



Ministerie van Infrastructuur
en Waterstaat

Wegwijzer

Datagedreven werken bij DGWB



Inhoud

Inleiding 5

Deel 1 Ga datagedreven werken! 6

Deel 2 Het dataproces 8

1. Beleidsvraag 8
2. Inspiratie 10
3. Inventarisatie data 12
4. Data analyse 12
5. Data resultaten 13
6. Data toepassen 13
7. Afsluiting 13

Deel 3 Factsheets van voorbeeldprojecten 14

Deel 4 Achtergrondinformatie 20

1. Handig om te weten 20
2. Theorie van Gartner 23

Auteurs:
Eric Harthoorn
Martine Ouwersloot

Inleiding

De wereld verandert in rap tempo. Nederland digitaliseert en dat biedt grote kansen om dingen slimmer te doen. We willen als overheid, in de meest brede zin van het woord, die kansen benutten. Zo wordt er veel gesproken over bijvoorbeeld data, kunstmatige intelligentie, blockchain, satellietdata en drones en de kansen die deze ontwikkelingen bieden. Overal zien we dat overheden, bedrijven en instellingen investeren in nieuwe technische mogelijkheden omdat ze kansen zien voor het verbeteren van hun taken, diensten of producten.

Deze wegwijzer is ontwikkeld voor beleidsmedewerkers van DGWB en bedoeld als hulp bij het gebruik van data in hun werkproces. Deze wegwijzer kan echter ook gebruikt worden door medewerkers van andere DG's. Dit document zal uitleg geven over (werken met) data, voorbeelden geven, en informatie verschaffen over hoe men zelf aan de slag kan met een dataproject en waar ondersteuning te vinden is.

Een afgeleid doel van de wegwijzer is de medewerkers van DGWB bewust te maken van de mogelijkheden die data kan bieden bij het ontwikkelen/bijstellen van beleid. Data moet gezien worden als een aanvulling op de al bestaande mogelijkheden/instrumenten waar een beleidsmedewerker in zijn werk al gebruik van maakt. Deze wegwijzer helpt beleidsmedewerkers van DGWB van onbewust onbekwaam naar bewust bekwaam te brengen op het gebied van datagedreven werken.

De toegevoegde waarde van het gebruik van data is:

- **Beter fundament:** Betere fundering en onderbouwing van beleid mogelijk maken, omdat data inzicht kan geven in het effect van eerdere interventies en voorspellingen kan doen over het effect van beleidsbeslissingen.
- **Iedereen betrokken:** Interdisciplinaire samenwerking van beleidsmensen, medewerkers beleidsuitvoering, en data-deskundigen;
- **Grotere voorspelbaarheid:** Het beter kunnen voorspellen van toekomstige situaties, waardoor beleid eerder kan worden bijgesteld op gewenste beleidsopgaven en niet alleen maar achteraf analyses te kunnen maken;
- **Meer detail:** Het blootleggen van nieuwe, en/of betere, en/of gedetailleerdere informatie die anders onder de radar was gebleven, en inzicht in specifieke behoeften en op basis daarvan maatwerkoplossingen te kunnen bieden.
- **Leggen van verbanden:** Het mogelijk maken om verbanden te leggen (door bijvoorbeeld databestanden aan elkaar verbinden) die anders niet zouden zijn gemaakt.

Leeswijzer

Deze wegwijzer hoeft je niet keurig van voor naar achter door te lezen. De wegwijzer is opgedeeld in vier delen. Het eerste gedeelte geeft uitleg over het wat en het waarom van datagedreven werken. Het tweede gedeelte beschrijft het proces van hoe je met data zou kunnen werken, hoe je daar mee begint en welke stappen daar in het algemeen bij horen. Het derde gedeelte van de wegwijzer laat een vijftal voorbeelden zien van dataprojecten en -pilots die worden of zijn uitgevoerd, waar je inspiratie uit op kan halen. Het laatste gedeelte biedt wat achtergrondinformatie bij veelvoorkomende begrippen rondom data.

Het zijn dus verschillende onderdelen waarin je de informatie kan halen die je op een bepaald moment nodig hebt. Shop dus gerust rond in de informatie die hier te vinden is.

Deel 1

Ga datagedreven werken!

Wat is datagedreven werken?

Datagedreven werken betekent het systematisch verzamelen, beheren, analyseren, interpreteren en toepassen van data. Als beleidsmedewerker bij DGWB zou je je moeten afvragen over welke data er beschikbaar is en hoe je deze kan toepassen om een (maatschappelijk) probleem op te lossen. Werken met data is echter nooit een doel op zich, het is een middel om een doel te bereiken.

Data kan je vermoedens bevestigen en je helpen je argumentatie te onderbouwen met de feitelijke informatie die data geeft. Of de data levert nieuwe inzichten op, informatie die je niet had verwacht. Data kan een schat aan informatie opleveren, als je maar weet waar en hoe je moet zoeken. Wij nodigen je uit om aan de hand van deze wegwijzer mee te gaan op deze zoektocht.

Waarom datagedreven werken?

Wij kunnen ons voorstellen dat je je afvraagt waarom je met data zou moeten gaan werken, als de manier waarop je voorheen je werk deed ook goed was. Wij twijfelen ook niet aan je deskundigheid, of aan de vakkennis en intuïtie die je hebt opgebouwd om vraagstukken te beantwoorden. Wij willen slechts laten zien dat data een hulpmiddel kan zijn in je werk, die juist de opgebouwde deskundigheid, vakkennis en intuïtie kan onderbouwen. Daarnaast kan data ook nieuwe informatie bieden. Er is namelijk enorm veel data, die extra kansen kan bieden. Het biedt een nieuwe, innovatieve manier van werken die je kan helpen je werk beter te doen.

Hoe datagedreven werken?

Deze wegwijzer is geschreven om handvatten te bieden bij het datagedreven werken proces. Elke vraag, elke case, elke dataset is weer anders, dus deze wegwijzer zal nooit het antwoord geven op al je vragen. Deze wegwijzer kan je een idee geven waar je aan kan denken, wat mogelijke stappen zijn en waar je terecht kan met je vragen. Onze eerste tip is dan ook om met de informatieadviseur van het DG te gaan praten, hij kan je verder helpen met je vragen door in gesprek met je te gaan en je eventueel door te verwijzen naar de juiste mensen. Maar de informatieadviseur is er niet alleen voor de eerste stap, ze zullen je bijstaan gedurende het hele proces (mocht je dat willen).

Privacy & ethiek

Als ministerie moeten we verantwoord omgaan met data en transparant zijn over het gebruik daarvan. Zo speelt bij het werken met datasets altijd de kans dat er persoonsgegevens in staan. Daar moet je zorgvuldig mee omgaan.

Het gebruik en persoonsgegevens en het toepassen van privacy aspecten is in de wet AVG (Algemene Verordening Persoonsgegevens) vastgelegd. De belangrijkste principes van de AVG zijn: rechtmatigheid, doelbinding, dataminimalisatie, en transparantie. Deze betekenen dat het verzamelen van persoonsgegevens een rechtmatige grondslag moeten hebben, voor een specifiek doel, dat er niet meer data mag worden verzameld dan nodig en dat men transparant is over hoe, wanneer, welke, waarom en door wie data verzameld wordt. Het is dus belangrijk vooraf na te gaan of er persoonsgegevens voorkomen in de datasets, en of deze nodig zijn om te werken met de data.

