaliteit van data één van de oorzaken was van falen?

Nee, niet in zin van extreme fouten, wel veel bijna fouten, die op laatste moment voorkomen zijn.

nb: er is geen systeem (nog niet) voor melden, registreren en voorkomen van fouten: wordt ingericht, nog niet klaar.. zou erg goed zijn omdat je dan ook meer pro-actief fouten kan proberen te voorkomen: wens voor toekomst.

ps: er is wel ervaring opgedaan met systeem voor melden van issues rondom datakwaliteit: github, later topdesk. werkt goed, maar gestopt door capaciteitsgebrek van beheerders (wordt als dit is opgelost, ook weer opgepakt)

meest gemist op dit moment: volledige organisatie regie en DAP (data acquisitie planning).

Datagovernance nu ingericht in meest data-intensieve eenheden, maar moet eigenlijk organisatiebreed.

Stappen hiervoor zijn wel gezet, maar niet afgerond.

**OP WELKE WIJZE IS ER IN JULLIE PROGRAMMA, ORGANISATIE OF DIENST
VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?**

Binnen onze afdeling is toezicht en handhaving op de geobasisregistraties BAG, BGT en BRO georganiseerd. Binnen ons programma DiS Geo (doorontwikkeling in samenhang van de geobasisregistraties) zijn we het kwaliteitsregime aan het doorontwikkelen. De ambitie is om nog beter op de gerealiseerde gegevenskwaliteit te sturen door deze beter meetbaar te maken, gericht het gesprek erover met gebruikers te voeren en nog relevantere accenten te leggen met onze kwaliteitszorg (ondersteuning van bronhouders) en toezicht.

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIG HOUDEN?

Ben vooral bekend met instellingen binnen de context van het stelsel van basisregistraties en dan meer specifiek de collega-toezichthouders (RvIG, Waarderingskamer, Logius, etc.).

HEBBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN? ZO JA, Zouden Jullie daar iets van toe kunnen sturen ter voorbereiding?

We hebben een jaarplan kwaliteit voor BAG en BGT (BAG als voorbeeld bijgevoegd als bijlage) waarin we exact aangeven op welke manier wettelijke kwaliteitsnormen meetbaar zijn gemaakt in de vorm van indicatoren, welke voor komend jaar i.s.m. gebruikers geprioriteerd zijn (op maatschappelijk relevante data wordt strenger toegezien) en op welke manier bronhouders worden afgerekend op de bijhouding van de registraties. Het is belangrijk aan te durven geven dat niet alles van even hoge kwaliteit hoeft te zijn en duidelijk te zijn waarop met voorrang wordt gehandhaafd.

HEBBEN JULLIE EEN VOORBEELD VAN EEN SITUATIE WAARIN KWALITEIT VAN DATA ÉÉN VAN DE OORZAKEN WAS VAN FALEN?

Hulpdienst die niet (op tijd) bij een calamiteit kon komen vanwege een onjuiste koppeling tussen BAG- en BGT-openbare ruimten.

**OP WELKE WIJZE IS ER IN JULLIE PROGRAMMA, ORGANISATIE OF DIENST
VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?**

Rijkswaterstaat is een organisatie die sterk data-gedreven is. Er is nauwelijks een proces denkbaar waarbij er geen behoefte is aan input van data.

Bij het verzamelen van data is de kwaliteit een integraal aandachtspunt .i.r.t. vereiste kwaliteit vanuit toepassing en kosten efficiëntie

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIG HOUDEN?

Voor het waterbeheer-domein (waarin ik werkzaam ben) gaat het o.a. om: IHW (Aquo-standaard), Geonovum, Informatiehuis Water, enz.

HEBBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN? ZO JA, Zouden Jullie daar iets van toe kunnen sturen ter voorbereiding?

Weet ik niet direct een voorbeeld van, of misschien kunnen de RWSV's (RWS-standaardvoorschriften) als zodanig gezien worden. [Standaardvoorschriften metingen water | Rijkswaterstaat](#)

HEBBEN JULLIE EEN VOORBEELD VAN EEN SITUATIE WAARIN KWALITEIT VAN DATA ÉÉN VAN DE OORZAKEN WAS VAN FALEN?

Daar is in het waterbeheer-domein (waterkwaliteit/-kwantiteit) niet snel sprake van, want betreft geen missie-kritische systemen/data.

Dat is anders bij stuwen, sluizen en keringen, maar dat ligt buiten mijn directe werkterrein (wordt zo goed mogelijk afgedekt door redundantie van sensoren/systemen)

Marien de Bakker – HAS green academy

**OP WELKE WIJZE IS ER IN JULLIE PROGRAMMA, ORGANISATIE OF DIENST
VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?**

In ons onderwijs wordt aandacht gegeven aan de kwaliteit van data, als een noodzakelijke check in relatie met de vraag (fitness for use). Daarnaast hebben we een cursus Datagedreven werken, waar in de data kwaliteit ook aan de orde komt. <https://www.has.nl/meer-has/nieuws/cursus-datagedreven-werken-van-start/>

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIGHOUDEN?

