

Datagestuurd werken: 15 lessen uit Zaanstad

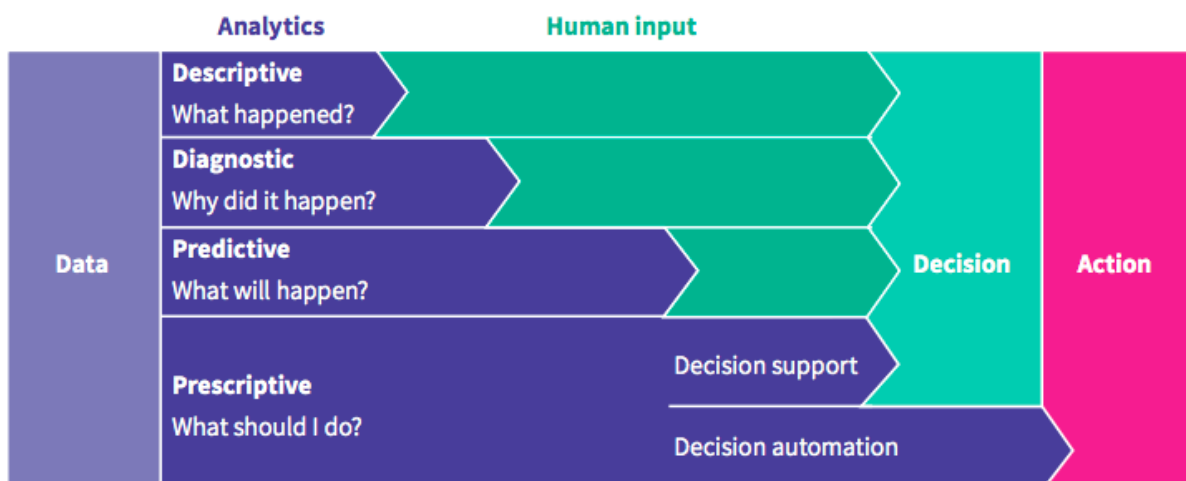
Bij veel publieke organisaties bestaat het geloof dat slim omgaan met data leidt tot betere resultaten voor burgers, zoals schonere straten, veiliger buurten, betere voorzieningen, etc. Maar hoe kom je van dit geloof tot daadwerkelijk betere resultaten op basis van data? Dit vraagt meer dan een data-analist aannemen en die een paar opdrachten laten uitvoeren. Waar begin je? Wat zijn essentiële onderdelen van je aanpak? Hoe begeleid je deze verandering? Wat moet je organiseren? Wat zijn belangrijke valkuilen en succesfactoren? In Zaanstad zijn we in 2016 begonnen. We hebben met vallen en opstaan mooie resultaten geboekt en veel geleerd. Begin 2019 heb ik de belangrijkste lessen gedeeld in een artikel. Het is nu 2023 en tijd voor een update. In dit artikel vind je de belangrijkste lessen van de afgelopen jaren.

Les 1: Creëer een gemeenschappelijke taal

In Zaanstad is door alle leidinggevende in de organisatie besloten in onze informatievoorzieningsstrategie (zie les 6) om meer datagestuurd te gaan werken. Maar datagestuurd werken is een gekke abstractie, net als veel andere abstracties die we in publieke organisaties vaak gebruiken, zoals 'de lerende organisatie', 'opgave gestuurd werken' of 'gebiedsgericht werken'. Ik weet nooit zo goed wat ze betekenen. Het bijzondere is dat als je een aantal willekeurige mensen in jouw organisatie vraagt wat dit soort termen betekenen dat je heel verschillende antwoorden krijgt. Dit gold ook voor datagestuurd werken. De vraag is: als je niet weet wat is het is hoe word je het dan?

Bovendien zagen we in de eerste datapilots dat we niet alleen verschillende betekenissen hadden bij datagestuurd werken, maar dat we ook allerlei termen, zoals big data, datagestuurd werken, business rules, data science, machine learning, onderzoek & statistiek, algoritmes, business intelligence, kriskras door elkaar gebruikten met wisselende betekenissen.

De vraag is dus, wat is datagestuurd werken? Binnen de gemeente Zaanstad gebruiken we onderstaand model van Gartner uit te leggen wat datagestuurd werken kan betekenen in het dagelijks werk. Je kan elk maatschappelijk vraagstuk hiervoor gebruiken. Ik doe dit aan de hand van fraude:



- Beschrijvende analyse: wie, wat, waar, hoeveel met behulp van dashboards of monitors. Wie fraudeert, waar komt het voor en hoe groot is de omvang van fraude?

- Diagnostiserende analyse: waarom met behulp van lokaal of wetenschappelijk onderzoek. Waarom wordt in die wijk veel en in die wijk weinig gefraudeerd, wat zijn de oorzaken van fraude, waarom doen mensen het en hoe doen ze het?
- Voorspellende analyses: wanneer met behulp van machine learning. Kunnen we voorspellen waar fraude gaat plaatsvinden en in welke situaties komt het vaak voor?
- Voorschrijvende analyse: voorschrijven van acties op basis van voorgaande analyses. Wat moeten we doen om fraude te voorkomen?

Datagestuurde werken in de echte wereld is dus data-analyses (beschrijvend, diagnostiserend, voorspellende en/of voorschrijvende) gebruiken bij jouw dagelijkse werkzaamheden. Bij het onderdeel 'organiseer capaciteit: datasetatelier en data-analyseteam' wordt nader toegelicht wie deze analyses kunnen uitvoeren.