Mocht je vragen of twijfels hebben, dan kun je deze het beste bespreken met de privacy-adviseur. Hun principe is 'beter een keer te veel dan te weinig gevraagd', dus schroom vooral niet.

Daarnaast zijn ethische waarden van groot belang bij werken met data. Hoe wordt bijvoorbeeld bepaald waar een zelfrijdende auto voor kiest als er plotseling een kind oversteekt en remmen niet meer helpt? Uitwijken en een wandelaar op het trottoir aanrijden, of toch zo hard mogelijk remmen en het kind aanrijden? Oftewel: door machine learning maken computerprogramma's zelf algoritmen, maar welke regels geef je het programma mee en wie controleert het algoritme? Omdat dit allemaal relatief nieuwe vraagstukken zijn, hebben we daar geen eenduidige antwoorden voor. Het is dus goed om met de betrokken van het dataproject eventuele ethische kwesties goed te bediscussiëren.

Om zo'n discussie op gang te brengen, kun je de [Ethische Data Assistent](#) van de Dataschool van Universiteit Utrecht volgen. Deze stelt je vragen over de kwaliteit van data, communicatie over de data, doel van het project, de algoritmen die worden gebruikt etc. Aan de hand hier van kun je zelf bepalen of je project ook ethisch goed in elkaar zit, of dat er bepaalde aspecten nog aangepast moeten worden.

Wat gaan we anders doen?

Enkele voorbeelden hoe data leidt tot anders werken:

- Datagedreven werken kan leiden tot *fact based policy making*, oftewel het maken van beleid op basis van feiten. Dit heeft invloed op de gehele beleidscyclus, van beleidsvoorbereiding tot beleidsevaluatie. In elke stap kan structureel data worden gebruikt als feitelijke informatie om beleid op aan te passen, evalueren, monitoren, etc.
- Daarnaast verandert de maatschappij, en is het belangrijk dat wij ons ook aanpassen. Door de opkomst van bijvoorbeeld citizen science* is het belang van op de hoogte te zijn van wat men in de omgeving meet, om zo niet overvallen te worden als hier vragen door komen.
- Verder kan het tot je beschikking hebben van actuele en juiste data helpen bij het beantwoorden van Kamervragen of WOB-verzoeken.

Wat als we het niet doen?

Uiteindelijk komt er het op neer dat werken met data geen disruptieve manier van werken is, het is slechts een aanvullend middel voor het maken van beleid. Het is echter wel belangrijk om data mee te gaan nemen in je beleidsproces. Data wordt veel makkelijker gemaakt en gedeeld. Niet alleen jij als beleidsmaker, maar ook burgers en belangenorganisaties hebben toegang tot data (denk aan open data, of data die ze zelf verzamelen) en kunnen daaruit eigen conclusies trekken. Voorheen was het voor personen buiten het ministerie lastiger om aan dit soort informatie te komen, maar het is nu veel meer in opkomst. Als we niets doen, we worden ingehaald door burgers en belangenorganisaties.

* Citizen science is het proces waarbij burgers op vrijwillige basis data verzamelen voor een bepaald doel of project. Een goed voorbeeld van citizen science is de Tuinvogeltelling van Vogelbescherming Nederland. De Vogelbescherming analyseert de data, maar die wordt door burgers zelf verzameld en doorgegeven.

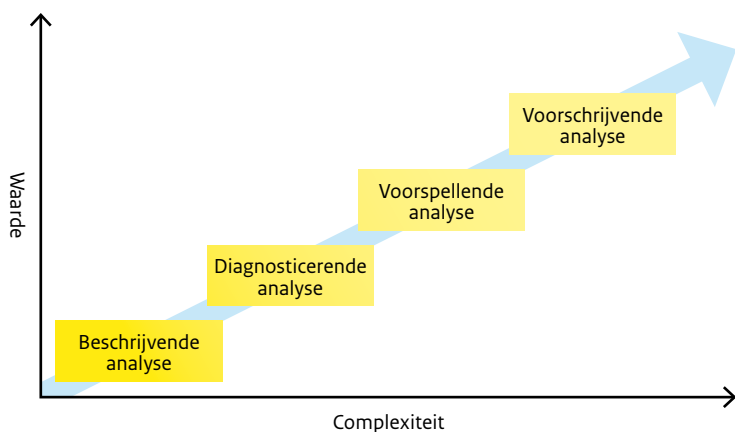
Deel 2

Het dataproces

1. Beleidsvraag

Wat is mijn beleidsopgave? Welke vragen heb ik daarvoor te beantwoorden?

Of het nou om beleidsvoorbereiding, -vorming, -uitvoering, of -evaluatie gaat, data kan bij allen een rol spelen in het beantwoorden of onderzoeken van de beleidsvraag. Echter zal het niet altijd even duidelijk zijn of, hoe en op wat voor manier data een rol kan spelen bij de beleidsopgave. Om dat helder te krijgen, is het goed om eerst na te gaan wat de beleidsopgave is en welke informatie daar bij hoort.



Bron: Gartner

Uit ervaring van de gemeente Zaanstad (koploper op het toepassen van data bij gemeentelijke vraagstukken) blijkt dat de eerste stap het belangrijkste is. Terugkijkende of beschrijvende analyses zijn van groot belang. Als deze niet bestaan is het immers heel moeilijk om voorspellende analyses te doen. Voorspellende analyses moeten namelijk ergens op gebaseerd worden, en dat is meestal op resultaten uit het verleden. Onze tip is dan ook om eerst te beginnen met een beschrijvende analyse.

Datagedreven werken proces



Bij welke vragen kan ik data gebruiken?

Om je op weg te helpen kun je kijken of je vraag bij een van de volgende typen vragen hoort:

• Beschrijvend: Wat is er gebeurd?

Voorbeeld: waar komt bodemdaling voor, en hoe groot is de omvang van bodemdaling?

Dit is vaak de eerste vraag je beantwoord wilt hebben. Zonder de antwoorden bij dit type vragen is het lastiger om een diepere analyse te maken.

Als of wanneer je een goed beeld hebt van wat het fenomeen is dat je wilt onderzoeken, dan zou je naar de volgende stap kunnen gaan. Echter kan het ook zo zijn dat wanneer deze beschrijvende analyse voltooid is, je het antwoord hebt dat je zocht. De andere analyses zou je dan achterwege kunnen laten, of in het achterhoofd houden voor een later moment.

• Diagnosticerend: Waarom is het gebeurd?

Voorbeeld: wat zijn de oorzaken van bodemdaling?

Bij dit soort vragen komt al wat meer data analyse kijken. Eén kaart over waar de bodemdaling voorkomt zal nog niet zo veel zeggen over waarom het op bepaalde plekken meer voorkomt dan anderen. Daar heb je ook andere informatie bij nodig. In het geval van de bodemdalingssignaleringskaart werd er ook gekeken naar grondwater kaarten en andere relevante kaarten in de Klimaateffectatlas. Door deze kaarten te vergelijken werden er nieuwe verbanden gelegd.