Niet specifiek, ga uit van providers, maar ook users

HEBBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN? ZO JA, Zouden jullie daar iets van toe kunnen sturen ter voorbereiding?

Niet heel specifiek. We proberen vooral in de vraagstelling al aan de orde te laten komen wat de noodzakelijke kwaliteit (schaal, ouderdom, inhoud, volledigheid, metadata, etc) van het resultaat zou moeten zijn. Vaak is dat lastig, omdat de reactie van de vraaghebber varieert tussen, “het best mogelijke” of “dat weten jullie toch het beste” of “ik heb geen enkel idee”.

HEBBEN JULLIE EEN VOORBEELD VAN EEN SITUATIE WAARIN KWALITEIT VAN DATA ÉÉN VAN DE OORZAKEN WAS VAN FALEN?

Recent een afstudeeropdracht voor het Ministerie van W&I over machine learning rond gastanks. Het lukte niet om de juiste volledigheid te krijgen vanuit onze kant, maar opdrachtgever was tevreden om te zien dat de aanpak van ML versterkt werd door allerlei andere methoden. Gaan daar in februari mee door.

**OP WELKE WIJZE IS ER IN JULLIE PROGRAMMA, ORGANISATIE OF DIENST
VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?**

Voor een aantal basisregistratie is kwaliteitsborging / toezicht voorzien. Er zijn dan middelen voor controles, terugmeldingen, monitoring en kwaliteitsdashboard. Tevens kennen we voor een aantal componenten de ENSIA-zelfevaluatie. Formele verantwoordelijkheid (wet) ligt in de meeste gevallen bij het ministerie van BZK in enkele gevallen (BRK en BRT) bij het Kadaster.

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIG HOUDEN?

Kadaster, WENR, Geodelta.

HEBBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN? ZO JA, Zouden Jullie daar iets van toe kunnen sturen ter voorbereiding?

KTH-Systematiek, Kwaliteitszorg BRK – BRT, Kwaliteitszorg luchtfoto's.

HEBBEN JULLIE EEN VOORBEELD VAN EEN SITUATIE WAARIN KWALITEIT VAN DATA ÉÉN VAN DE OORZAKEN WAS VAN FALEN?

De Kadastrale Kaart is nu niet goed te gebruiken in combinatie met andere bronnen (luchtfoto, BGT) omdat de ligging van grenzen (i.v.m. historie) op de kaart kan afwijken van de feitelijke ligging in het terrein. Dit wordt nu opgepakt in het kader van Kadastrale Kaart Next. Geeft nog steeds veel aanleiding voor disputen en misverstanden. O.a. bij bouwen over de grens.

Richard Witmer, Kadaster.

**OP WELKE WIJZE IS ER IN JULLIE PROGRAMMA, ORGANISATIE OF DIENST
VERANTWOORDELIJKHEID/TAAK/TIJD/GELD VOOR KWALITEIT VAN DATA VOORZIEN?**

Voor de basisregistraties BRK, BRT, BAG en BGT is verantwoordelijkheid/taak/tijd/geld geregeld dmv. kwaliteitsmanagement als onderdeel van de taken van het Kadaster.

Via PDOK en NGR worden de metadata van veel datasets vastgelegd. Terugmeldsysteem Verbeter de Kaart, het Klant Contact Center en BKK (Bezwaren en Klachten) zorgen ook voor verbetering van de kwaliteit van data.

WELKE INSTELLINGEN KENNEN JULLIE DIE ZICH MET KWALITEIT VAN DATA BEZIG HOUDEN?

WEnR, andere universiteiten, Geonovum, Kadaster, ministerie van BZK, SVB-BGT / VNG, andere overheden, bedrijfsleven, alle gebruikers van de data.

HEBBEN JULLIE BEST PRACTICES OF BEST-LEARNINGS DIE JULLIE KUNNEN DELEN? ZO JA, Zouden Jullie daar iets van toe kunnen sturen ter voorbereiding?

Terugmeldsysteem Verbeter de Kaart is zeer laagdrempelig en wordt al jarenlang veel gebruikt om de data van BRT, BAG en BGT te verbeteren. Het systeem wordt uitgebreid met andere registraties.

HEBBEN JULLIE EEN VOORBEELD VAN EEN SITUATIE WAARIN KWALITEIT VAN DATA ÉÉN VAN DE OORZAKEN WAS VAN FALEN?

Nee, gelukkig niet meteen voorhanden.