Les 2: Start met de juiste analyse

We zijn met datagestuurde werken gestart op de hype van big data. De ontwikkeling van big data gaat vooral over vooruitkijkend (voorspellend en voorschrijvend). Het vakgebied data science werd door veel organisaties gezien als een zeer interessante ontwikkeling. Met behulp van machine learning werd geprobeerd voorspellende analyses te maken op basis van veel data. Organisaties waren aan het experimenteren met data science pilots. Wat is de waarde van voorspellende analyses voor onze organisatie? Ook wij wilden dit ervaren en zijn bij de start drie data-science pilots gestart in het maatschappelijk domein. We kwamen er in het proces al snel achter dat de eerste stap altijd eerst goed terugkijken is met behulp van dashboards (beschrijvend) en lokaal en wetenschappelijk onderzoek (diagnostiserend). Als de beschrijvende en diagnostiserende analyses niet bestaan of gebruikt worden, dan ben je bij voorspellende en voorschrijvende analyses kansloos vanaf het begin.

Les 3: Machine learning vraagt om afgebakende context en data

Gemeenten hebben in de afgelopen jaren geoefend met data science pilots, maar de voorspellende analyses met behulp van machine learning bleken keer op keer te mislukken. Het aantal succesvolle machine learning oplossingen in gemeenteland zijn op één hand te tellen. Dit heeft twee belangrijke oorzaken:

1. Het ontbreken van een afgebakende context.

Niemand twijfelt dat zelfrijdende auto's er komen, maar voorlopig alleen op de snelweg en niet in de stad. Zelfrijdende auto's zijn goed tot de rand van de stad. De sensoren op een auto registreren wel waar andere verkeersgebruikers zich bevinden, maar kunnen slecht voorspellen wat ze gaan doen. De stad is als context te ingewikkeld. De snelweg is een afgebakende context, want iedereen gaat dezelfde kant op met ongeveer dezelfde snelheid. De context van gemeentelijke vraagstukken is vaak te complex om machine learning succes te kunnen toepassen.

2. Bestaande gemeentelijke data niet geschikt voor machine learning

Het succes van data science is afhankelijk van de data die gebruikt wordt. Waarom kan Prorail voorspellen wanneer er onderhoud aan het spoor noodzakelijk is? Sensoren in het spoor. Hoe kon het voorspellende model rond kindermishandeling ontwikkeld worden? Grote hoeveelheid open tekst velden van artsen. En onlangs hebben de onderzoekers van de Massachusetts Institute of Technology (MIT) machine learning toegepast om kanker te voorspellen. Het model is getraind op data van bijna 30.000 patiënten die zijn gediagnosticeerd met 1 van 22 soorten kanker. Hier was

het mogelijk om met de beschikbare data en machine learning voorspellingen te doen. Echter, het merendeel van de bestaande gemeentelijke datasets zijn niet geschikt om voorspellende analyses mee te ontwikkelen voor complexe maatschappelijke vraagstukken.

Een afgebakende context inclusief daarbij horende data zijn twee belangrijke randvoorwaarden voor het succesvol toepassen van machine learning. Sociale en maatschappelijke contexten zijn te complex, waardoor in deze situaties machine learning tot op heden maar beperkt succesvol is.

Les 4: Stop met het voorspellen van menselijk gedrag

Machine Learning is een onderdeel van Artificial Intelligence (AI) dat zich toespitst op technieken waarmee computers kunnen leren op basis van data. Veel successen met AI zijn behaald in situaties waarin de spelregels eenduidig zijn (bijvoorbeeld bij schaken of Alpha Go) of als de gewenste uitkomst makkelijk als goed of fout te classificeren is. Arvind Narayanan, hoogleraar aan de Princeton-University, verdeelt de toepassing van AI in drie groepen:

1. Waarneming: oprecht snelle technologische vooruitgang

Content identificatie (zoals Shazam), gezichtsherkenning, medische diagnoses van scans, spraak naar tekst en deepfakes.

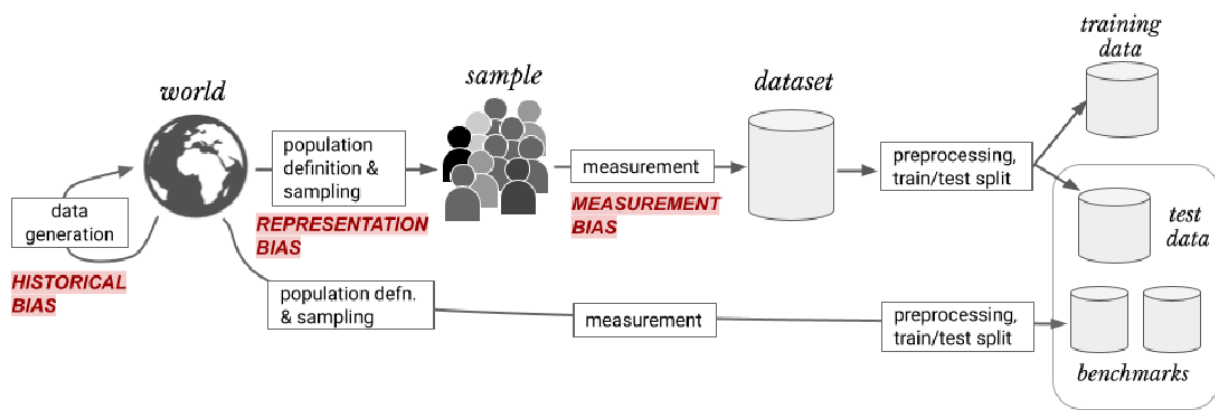
2. Geautomatiseerd oordeel: verre van perfect, maar het verbeterd