- **Voorspellend: Wat en waar gaat er het gebeuren?**

Voorbeeld: kunnen we voorspellen waar bodemdaling gaat plaatsvinden en wat zijn de risico's?

Als je weet waarom en hoe in het verleden het verschijnsel heeft plaatsgevonden, kun je wellicht ook vooruit kijken naar wat er in de toekomst kan gebeuren. In het geval van de bodemdalingskaart zal er worden gekeken naar de grondwaterkaarten en daar mogelijke nieuwe hotspots uit gehaald. Door die nauwlettend in de gaten te houden kun je wellicht bodemdaling sneller signaleren op die plekken, en aan de hand van die informatie je voorspelling bijstellen.

- **Voorschrijvend: Hoe kunnen we het laten gebeuren/voorkomen?**

Voorbeeld: welke maatregelen (beleid) kunnen we maken om bodemdaling te voorkomen?

Wanneer je weet of vermoedt dat de kans op bodemdaling op een bepaalde plek aanzienlijk groot is, kan men maatregelen nemen dit te voorkomen. Het stukje data-analyse dat hier bij komt kijken is dat de plekken waar maatregelen/beleid worden toegepast in de gaten worden gehouden om te kunnen bepalen of de maatregelen/beleid effect heeft gehad. Als er op meerdere plekken beleid wordt toegepast, kan deze informatie worden gebundeld en daar weer meer informatie uit voortkomen.

Als je de vraag helder(er) hebt, kun je bepalen wie je zou kunnen helpen met het verzamelen van de data. Dit is natuurlijk afhankelijk van de data die je nodig hebt. Als je wel al weet wie de beoogde dataset heeft, kun je deze benaderen. Gedacht kan worden aan een (uitvoerings-)organisatie als RWS, ILT of KNMI, waar ze ook actief met data bezig zijn in hun datalabs.

Helpen bovenstaande vragen niet om je vraag te verduidelijken, dan raden we je aan om in gesprek te gaan met de informatieadviseur van je DG. Vaak helpt het al om samen te brainstormen over de datavraag. Daarnaast kan de volgende stap wellicht helpen.

2. Inspiratie

Datapilots binnen & buiten IenW

Voorbeelden van dataprojecten zijn zowel binnen als buiten het departement te vinden. Andere directies, DG's, departementen, bedrijven, belangenorganisatie, burgers; bijna iedereen is bezig met werken met data. Het is goed eens rond te kijken, en te bedenken of iets aanspreekt. Hier volgt een kort overzicht van de projecten die binnen IenW lopen (meer informatie over de projecten staat in de factsheets vanaf pagina 14). Verder is het goed om te beseffen dat het idee voor jouw dataproject wellicht ook al ergens anders verkend of uitgevoerd wordt. Dat kan mogelijkheden bieden. Wellicht kan er samengewerkt worden, of in ieder geval van elkaar geleerd. Het is namelijk ook mogelijk dat een partij het onderwerp op een andere manier benadert dan jij voor ogen hebt. In zo'n geval hoeft je je eigen werkwijze natuurlijk niet aan te passen, maar kunnen de projecten worden gezien als aanvulling op elkaar.

In gesprek gaan met collega's

Binnen IenW zijn er al datapilots geweest bij verschillende DG's. Hieronder volgt een korte samenvatting van de projecten. Een uitgebreidere toelichting op de projecten kun je vinden in deel 3: factsheets.

Kennismaken met datalabs

Ook door een kijkje te nemen bij een datalab kom je achter mogelijkheden van werken met data. Een datalab is een plek waar gewerkt wordt met data, waar dus de data wordt verzameld en geanalyseerd. Vaak is het ook een plek waar geëxperimenteerd kan worden met verschillende technieken en waar er ruimte is om nieuwe ideeën uit te werken. Een goed startpunt dus voor een dataproject.

1 Bodemdalingssignaleringskaart

Doel van dit dataproject is te onderzoeken of er nieuwe relevante informatie/indicatoren kunnen worden verwacht uit satelliet afgeleide informatie over bodemdalingsdeformatie, en daardoor beter inzicht te krijgen in de trends van bodemdaling. Bij dit project is er dus vooral gebruik gemaakt van satellietdata, en verschillende satellietdata aan elkaar gekoppeld om nieuwe verbanden te leggen.

2 Eems-Dollard

Door satellietdata te gebruiken komt er gebiedsdekkende informatie beschikbaar over de waterkwaliteit en primaire productie. Dit is in tegenstelling tot hoe het hiervoor ging, namelijk een beperkt aantal fysieke meetpunten dat informatie verschafte.

3 Hittestress

In deze pilot werd er gekeken naar hittestress in steden, en wat voor invloed dat heeft op de bouw van nieuwe woningen. Temperatuurbestanden en klimaatmodellen werden hiervoor geanalyseerd.

4 Opkomende stoffen

Op verschillende manieren werd er gekeken naar bestaande informatiebronnen, om te kijken of er nieuwe informatie uit de bestaande data gehaald kon worden met betrekking tot opkomende, en potentieel gevaarlijke stoffen. Gebruikte data zijn onder andere grote tekstbestanden, buitenlandse data, data uit meetstations van RWS. Er werden verschillende technieken toegepast om hier meer informatie uit te halen.

5 Luchtkwaliteit

In deze pilot werd gekeken naar een verband tussen het weer, de doorstroming van het wegverkeerd, en de luchtkwaliteit. Er werd gebruik gemaakt van realtime monitoring daarvoor relevante data, en gekeken of hieruit een voorspellend model afgeleid kon worden.

Hier volgt een korte beschrijving van de datalabs die onder de beleidsterreinen van IenW vallen:

ILT

Sinds 2017 heeft de ILT een innovatie- en datalab waar analisten en wetenschappers werken aan oplossingen voor analyse vraagstukken en maatschappelijke problemen op de diverse terreinen van de ILT. Bij de ILT kun je terecht voor o.a. statistiek, netwerkanalyse, speltheorie, textmining, webscraping of Geografische Informatie Systemen. DGWB zou bij de ILT terecht kunnen voor vragen rondom binnenvaart, bodemtoezicht, drinkwater, hoogwaterveiligheid, en milieu.

RWS

Rijkswaterstaat is, na het CBS, de grootste overheidsleverancier van Open Data. RWS heeft veel data in haar bezit, die ze voor haar eigen werk gebruikt en nodig heeft. Dit zijn vooral gegevens op het gebied van water, weg en spoor.

KNMI

Het datalab van KNMI richt zich op het moment op twee hoofdthema's. Het eerste is 'KNMI in de straat' met als doel het verkrijgen van hoge resolutie weer- en luchtkwaliteitswaarnemingen en prognoses. Het tweede thema is 'Weer en Verkeer' met als doel een hoge resolutie van waarnemingen en prognoses op de weg- en waterinfrastructuur. Het KNMI heeft naast het datalab ook een datacentrum, waar data opgevraagd kan worden over het weer, klimaat, seismologie, klimatologische gegevens, satelliet- en radardata en gegevens over aardbevingen. De meeste data is vrij beschikbaar en wordt als Open Data verstrekt.