Spamdetectie, detectie van auteursrechtelijk beschermd materiaal, geautomatiseerde beoordeling van essays, detectie van haatspraak en content aanbevelingen

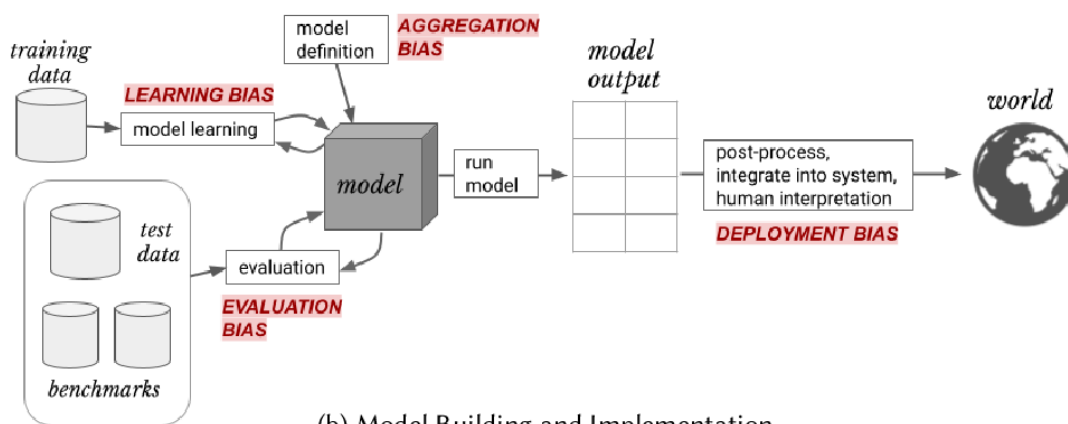
3. Voorspellen van sociale uitkomsten: fundamenteel dubieus

Voorspellen van criminele recidive, voorspellen van werkprestaties, voorspellen criminaliteit voorspellen van terroristische risico's, voorspellen van risicokinderen

Het voorspellen van menselijk gedrag met AI blijkt dus fundamenteel dubieus. Wat is hiervoor de reden? Bias, ofwel vooroordelen. Suresh en Gutttag van MIT beschrijven 7 vooroordelen in het proces van het creëren van data tot het ontwikkelen van voorspellende analyses.



(a) Data Generation



(b) Model Building and Implementation

1. Historische vooroordelen

Als je in google bijvoorbeeld raad van bestuur of engineer intypt, dan krijg je veel mannen.

2. Representatie vooroordelen

Als je geen goede representatie van de doelgroep gebruikt bijvoorbeeld bij gezichtsherkenning, dan doet de toepassing het wel bij blanke mensen en veel minder bij mensen met een donkere huidskleur.

3. Meet vooroordelen

Als je te simpele indicatoren (bijvoorbeeld om een succesvolle student te meten) gebruikt, als je op een verschillende manier meet per groep (bijvoorbeeld voor het voorspellen criminaliteit) als de accuraatheid van de data verschilt per groep (bijvoorbeeld registraties van blanke agenten)

4. Leer vooroordelen

De reeks aannames die gebruikt worden om de output van bepaalde inputs te voorspellen die je niet bent tegengekomen. Dit gebeurt omdat het model leert, vanwege de beperkte capaciteit, om de meest voorkomende functies te bewaren.

5. Aggregatie vooroordelen

Als je een one-size-fits-all model gebruikt voor de hele populatie met verschillende subgroepen (bijvoorbeeld bij de analyse van twitter, dan ziet een algemeen model niet de verschillende dialecten of straattaal)

6. Evaluatie vooroordelen

Een niet-representatieve benchmark stimuleert de ontwikkeling en implementatie van modellen die alleen goed presteren op de subset van data die worden vertegenwoordigd door de benchmarkgegevens (bijvoorbeeld foto's van vrouwen met donkeren huidskleur waren ondervertegenwoordigd in de benchmark)

7. Inzet vooroordeel

Dit ontstaat wanneer er een mismatch is tussen het probleem dat een model moet oplossen en de manier waarop het is daadwerkelijk gebruikt (bijvoorbeeld het model dat voorspelde of een persoon terug viel in crimineel gedrag werd gebruikt om de duur van de straf te bepalen)

Arvind Narayanan geeft aan dat machine learning grote kansen laat zien op het gebied van waarneming en geautomatiseerd oordeel, maar niet voor het voorspellen van sociale uitkomsten. Op basis van bovenstaande vooroordelen hebben we in Zaanstad het uitgangspunt dat machine learning algoritmes niet gebruikt worden voor het voorspellen van menselijk gedrag. In algemene zin zijn wij op dit moment nog terughoudend in het inzetten van voorspellende en voorschrijvende analyses. Bijvoorbeeld bij de voorschrijvende analyse is menselijke handelen vaak grotendeels 'out of the equation' is. Dit is voor veel maatschappelijke vraagstukken onwenselijk, maar niet voor alle processen. Een mooi voorbeeld zijn chatbots die op dit moment in de publieke sector bij klantcontactcentra massaal hun intrede doen. Hier is deze techniek heel kansrijk en zo zijn meer voorbeelden in de publieke sector.