3. Inventarisatie data

Zoeken naar data

Wanneer je vraag meer vorm heeft gekregen, kun je bekijken in hoeverre de informatie die je zoekt al beschikbaar is bij relevante organisaties (zoals bijvoorbeeld Rijkswaterstaat). Wellicht bestaan er al factsheets of zijn er in het verleden al onderzoeken gedaan naar het onderwerp. Het is goed om in beeld te hebben welke kennis er al is, en waar nog verdiepende analyses uitgevoerd moeten worden. Goed om te weten is dat het lastig kan zijn om data van partijen los te krijgen, omdat men niet graag heeft dat je zo maar met hun data aan de haal gaat. Laat je op zo'n moment niet gelijk afschrikken. Ga met de partij in gesprek en maak duidelijk waarom je data van een partij kan hebben.

Daarnaast is er een IenW dataregister in ontwikkeling. In deel 4 'Handig om te weten' vind je alvast een overzicht van beschikbare data specifiek voor het domein Water & Bodem. Mocht je een beoogde dataset hier niet tussen staan, dan kun je informeren via de informatieadviseur of er bij de uitvoerende organisatie al data verzameld wordt, maar wellicht nog niet gepresenteerd wordt. Daarnaast kun je via de open data website van de rijksoverheid een [open data verzoek doen](#).

Data vergaren en/of data laten genereren?

Als de beoogde data (nog) niet bestaat is er natuurlijk ook nog de optie om de opdracht te geven de data te laten genereren door een partij. Bekijk dan in overleg wat er mogelijk is. Natuurlijk zorgt het zelf creëren of verzamelen van data wel voor een mogelijke langere doorlooptijd van het project.

Bij het laten genereren van data is het van belang dat je nadenkt op welk detailniveau je data wilt gebruiken, en op welk detailniveau je betrouwbare uitspraken wilt kunnen doen. Pas op voor schijnnaauwkeurigheid: Meer gedetailleerde data is niet altijd meer betrouwbaar, de kans op meetfouten kan bijvoorbeeld groter zijn. Daarnaast is een groot detailniveau misschien niet eens nodig. Bijvoorbeeld, is het nodig om bodemdaling op de vierkante meter te weten, of volstaat een groter detailniveau ook?

4. Data analyse

Als de opdracht en vraag duidelijk zijn kan er begonnen worden met het verzamelen van de nodige data en de analyse ervan. Deze stappen worden meestal bij een datalab of een ander data-expertise centrum uitgevoerd. Bij IenW kan er gebruik worden gemaakt van de datalabs van het KNMI, RWS, of ILT. Echter kan dit ook door een universiteit of ingenieursbureau worden gedaan. Dit wordt samen bepaald in gesprek met de I-adviseur. Daarnaast is er een eenvoudige datalab omgeving voor de bestuurskern beschikbaar, hiermee kunnen snel eenvoudige analyses worden uitgevoerd. Soms zullen de analisten gelijk aan het werk kunnen omdat de data waarmee moet worden gewerkt al bestaat. In andere gevallen moet deze nog gegenereerd worden. Op wat voor manier dit gebeurd etc. zal natuurlijk van te voren worden besproken.

Data vergaren, harmoniseren, bewerken en analyseren

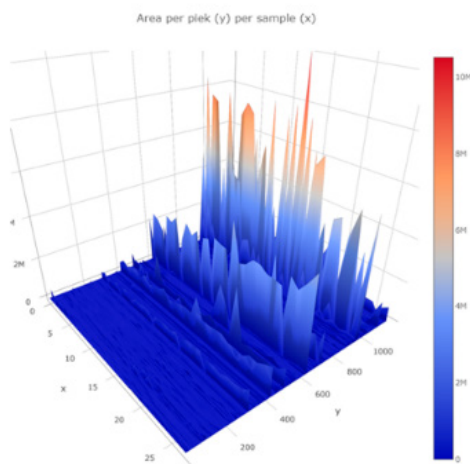
De analyse door het datalab of een andere derde partij is een stap waarbij soms geduld vereist zal zijn. Uit ervaring blijkt dat het analyseren van data nogal wat voeten in de aarde kan hebben. Zeker op het moment dat er met verschillende datasets gewerkt moet worden. Om deze datasets aan elkaar te kunnen koppelen, zullen ze eerst compatibel gemaakt moeten worden bijvoorbeeld. Als in de ene dataset bijvoorbeeld afstanden worden gemeten in meters, en in de andere in kilometers, moet dit eerst worden aangepast. Maar ook in een enkele dataset zal er moeten worden gecontroleerd op gebruikte invoer consistent is. Verder blijkt het verzamelen van datasets bij verschillende instanties en eigenaren soms lastig te zijn, omdat ze hun data niet of niet volledig vrij willen geven. Dit zijn allemaal zaken waarmee rekening gehouden moet worden. Gelukkig is al gebleken dat dit niet alleen maar belemmeringen hoeven te zijn. Het kan ook tot gevolg hebben dat er door met andere datasets te werken dan vooraf was voorgeschreven, ook andere uitkomsten zijn die men niet verwacht had maar wel tot nieuwe inzichten kunnen leiden. Ook bij deze stap is betrokkenheid belangrijk. Er zal geregeld feedback komen over het verloop van de analyse of verkrijgen van data, of er moeten vragen van de uitvoerders van de data-analyse verduidelijkt worden.

5. Data resultaten

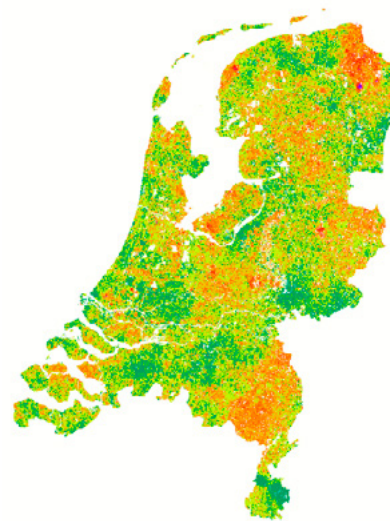
Data visualisaties: tabellen, grafieken, etc.

Het visualiseren van de uitkomsten van de data-analyse (zoals kaarten, grafieken, dashboards etc.) zijn erg belangrijk om de informatie uit de data te halen. Dit wordt gedaan door de organisatie die ook de data heeft geanalyseerd, eventueel in overleg met jou. Uit de visualisatie kun je dan de informatie halen die uit de data is voortgekomen.

Enkele voorbeelden staan op deze pagina.



1. Gaschromatografie bij de datapilot Opkomende Stoffen



2. Bodemdalingsignaleringskaart

6. Data toepassen

Beantwoording van beleidsvragen

Bij het implementeren van de resultaten kun je jezelf bijvoorbeeld volgende vragen stellen:

- Wat zijn de opvallende inzichten n.a.v. de verschillende soorten analyses?
- Wat betekenen deze inzichten voor de praktijk?
- Wat zijn de mogelijke oplossingsrichtingen voor het vraagstuk?

Aanpassen/bijstellen beleid

De implementatie van de uitkomsten lijkt een laatste stap in het dataproces, maar dat hoeft het niet te zijn. In een proces van continue of real-time monitoring zal deze stap constant opnieuw worden uitgevoerd, al dan niet vooraf gegaan in bijsturingen in het proces dat er aan vooraf gaat.