Les 5: Creëer een paar successen en bouw daar een krachtig verhaal omheen

Zoals gezegd zijn we gestart met een groot aantal pilots in verschillende domeinen. We waren op zoek naar een paar lokale successen die onomstotelijk de waarde van datagestuurde werken laten zien. Hieronder twee voorbeelden in de begin periode veelvuldig verteld zijn:

- In Keulen zaten vluchtelingen aan vrouwen. In de Europese media werd al snel geroepen dat er alleen testosteronbommen binnen kwamen. Dit werd door de landelijke media overgenomen. We hebben toen uitgezocht hoe het voor Zaanstad zat en uit de beschrijvende analyse bleek dat het aantal mannen en vrouwen ongeveer gelijk was. Bovendien was het gedeelte dat als testosteronbom kon worden bestempeld, maar een klein gedeelte van het geheel. Uit de analyse bleek echter ook dat 75% van alle statushouders in het verleden niet uit de bijstand is gekomen. Deze feiten hebben we in het politieke debat gebracht, waardoor het niet over testosteronbommen ging, maar over hoe krijgen we ze aan het werk. Een hele simpele beschrijvende analyse die echt geholpen heeft.
- In Zaanstad staan veel panden op (korte) houten funderingspalen. Met name grenenhout is kwetsbaar voor bacteriële aantasting. Hierdoor neemt de draagkracht van de fundering af en ontstaat er schade aan een pand. Funderingsherstel is dan noodzakelijk. We hebben met machine learning een voorspellend model ontwikkeld waarmee funderingsproblematiek in Zaanstad voorspeld kon worden..

Het doel van deze verhalen was om andere inhoudelijke teams te verleiden om ook proberen een maatschappelijk vraagstuk aan te pakken met data en de waarde van datagestuurde werken te tonen aan de organisatie.

Les 6: Creëer ruimte door de macht van de organisatie te laten kiezen voor data

Binnen de gemeente Zaanstad kwam data een aantal jaar geleden op de agenda zoals bij veel organisaties. 'Big data' was op dat moment één van meest besproken onderwerpen op congressen. De leidinggevende van de IT-afdeling had dit opgemerkt en besloot dat we hier iets mee moesten. Aan mij de vraag om dit op te pakken en te onderzoeken wat de kansen en mogelijkheden waren voor onze organisatie.

Aangezien de keuze om iets te doen met datagestuurd werken geen bewuste keuze van de hele organisatie was wilde ik die expliciete keuze ophalen in een gedragen informatievoorzieningsstrategie. In de IV-strategie dwingen we onszelf om jaarlijks naar nieuwe ontwikkelingen te kijken en samen te duiden: willen we er iets mee of niet?

Om te komen tot een gedragen IV-strategie zijn we gestart met het interviewen van college, raad, directie, strategen, informatiemanagers, IT-leverancier, bank, sociaal wijkteam, zorginstelling, woningbouwcorporatie met de vraag welke relevante ontwikkelingen komen op de gemeente af en waar moeten wij iets mee? Uit deze ronde zijn twaalf thema's naar voren gekomen, die vervolgens aan alle leidinggevendenden van de gemeente Zaanstad zijn voorgelegd met de vraag: welke drie thema's moeten opgepakt worden en welke negen gaan we niet doen? Bijna net zo belangrijk als de top 3 is de keuze om negen thema's niet te doen.

Er zijn drie thema's geprioriteerd, waarbij het thema datagestuurd werken met stip op 1 stond. Het risico van deze interventie is dat het thema dat jij persoonlijk heel relevant vindt niet geprioriteerd kan worden. Het proces om thema's binnen en buiten de organisatie op te halen en daarna te laten prioriteren door alle leidinggevendenden had een aantal voordelen

1. Het was niet meer mijn thema, maar dat van de organisatie.
2. Er stond een hek om de verandering, want de macht van de organisatie had gekozen
3. Inzet van capaciteit (geld en mensen) werd veel eenvoudiger

Het gevolg was dat ik twee jaar na de keuze van de leidinggevende van de IT-afdeling om iets met datagestuurd werken te gaan doen niet alleen verantwoordelijk was om resultaten te boeken. We waren met 10-15 personen onze ambities op het gebied van datagestuurd werken aan het realiseren. Elk jaar scherpen we als team onze ambities voor het komende jaar aan. Wat willen we bereiken in het komende jaar?

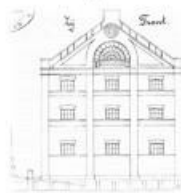
Individueel idee, onze keuze

Beoogde resultaten programma datagestuurd werken in 2023-2024

1. Data Lab



2. Datapakhuis



3. Data-academy



4. Informatieketens verbeteren



5. Toevoegen data-analyses aan teams door datateam



Les 7: Doelen zijn altijd het startpunt

Eén van de grootste valkuilen bij het maken van data-analyses (in wat voor vorm dan ook) is het starten met het spel van vraag en aanbod. De meeste dataprojecten beginnen met de vraag: wat willen we weten? En dat is niet de juiste vraag. De eerste vraag moet zijn: wat willen we bereiken? Vaak zijn de doelen niet duidelijk, abstract of zelfs afwezig. Het is een reflex om bij een vraag direct aan de gang te gaan en data te leveren en analyseren. Ik heb gemerkt dat als je niet weet wat je wilt bereiken, je zeker niet weet wat je wilt weten en dan is de kans op goed gebruikte data-analyses (in wat voor vorm dan ook) heel klein. Het startpunt zijn de concrete doelen op het voorliggende maatschappelijke vraagstuk. Als een team of afdeling de doelen niet scherp heeft, dan gaan we niet van start. Een fictief voorbeeld om de relatie tussen doel en data te verhelderen:

1) Doel: wat willen we bereiken?