7. Afsluiting

Werken met data wordt ook wel een zoektocht genoemd. Het is vooraf niet altijd even helder welke data beschikbaar is, de vraag kan tijdens het dataproces veranderen en de resultaten zijn soms verrassend. Dit maakt het werken met data nu zo boeiend en uitdagend. We hopen dan ook dat deze wegwijzer je iets wijzer maakt over het werken met data, wat er allemaal mogelijk is en hoe dit aan te pakken.

Deze wegwijzer zal nooit helemaal af zijn. Nieuwe datatechnieken, nieuwe voorbeelden, nieuwe wet- en regelgeving, en nieuwe ambities kunnen allemaal van invloed zijn op het werken met data. Dat zorgt er voor dat data constant in ontwikkeling is en dat er steeds meer mogelijk zal zijn, zolang men de kansen ziet en deze durft te grijpen!

Deel 3

Factsheets van voorbeeldprojecten

Factsheet Dataproject/pilot

Bodemdalingsignaleringskaart (DGWB)

Naam: Bodemdalingsignaleringskaart

Omschrijving/doel: Onderzoeken of nieuwe relevante informatie/indicatoren kunnen worden verwacht uit satelliet afgeleide informatie over bodemdalingsdeformatie en daardoor beter inzicht krijgen in de trends van bodemdaling.

Organisatie onderdeel/opdrachtgever:

Hanneke Brouwer (DGWB/WOM/BOW)

Contactpersoon: Douwe Jonkers, Tommy Bolleboom (RWS)

Uitvoerder/datalab: RWS/Geomatics ventures Ltd

Gebruik van databronnen/datatechnieken etc.:

Technieken:

- Satellietdata

Databronnen:

- Hoogtekaarten
- Bodemkaarten
- Grondwaterkaarten
- Datasets van stedelijke bodemdaling uit projecten uit het verleden
- Klimateffectatlas



Resultaat: Uit de resultaten blijkt dat grondwater een belangrijke factor is in bodemdaling. Daarnaast kunnen met de data genomen maatregelen geëvalueerd worden.

Het project levert informatie over bodemdaling die medeoverheden nodig hebben om de verplichte stresstesten in het kader van de Delta aanpak klimaatadaptatie te kunnen uitvoeren.

In de pilot is gewerkt met INSAR (satelliet) technologie, die nauwkeurig bodemdaling hotspots kan bepalen. Validatie van data door andere data blijft echter belangrijk, om te kunnen bepalen of de orde van grootte van de bodemdaling correct is. Data die werd gebruikt voor validatie is afkomstig van bijvoorbeeld de waterschappen, RWS, en Deltares.

Toegevoegde waarde: Zonder satellietdata te gebruiken was het opstellen van deze kaart veel moeilijker en tijdrovender geweest, want dan had men over langere tijd handmatig moeten weten. Daarnaast door verschillende kaarten over elkaar te leggen kon er nieuwe informatie worden ontsloten, bijvoorbeeld dat grondwater een belangrijke factor is in bodemdaling.

Factsheet Dataproject/pilot

Satellietdata monitoren kwaliteit Eems-Dollard (DGWB/WOM)

Naam: Eems-Estuarium

Omschrijving/doel: De huidige monitoring heeft een tweewekelijkse frequentie op enkele puntlocaties in het estuarium. Dit is ontoereikend om voldoende gebiedsdekkende informatie over waterkwaliteit en primaire productie te verkrijgen. Omdat een sterke uitbreiding van het grondmeetnetwerk onbetaalbaar is, is RWS op zoek naar nieuwe (kosten)efficiëntere meetstrategieën. Daarbij kunnen satellietdata mogelijk een oplossing bieden. De inzet van dergelijke nieuwe technieken lijkt veelbelovend, maar is zeker nog niet bruikbaar. Met name door de complexiteit in ondiep troebel water met getij, door de grote variatie in ruimte en tijd en de interacties met bodemreflecties, ligt voor dit gebied een grote uitdaging bij het bepalen van de basisvariabelen in het algemeen en de primaire productie in het bijzonder.

Organisatie onderdeel/opdrachtgever:

Hanneke Brouwer (DGWB/WOM/BOW)

Contactpersoon: Bert Doze en Charlotte Schmidt (RWS)

Periode: 2017/2018

Uitvoerder/datalab:

RWS in samenwerking met het Netherlands Space Office (NSO). Het NSO is het ruimtevaartagentschap van de Nederlandse rijksoverheid, IenW is een van de opdrachtgevers van het NSO. Er is een SBIR (EZK subsidieregeling voor innovatieve projecten) call uitgezet, waarop bedrijven zich konden inschrijven. Uiteindelijk is het bedrijf Water Insight BV geselecteerd om deze opdracht uit te voeren.



Gebruik van databronnen/datatechnieken etc.: De Europese Commissie en het Europees ruimtevaartagentschap ESA hebben in 2001 het gezamenlijke initiatief gelanceerd voor een mondiaal monitoringsysteem ten behoeve van milieu en veiligheid: Copernicus. De European Space Agency lanceert hiervoor een serie satellietmissies, de Sentinels genaamd. Er is hier gebruik gemaakt van Sentinel-3. Sentinel-3 (oceanografie en landgebruik) is een multi-instrument missie voor het meten van zeeoppervlak topografie, zee- en landoppervlaktetemperatuur, oceaan kleur en landoppervlak, met radar-alimetrie, infrarood radiometer en hyperspectrale sensoren.

Resultaat: Het is duidelijk geworden dat de bouwstenen, nodig voor de succesvolle toepassing van aardobservatie in de schatting van primaire productie, gerealiseerd kunnen worden. Een dataset van kaarten van alle relevante en gevraagde parameters voor de berekening van de primaire productie voor de Eems-Dollard regio is geleverd en beschikbaar gesteld. Algemeen is geconcludeerd is dat met Sentinel 3 voldoende grote wateren met hoge frequentie gemonitord kunnen worden en dat detailopnamen van Sentinel 2 voldoende vaak per jaar langskomen om de lokale structuren, stromingspatronen, hoogtij- laagtij dynamiek etc. vast te leggen.

Toegevoegde waarde: Satellietdata zijn van toegevoegde waarde voor het meten van de waterkwaliteit van de Eems-Dollard. Satellietdata leveren vlakdekkende informatie in tegenstelling tot de huidige beperkt aantal puntmetingen. Met deze vlakdekkende informatie kunnen gefundeerde afwegingen en beleid worden ontwikkeld om de waterkwaliteit te verbeteren.

Factsheet Dataproject/pilot

Big data, 1 miljoen woningen en hitte stress (DGRW/RO)

Naam: Hitte stress

Omschrijving/doel: De beleidsopgave van Nederland is om tot 2040 1 miljoen additionele woningen te bouwen. Deze nieuwe woningen zullen naar verwachting voornamelijk binnen de grote steden in Nederland worden gerealiseerd. En uitgaande van een trek naar de stad, wat betekent dat voor de druk op de bebouwde omgeving in de grootste Nederlandse gemeenten? Verder zullen burgers steeds meer worden blootgesteld aan het specifieke (lokale) klimaat in de stad. Dit laatste is een extra uitdaging als gevolg van een veranderend klimaat en de temperatuur die naar verwachting zowel ruraal als in de stad toeneemt. Wat betekent dit voor de bouwopgave en op wat voor manier moet er dan gebouwd worden? Hoe om te gaan met hittestress?

Organisatie onderdeel/opdrachtgever:

Hans Tijl, directeur Directie Ruimtelijke Ontwikkeling van DGRW

Contactpersoon: Rien Bout, Nieske Bisschop (DGRW), Jan-Willem Nooteboom (KNMI)

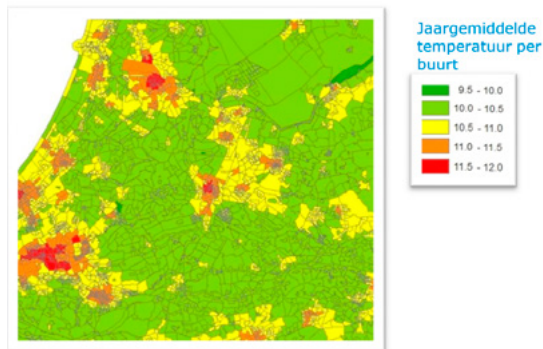
Periode: 2017

Uitvoerder/datalab: KNMI, PBL, WUR

Gebruik van databronnen/datatechnieken etc.: Temperatuurbestanden en klimaatmodellen in combinatie met data analyse.

Resultaat: De beleidsmatige meerwaarde van het onderzoek is dat het onderzoek gebaseerd is op metingen, dus feitelijke data in plaats van modelberekeningen. Fictieve bebouwingsscenario's zijn meegenomen in de analyse. De casus is voor een tweetal gebieden in Den Haag concreet uitgewerkt. Met de gehanteerde methode kun je flexibel, snel en effectief zaken doorrekenen. Hiermee kun je ook helder maken welke aspecten niet zijn meegenomen. Zo is in de analyse het effect van water in de stad niet meegenomen omdat dit wetenschappelijk (methodologisch) niet robuust te onderbouwen is. Ook het effect van wind op bebouwing is niet meegerekend. Ondanks dat een aantal aspecten niet konden worden meegenomen heeft het project wel uitermate nuttige informatie opgeleverd. Voor de gemeente Den Haag kunnen de inzichten gebruikt worden in het proces van besluitvorming rond nieuw te bouwen woningen. Een van die inzichten is dat extra hoogbouw tot minder hittestress leidt dan het bebouwen van groen in de stad. Bij extra hoogbouw kan de groenvoorziening naar het lijkt nl. behouden blijven.

Toegevoegde waarde: Door inzet van data-analyse is gebleken dat extra hoogbouw tot minder hittestress leidt dan het bebouwen van groen in de stad. Met deze kennis kan de gemeente Den Haag een gefundeerd beleid op het gebied van woningbouw ontwikkelen.



Factsheet Dataproject/pilot

Opkomende Stoffen (DGWB)

Naam: Opkomende stoffen

Omschrijving/doel: Opsporen van opkomende, potentieel gevaarlijke stoffen, door de toepassing van big data technieken. Op de verschillende manieren die hieronder beschreven staan wordt getracht uit bestaande data meer informatie te halen. Er is veel data, maar in dit project wilde men met een helicopterview bekijken wat voor informatie er nog meer in deze bestaande data 'verborgen' zou kunnen zijn. Er werd gezocht naar een patroon in het voorkomen van stoffen in plaats van langdurige hoge concentratie, om eens andere stoffen te ontdekken dan via de gebruikelijke methode.

Organisatie onderdeel/opdrachtgever:

Landelijke Werkgroep Opkomende Stoffen
Marjan van Giezen (DGWB/WOM/WKK)
Contactpersoon: Saskia Onnik,
Arjan van der Put (ILT)



Uitvoerder/datalab: Datalab RWS

Gebruik van databronnen/datatechnieken etc.:

Technieken:

- **Machine learning:** In grote datasets met informatie over stoffen, of verschillende bestanden met informatie over stoffen kan door gebruik te maken van algoritmes gezocht worden naar verbanden en patronen in de datasets. Dit heeft als doel meer zicht te krijgen op eventuele schadelijkheid en relevantie van een nieuwe stof.
- **Non-target screenings:** Het relateren van screening resultaten van meetstations aan ecologische data om zo nieuwe inzichten of patronen te ontdekken in de gevonden mengsels.
- **Internet scrapings:** Internationaal wordt er ook gemeten op zorgwekkende (en deels niet-genormeerde) stoffen. Deze data van het bovenstrooms gebied van de Maas wordt vergeleken met data die wordt opgehaald in het benedenstrooms gebied. Verwacht wordt dat de verschillen tussen beide gebieden inzicht zullen opleveren.
- **Tekstmining:** Het onderzoeken van relevante zoektermen in verschillende tekstbestanden (patentaanvragen, innovatiesubsidieaanvragen en vergunningaanvragen, inclusief de bijbehorende inhoudelijke rapporten). Deze resultaten worden voornamelijk gebruikt om nieuwe opkomende stoffen te vinden die argwaan wekken.

Databronnen:

- Screening resultaten meetstations RWS
- Data internationaal Maasverband
- Patentaanvragen, innovatiesubsidie aanvragen, vergunningaanvragen, inhoudelijke rapporten

Resultaat: Sommige technieken hebben al nieuwe inzichten opgeleverd, andere technieken blijken lastiger uitvoerbaar. Door het er echter met projectpartners over te hebben worden ook weer nieuwe inzichten vergaard. Tegenslagen kunnen dus ook wat opleveren.

Toegevoegde waarden: Door de informatie die er al was te combineren en te gebruiken als databron kon er informatie uit gehaald worden die anders onder de radar was gebleven.

Factsheet Dataproject/pilot

Luchtkwaliteit (DGMI)

Naam: Luchtkwaliteit en smart/green mobility

Omschrijving/doel: Onderzoek op de relatie tussen het weer en doorstroming van het wegverkeer, en de relatie tussen de doorstroming van het wegverkeer en de luchtkwaliteit onderzoeken.

Organisatie onderdeel/opdrachtgever: DGMI

Uitvoerder/datalab: Datalab RWS & ILT

Gebruik van databronnen/datatechnieken etc.:

Technieken:

- Machine learning
- Realtime monitoring

Databronnen:

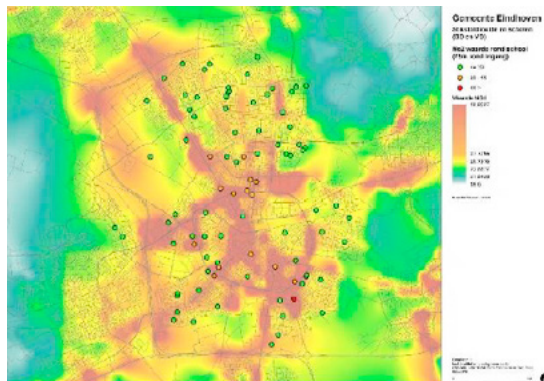
- Satellietdata en -modellen
- Weerdata en -modellen (KNMI)
- Verkeersdata (bijv. stoplichtdata, maar ook real-time data van bijv. Google, verkeerscentrales RWS, RDW gegevens)
- Luchtkwaliteitsdata (van RIVM)

Resultaat: Bij de collega's van DGMI hebben ze al een datapilot met betrekking tot luchtkwaliteit uitgevoerd. De startvragen hadden betrekking op de relatie tussen het weer en doorstroming van het wegverkeer, en de relatie tussen de doorstroming van het wegverkeer en de luchtkwaliteit. Op basis van deze relaties werd er voorspeld wat de luchtkwaliteit zou zijn. Deze pilot werd uitgevoerd in een gebied van 10 bij 12 km bij Eindhoven vanwege de hoge data-dichtheid die daar gevonden werd. Uit de pilot bleek dat de beoogde voorspelling over de luchtkwaliteit mogelijk lijkt, onder voorbehoud van satellietdata, weerdata en verkeersdata.