We willen mensen die in de bijstand zitten naar werk begeleiden.


2) Data: wat willen we weten?

- Hoeveel mensen zitten in de bijstand? (beschrijvend)
- Wat zijn de kenmerken van die mensen? (beschrijvend)
- Hoeveel mensen zijn uitgestroomd? (beschrijvend)
- Welke interventies hebben wel/geen effect? (diagnosticerend)
- Waarom zitten mensen in de bijstand? (diagnosticerend)
- Hoeveel mensen zitten in de bijstand in 2025? (voorspellend)
- Wie komt er in de bijstand? (voorspellend)
- Etc.

Les 8: Ga op zoek naar kansrijke dataprojecten

Ik heb in het verleden heel vaak in kansloze dataprojecten gezeten zonder dat ik het wist en om uiteenlopende redenen: ontbrekende doelen, data ongeschikt, niet duidelijk wat het team wil weten, etc. Op basis van die ervaring heb ik samen met professor Frans Feldberg het Datagedreven Innovatiecanvas (www.DatCan.nl) voor de publieke sector ontwikkeld. DatCan zorgt dat je als team een gemeenschappelijk begrip ontwikkelt van het 'waarom', 'wat' en 'hoe' van een dataproject en dat je deze gedurende de projectuitvoering voortdurend in relatie tot elkaar beschouwt. Deze drie kernvragen zijn essentieel voor ieder dataproject, maar makkelijker gesteld dan beantwoord. Helemaal in de context van dataprojecten die per definitie multidisciplinair zijn, waar 'technuten' en 'domein experts' ieder op hun eigen manier de antwoorden invullen. Hoe borg je dat alle betrokkenen over hetzelfde praten en elkaar begrijpen? De structuur van DatCan ondersteunt een proces waarin deze drie kernvragen aan de hand van deelvragen, die in logische clusters zijn gegroepeerd, concreet worden ingevuld. Het canvas is dus geen doel, maar een hulpmiddel om gedurende het dataproject zo veel mogelijk te borgen dat er gedeeld begrip bestaat over het 'waarom', 'wat' en 'hoe'. Een belangrijke randvoorwaarde om samen succesvol te zijn.

Hoe			Wat	Waarom		
Welke partijen hebben we nodig om het vraagstuk op te lossen?	Welke data hebben we hiervoor nodig?	Wat moeten we hiervoor doen?	Wat willen we weten om onze doelen beter te kunnen realiseren? <ul style="list-style-type: none"> • Beschrijvend • Diagnosticerend • Voorspellend • Voorschrijvend 	Wat zijn onze doelen?	Wat is de kern van het vraagstuk?	Aan welke opgave is dit probleem gerelateerd?
Welke partijen willen we erbij betrekken?			Wat zijn de ethische en privacy aspecten?		Voor wie is dit een probleem?	Wie is opdrachtgever? Wie is eigenaar van het vraagstuk?
Welke kennis hebben we hiervoor nodig?			Wat moeten we organiseren om de oplossing succesvol te kunnen gebruiken?			
Wat is het Budget? Wat zijn de benodigde investeringen?						

 datcan.nl
datagedreven innovatie canvas

Les 9: Werken met data betekent je onder dompelen in de data

Ik zie heel vaak dat mensen denken dat datagestuurd werken is als voor een groot scherm gaan staan en wachten tot de inzichten tot je komen. Zo werkt het natuurlijk niet. Na het ontwikkelen van een dashboard (beschrijvende analyse), het uitvoeren van een onderzoek (diagnosticerende analyse) is één ding zeker: het gaat niet zomaar goed gebruikt worden. Niet voor niets hebben wij een rijke historie van niet gebruikte dashboards en onderzoeken in onze organisatie. Dit is geen opzet, maar het gevolg van het ontbreken van een werkwijze na de data-analyses om deze goed te gebruiken. Werken met data betekent je onder dompelen in de data en samen inzichten uit de analyses halen, samen tot betere oplossingen komen en deze samen realiseren. Dit vraagt om een geloofwaardig proces met het team:

Proces na een data-analyse

Van data naar realiseerbaarheid



1. Van data-analyse naar inzichten

Welke inzichten halen de teamleden op basis van de doelen uit de data-analyses? Deze inzichten worden gedeeld op een muur en met de gebruikers besproken, zodat er een lijst met gemeenschappelijke inzichten ontstaat.

2. Van inzichten naar grenzeloze oplossingen

Wat zijn op basis van de inzichten grenzeloze oplossingsrichtingen voor het vraagstuk? In deze vraag gaat het om de logica tussen het inzicht en oplossingsrichting. Deze hoeft dus niet haalbaar te zijn. Als het maar logisch is op basis van het inzicht.

3. Van oplossingen naar haalbaarheid

Om de haalbaarheid te bepalen worden drie vragen beantwoord:

- a) Wat moet er grofweg gedaan worden om de oplossingsrichting te realiseren?
- b) Wie zijn de relevante stakeholders per oplossingsrichting? Zijn die al georganiseerd in een gremium? Zo ja, welk gremium?
- c) Waarom gaat het niet lukken om de oplossingsrichting te verbeteren?

4. Van haalbaarheid naar realisatie

Om de realisatie te starten staan de volgende vragen centraal:

- a) Wat moet er geregeld worden om de realisatie te starten?
- b) Wie gaat het trekken?
- c) Hoe gaan we voortgang volgen?

Op basis van deze werkwijze wordt 'afgedwongen' dat de opgeleverde data-analyses (in wat voor vorm dan ook) daadwerkelijk gebruikt worden.

Les 10: Ontwikkel een middel om complexe vraagstukken samen met partners aan te pakken op basis van data

Opgaven in de samenleving laten zich niet vangen door grenzen van een organisatie. Iedereen denkt na over datagestuurd werken. Kan data ons verbinden? Vanuit dat perspectief zijn we in de regio Zaanstreek Waterland een aantal jaar geleden gestart met het Data Lab. Het doel van de eerste pilot was om de waarde hypothese te bewijzen, dus kunnen we een middel ontwikkelen dat helpt om samen met partners maatschappelijke vraagstukken een beetje beter maken op basis van data. We hebben Parteon, GGD, Politie, gemeente Purmerend, Agora, Eveen, UWV en VU/UvA toentertijd gevraagd om mee te doen, samen de methodiek te ontwikkelen en de waarde voor andere maatschappelijke vraagstukken te beoordelen. Hieronder de aanpak op hoofdlijnen:

De methodiek van het Data Lab

Data Lab verbindt

1. Maatschappelijke opgave prioriteren
2. Samen feitelijke foto maken
3. Veilig delen data
4. Verdiepende data-analyses
5. Oplossingen bepalen vanuit eigen rol en positie deelnemers



1. Maatschappelijke vraagstuk prioriteren

In deze stap wordt duidelijk wat de doelen van de deelnemende organisaties zijn, wat de kern van het vraagstuk is en wat we willen weten om het vraagstuk aan te pakken. Na deze fase kan een organisatie bepalen in of uit te stappen. Extramuralisering & overlast is het eerste maatschappelijke vraagstuk dat opgepakt is.

2. Samen feitelijke foto maken

In de tweede stap worden alle relevante informatie uit factsheets en onderzoeken gehaald die al in de verschillende organisaties aanwezig zijn over het voorliggende thema, zodat we in beeld krijgen wat we samen al weten en welke verdiepende analyses nog uitgevoerd moeten worden. De inhoudelijk experts extramuralisering & overlast van de deelnemende organisaties zeiden na de feitelijke foto: “wat weten we al veel zonder dat we het weten, wat ligt er al veel zonder dat we het gebruiken”.

3. Veilig delen data

Bij elke data-analyse speelt privacy een belangrijke rol en deze opgave wilden voor alle toekomstige vraagstukken in het Data Lab slechten. Het is ons gelukt om een werkwijze te ontwikkelen waarin de privacy gewaarborgd is en het mogelijk maakt om data te gebruiken voor analyses. Het doorlopen van de aanpak veilig delen data zorgt ervoor dat de je erachter komt wat wel en niet mag. Een deelnemer zei na deze stap dat hij nog niet eerder had meegemaakt dat zulke gevoelige data tussen organisaties daadwerkelijk gedeeld werd op een veilige manier.

4. Verdiepende data-analyses

In de vierde fase worden verdiepende analyses (beschrijvend, diagnosticerend en voorspellend) uitgevoerd.

5. Oplossingen bepalen vanuit eigen rol en positie deelnemers

In fase 5 is het beschouwen afgelopen. Er mogen geen vragen meer gesteld worden. Nu gaan we aan de slag met inzichten en oplossingen (zie les 9)

1. Wat zijn inzichten (op basis van stap 2 - feitelijke foto en stap 4 - verdiepende analyses)?
2. Wat zijn grenzeloze oplossingsrichtingen op basis van deze inzichten?

3. Welke oplossingsrichtingen zijn haalbaar?
4. Welke oplossingsrichtingen zijn realiseerbaar?

Een paar voorbeelden van gemeenschappelijke inzichten uit het eerste Data Lab: beperkt gebruik van WMO door mensen met verward gedrag, slechts 10 personen zorgen voor grootste gedeelte van overlastmeldingen, meldingen zijn gecentreerd, maar zorg verspreid, beperkte regie op de meldingen en veel meldingen over vervuiling. Op basis van deze inzichten zijn er oplossingsrichtingen vormgegeven.