Een aantal belangrijke leerpunten uit de pilot waren:

- Visualisatie is een belangrijke sleutel in de communicatie over data;
- Machine learning in combinatie met realtime data heeft een hoge innovatie-potentie;
- Gereed maken van de data om ze te gebruiken en te combineren kost veel tijd.

Toegevoegde waarde: Door data op een nieuwe manier te combineren kan er meer informatie worden gegenereerd omdat er verbanden worden gelegd die eerder niet duidelijk waren. Daarnaast kan er aan de hand van deze pilot ook worden geëxperimenteerd met het voorspellen van luchtkwaliteit, in plaats van alleen achteraf of real-time resultaten op dit gebied te kunnen leveren.



Deel 4

Achtergrondinformatie

1. Handig om te weten

I-team

Bij elk DG is een informatievoorziening (IV)-team ingericht. Een I-team helpt beleid haar beleidsopgaven te realiseren door de inzet van informatievoorziening, data en ICT.. Het I-team van DGWB kun je op [intranet](#) vinden.

Opleidingen

Zelf meer leren over hoe te werken met data is natuurlijk ook goed mogelijk. Via het Leerportaal van IenW worden de volgende cursussen aangeboden:

- **Data Science voor Beleid – managerial track**

De cursus Data science voor beleid is één van de twee leerlijnen uit de “Data Science voor IenW” reeks. Deze managerial track richt zich op beleidsmedewerkers en zorgt ervoor dat deze na het volgen van deze cursus bekend zijn met concepten rondom data driven policy making, business modeling, en het transformeren naar een datagestuurde organisatie. Daarnaast stelt de cursus beleidsmedewerkers in staat om een klein datagestuurd proces of beleidsvraagstuk binnen IenW aan te pakken en wordt de cursist de handvatten geboden om met specialistische ondersteuning complexere datagestuurde processen in te richten.

- **Data Science voor IenW: technical track**

De technical track data science is één van de twee leerlijnen uit de “Data Science voor IenW” reeks. Deze technical track richt zich op de technische vaardigheden die nodig zijn om data science uit te voeren in de praktijk. Voor deze cursus is programmeerervaring en/of gedegen kennis van statistiek gesteld als voorkennis.

Data lab

Een datalab is een plek waar met data wordt gewerkt en waar dus de data wordt verzameld en geanalyseerd. Vaak is het ook een plek waar geëxperimenteerd kan worden met verschillende technieken en waar er ruimte is om nieuwe ideeën uit te werken.

AOMD team

Binnen IenW is er het 'Anders Omgaan met Data' team, dat zich bezig houdt met het stimuleren van dataprojecten. Zij opereren IenW breed, en hebben dus ervaring op verschillende beleidsterreinen.

Digishape

DigiShape is een samenwerkingsinitiatief in Topsector Water en Maritiem tussen overheidspartijen, waaronder IenW en RWS, het bedrijfsleven en kennisinstituten om data-innovatie in de watersector effectiever en efficiënter te laten plaatsvinden en de digitalisering van de watersector te versnellen. Momenteel worden innovatieprojecten opgezet (gelieerd aan en betaald uit bestaande beleidsprogramma's), wordt verbreding naar andere partners in de gouden driehoek gezocht, en wordt een governance-structuur ontwikkeld.

DCI datateam

Ook DCI heeft een data-team dat het AOMD team o.a. ondersteunt in het ontwikkelen van de DAT-omgeving, en het dataregister van IenW. De focus ligt in de toekomst meer op het ontwikkelen van gemeenschappelijke dienstverlening, en minder op het ontwikkelen van producten.

IenW dataportaal

IenW heeft een eigen dataregister dat op dit moment nog in ontwikkeling is. In het dataregister kan men zoeken naar datasets, maar worden er ook relevante datasets uitgelicht. Zodra de website live is zal hier een link naar het register worden opgenomen.

CBS dataportaal

Als Centraal Bureau voor de Statistiek heeft het natuurlijk veel data in huis. Deze data bieden ze veelal ook aan in de vorm van open data. Het CBS heeft verschillende websites waarop ze open data aanbieden. Via [Statline](#) kun je datasets vinden over bijvoorbeeld milieu en bodemgebruik.

Data Analyse Toolset (DAT)

De DAT biedt een omgeving waar toepassingen zijn verzameld waarmee data analyses kunnen worden uitgevoerd. Het gebruik van de toepassingen vergt wel enige kennis en vaardigheid op het gebied van data-analyse. Daarmee is de beleidsmaker niet de primaire doelgroep, maar meer de data analisten en data scientists. Het bijzondere van de DAT is dat het in de cloud draait en er dus geen programma's hoeven te worden geïnstalleerd op de computer van de gebruiker en dat informatie in de cloud kan worden opgeslagen. Informatie gaat hierdoor minder makkelijk verloren gaat.

Data scientist & data analyst

Deze twee begrippen worden vaak door elkaar gebruikt, en het verschil wordt niet door iedereen op dezelfde manier gezien. Het grootste verschil tussen de scientist en de analyst is dat er over het algemeen van de scientist wordt verwacht op een technischere manier om te kunnen gaan met data. Denk aan het zelf code kunnen schrijven om modellen en algoritmes te produceren. Analisten richten zich (zoals de naam al zegt) meer op de analyse van wat er dan uit die modellen en algoritmes komt.

Data methodes

Om te laten zien wat er zoal mogelijk is wordt hieronder een aantal manieren genoemd om data te verkrijgen en te verwerken.

- **Webscraping** (verzameltechniek): manier om (ongestructureerde) data van een website om te zetten in informatie (database of spreadsheet meestal). Het programma haalt de data van de database op de website en zet deze om in een eigen dataset.
- **Textmining** (analysetechniek): proces om grote hoeveelheden ongestructureerde tekst te analyseren en te verkennen. Textmining kan een vorm van webscraping zijn, als de teksten van bijvoorbeeld van internetfora worden gehaald om sentiment analyse te doen. Bij sentiment analyse worden bijvoorbeeld comments gescand op bepaalde emoties.
- **Monitoring** (verzameltechniek): analyseren van gegevens die een beeld geven van het recente verleden.

- **Realtime monitoring** (analysetechniek): analyseren van gegevens die een beeld geven van de actuele situatie. Dit kan door middel van bijvoorbeeld een dashboard met realtime data.
- **Predictive analysis** (analysetechniek): verbanden die toekomstige problemen en behoeften voorspellen op basis van in het verleden verzamelde data.
- **Profiling** (analysetechniek) analyseren wat er op individueel niveau speelt en in de toekomst zal spelen
- **Data-discovery** (analysetechniek): big data, ontdekken van verbanden tussen diverse gegevens door data-analyse.

Het project Opkomende stoffen maakte gebruik van bijv. textmining, data-discovery in big data en kan worden gebruikt voor (realtime) monitoring en predictive analysis.

De bodemdalingskaart is het resultaat van monitoring en kan daar in de toekomst ook voor worden ingezet.

Artificial Intelligence

Artificial Intelligence (AI), of Kunstmatige Intelligentie (KI), is alle intelligentie die niet biologisch is. Bij AI leren machines acties en beslissingen van mensen te kopiëren. Ze redeneren logisch, begrijpen taal en leren van fouten. AI is gebaseerd op machine learning-technologie: computers die complexe algoritmes toepassen met zelflerend vermogen.

Dataregister

Hieronder staat een aantal andere websites waarop data te vinden is die relevant kunnen zijn voor het domein water en bodem.

Naam + link	Eigenaar	Doel/inhoud	Type data	Hoe te gebruiken	Werkveld
Noordzeeloket	Rijksoverheid	Overheidsinformatie over de Noordzee regels, voorwaarden, vergunningen, (inter)nationaal beleid).	Atlas actueel (data over beleid en beheer, gebruik en watersysteem)	Online, interactieve kaart	Water
Atlas Leefomgeving	Rijksoverheid	Kaarten over o.a. bodem, water, ruimte, natuur	Kaartdata	Interactieve kaart	Water + bodem
Bodemloket	Rijkswaterstaat	In kaart brengen bodemkwaliteit.	Bodemkaart	Online, interactieve kaart	Bodem
CBS Open Data Statline Bodemgebruik	CBS	Informatie over bodemgebruik, monitor Nota Ruimte 2008	Metadata, onbewerkte dataset, dataset voor grafische weergave	Zelf data gebruik, mogelijkheid om op de site zelf grafieken te maken	Bodem
CBS Open Data Statline Milieu	CBS	Emissies naar water en waterzuivering	Metadata, onbewerkte dataset, dataset voor grafische weergave	Zelf data gebruik, mogelijkheid om op de site zelf grafieken te maken	Water
CBS Open Data Statline Verkeer en vervoer	CBS	Binnenvaart, zeevaart	Metadata, onbewerkte dataset, dataset voor grafische weergave	Zelf data gebruik, mogelijkheid om op de site zelf grafieken te maken	Water
Data.overheid.nl	Rijksoverheid	Open data m.b.t. o.a. landbouw, natuur en milieu, ruimte en infrastructuur	Datasets (ODS, CSV), websites	Data downloaden en zelf gebruiken	Water + bodem
Dataportaal Rijkswaterstaat	Rijkswaterstaat	Data mb.t. o.a. binnenwater, klimatologie, natuur en milieu	Datasets	Datasets downloaden, of op de websites de interactieve kaart gebruiken	Water + bodem

2. Theorie van Gartner

Het bekende onderzoeksbureau Gartner heeft een model ontwikkeld waar op vier niveaus wordt aangegeven hoe data kan helpen inzicht te krijgen in beleidsvraagstukken. Het model beschrijft hoe men kan gaan van terugkijken door middel van data, naar inzicht krijgen door data naar voorspellen aan de hand van data. Daarnaast beschrijft het model een ladder met verschillende stadia van werken met data. Het begint met het beschrijven van een situatie en werkt toe naar het inzetten van beleid en de effecten daarvan te kunnen voorspellen. Bij elke trede horen verschillende vragen die je kan stellen, waaronder hiervan een aantal voorbeelden staan.

Eerste niveau: descriptive analytics = beschrijvende analyse

Wat is het verschijnsel, wat is de omvang, hoe vaak komt het voor, etc.

Voorbeeld: *Wat is bodemdaling? Waar komt het voor?*

Tweede niveau: diagnostic analytics = diagnosticerende analyse

Wat zijn de oorzaken van het verschijnsel, waar komt het vandaan, waarom gebeurt het, hoe is het verschijnsel ontstaan, etc.

Voorbeeld: *Waarom daalt de bodem? Wat zijn de oorzaken?*

Derde niveau: predictive analytics = voorspellende analyse

Kunnen we voorspellen waar het verschijnsel gaat plaatsvinden, wat de omvang zal zijn, wat zullen de risico's zijn, etc.

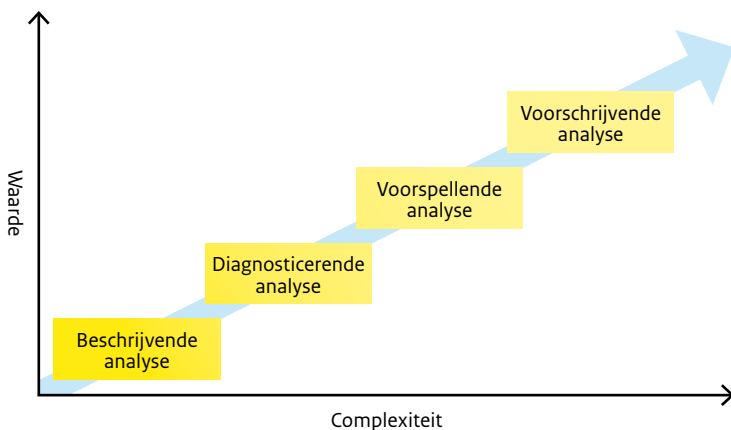
Voorbeeld: *Kunnen we voorspellen waar de bodem nog meer of nog verder gaat dalen? Wat voor gevolgen heeft dat? Welke risico's brengt dat met zich mee?*

Vierde niveau: prescriptive analytics = voorschrijvende analyse

Welke maatregelen kunnen we nemen/maken om het verschijnsel te voorkomen? Wat gebeurt er als we geen maatregelen nemen?

Voorbeeld: *Zijn er manieren om bodemdaling te voorkomen? Welk beleid kunnen we voeren om bodemdaling tot stilstand te brengen?*

Zoals je kan zien in het model, worden de vragen steeds lastiger te beantwoorden en wordt er naar gelang de vraag moeilijker wordt ook meer inhoudelijk inzicht gevraagd. De eerste vragen zouden met statistische modellen beantwoord kunnen worden, terwijl er voor het derde niveau al intelligentere modellen zouden moeten worden gebruikt.



Bron: Gartner

Voorbeeld

Het big data pilot project Opkomende stoffen is een voorbeeld van een voorspellende analyse. Door verschillende datasporen te gebruiken en naar het grotere geheel te kijken, wordt geprobeerd een methode te ontwikkelen aan de hand waarvan voorspellingen kunnen worden gedaan over nieuwe stoffen die onderzocht moeten worden op schadelijkheid.

Dit is een uitgave van het

Ministerie van Infrastructuur en Waterstaat

Postbus 20901 | 2500 EX Den Haag
www.rijksoverheid.nl/ienw

April 2019