Na het eerste Data Lab waren alle deelnemers positief over de waardehypothese van het Data Lab voor andere maatschappelijke vraagstukken. Het doel was vervolgens om het Data Lab door te ontwikkelen van pilot naar ecosysteem waar maatschappelijke organisaties (zorginstellingen, politie, GGD, woningcorporaties, scholen, etc.), wetenschap (VU/UvA) en gemeenten gezamenlijk aan opgaven in de samenleving werken op basis van data. In de eerste pilot hebben we de methodiek voor toekomstige Data Labs ontwikkeld en inmiddels hebben 35 gemeenten en maatschappelijke partners de intentie uitgesproken om onderdeel te zijn van het Data Lab en bij voor de organisatie relevante vraagstukken deel te nemen. Vraagstukken die in de opstart, uitvoering of afgerond zijn:

- In de regio Zaanstreek Waterland: extramuralisering & overlast, jongens/meisjes in het onderwijs, versnellen woningbouwopgave, bedrijventerrein molletjesveer Noorderveld, schuldenproblematiek, huiselijke geweld, brandveiligheid en kwetsbare burgers,
- Buiten de regio Zaanstad: de methodiek van het Data Lab is voor iedereen beschikbaar die samen met partners aan de slag wil met maatschappelijke vraagstukken op basis van data. De methode is inmiddels 5 keer toegepast bij kindermishandeling (associatie voor jeugd), Langer veilig thuiswonen (Tis), Verkleinen gezondheidsverschillen (Alphen aan den Rijn), veilige en leefbare werk- en woonomgeving en kwetsbare burgers (Zeist) en huiselijk geweld (Almere)

De methodiek van het Data Lab is uitgewerkt en voor iedereen beschikbaar die samen met partners aan een maatschappelijk vraagstuk wil werken op basis van data. Het doel is dat elk jaar zowel binnen als buiten de regio Zaanstreek Waterland meerdere Data Labs worden uitgevoerd voor maatschappelijke vraagstukken die zich niet laten vangen door de grenzen van een organisatie.

Les 11: Organiseer capaciteit om data te ontsluiten (datacluster) en analyseren (data-analisten) en stimuleer de samenwerking in het datateam

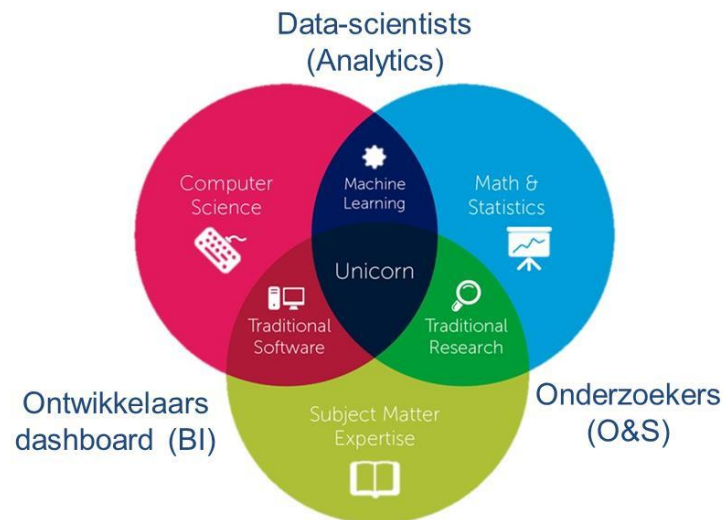
Als je aan de slag wil met maatschappelijke vraagstukken op basis van data dan heb je twee soorten capaciteit nodig. Wie gaat data ontsluiten? Wie gaat data analyseren? Beide vragen moeten verankerd worden in je organisatie:

1. Datacluster ontsluit data

Het datacluster zijn de data-architecten en data-engineers die datasets prepareren vanuit de bron en deze publiceren in het datapakhuis in twee verschijningsvormen (zie volgende onderwerp - ontsluit in samenhang alle beschikbare data en visualiseer die in dashboards).

2. Data-analist analyseert data

De volgende vraag is wie maken de beschrijvende, diagnosticerende en voorspellende analyses? In Zaanstad gebruiken we onderstaand model om deze vraag te beantwoorden.



- Dashboard ontwikkelaar voor beschrijvende analyses
Beschrijvende analyses zijn bij uitstek geschikt om te automatiseren in dashboards. Binnen de gemeente Zaanstad gebruiken we Qlikview voor de beschrijvende analyses.
- Onderzoekers voor diagnosticerende analyses
Combineer je wiskunde & statistiek met inhoud dan gaat het over traditioneel onderzoek. Relevant om de diagnosticerende analyse te maken.
- Data-scientists voor voorspellende analyses
Ten slotte is er een nieuwe discipline bijgekomen. Data-scientists die machine learning toepassen om voorspellende analyses te maken.

In Zaanstad ligt de focus op dit moment vooral op het goed gebruikt krijgen van beschrijvende (Qlikview dashboards) en diagnosticerende analyses (lokaal/wetenschappelijk onderzoek), waardoor we op dit moment geen data-scientists nodig hebben (zie les 1). De data-analisten in Zaanstad zijn nu dashboardontwikkelaars en onderzoekers. Met de snelheid waarmee we als samenleving digitaliseren en ongekende ontwikkeling van machine learning is de kans groot dat dit over een aantal jaar ook data-scientist zijn, maar voor nu zijn dit de twee data-analisten die we nodig hebben.

We hebben de ambitie om team voor team data-analyse beter te gebruiken in onze organisatie (zie les 13) en de samenwerking tussen het datacluster en de data-analisten is daar essentieel in. Data-analisten zitten in allerlei teams in onze organisatie. We hebben het datacluster en de data-analisten samengevoegd (niet hiërarchisch) in het datateam en we hebben samen een werkwijze ontwikkeld van vraag tot goed gebruik van data-analyse, hebben doelstellingen geformuleerd en bespreken periodiek alle inkomende datavragen.

Les 12: Ontsluit in samenhang alle beschikbare data en visualiseer die in dashboards

We beschikken als gemeente over een groot aantal gegevensbronnen, die tot een aantal jaar geleden beperkt, dan wel slecht ontsloten waren. Dit maakt het delen van data tussen medewerkers, laat staan met maatschappelijke partners en burgers een tijdrovende en complexe bezigheid. Bovendien is niets zo frustrerend als een data-project te willen starten, maar de datasets zijn niet beschikbaar. Onze data-architecten hebben toen het concept Datapakhuis bedacht. Het Datapakhuis faciliteert expliciet het delen van data met de binnen- en buitenwereld, waarbij de privacy gewaarborgd is.

Inmiddels hebben we alle datasets van de gemeente onder de knop zitten in twee vormen: ADS en ODS.

- Operational Data Store (ODS): datasets met direct herleidbare gegevens.
- Analyse Data Store (ADS): datasets met geaggregeerde en gepseudonimiseerde data zodat datasets gecombineerd kunnen worden waarbij privacy gewaarborgd is.

Daarnaast is elke dataset in samenhang ontsloten, dus bsn (versleuteld), maand (tijd), wijk en buurt (plaats). Doordat alle datasets in samenhang zijn ontsloten is data-analist niet meer vijf dagen bezig met datapreparatie, maar heeft een dashboardontwikkelaar binnen een paar uur beeld en kan een onderzoeker veel sneller aan de slag.

Les 13: Creëer de ruimte om informatieketen te mogen verbeteren

Op een gegeven moment explodeerde de vraag naar dashboards in onze organisatie. Een groot aantal teams wilde beschrijvende analyses om tot betere oplossingen te komen voor het vraagstuk waar zij aan werken. Als de onderkant van een informatieketen op orde is dan is een dashboard als een muurtje schilderen in je huis, ongeveer een halve dag werk. Echter, als de onderkant van de informatieketen niet op orde is dan is het als een huis bouwen op vervuilde grond, na 1,5 jaar werk heb je nog steeds niets, terwijl de vraag exact hetzelfde is. Vaak heeft de vrager het verschil tussen 3 dagen en 1,5 jaar niet in de gaten.

Als een informatieketen niet goed functioneert herken je dit aan het volgende: het kost keer op keer heel veel tijd om relatief simpele datavragen te antwoorden, het is onduidelijk aan wie de vraag gesteld kan worden, er zijn wel dashboards (of analyses) die niemand gebruikt en begrijpt, als na veel pijn en moeite toch een antwoord komt wordt continu getwijfeld of het antwoord klopt.

Een informatieketen bestaat uit 6 lagen van onder naar boven:

1. Administratieprotocol

Het administratieprotocol is een samenhangende set aan afspraken die gelden voor het voeren van een kwalitatief hoogwaardige administratie in een bronapplicatie. Het is de basis voor registratie in de bronapplicaties, processchema's, werkinstructies en contracten met leveranciers.

2. Registraties in bronapplicaties

In de bronapplicaties wordt de administratie gevoerd van de informatieketen, processen uitgevoerd en wordt data gecreëerd

3. Datapakhuis

In het datapakhuis wordt data ontsloten en gevalideerd, zodat het gebruikt kan worden.

4. Visualisaties in dashboards

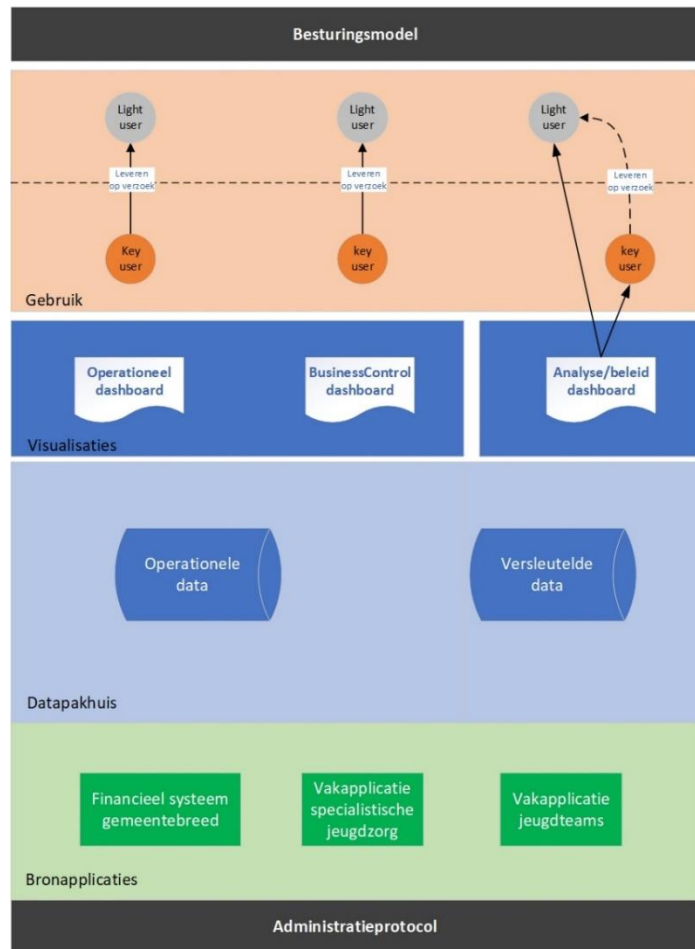
Op basis van gevalideerde datasets uit het datapakhuis kunnen visualisaties van datasets gemaakt worden in dashboards of in eenmalige beschrijvende analyses.

5. Gebruik

In het gebruik gaat het over de kennis en kunde om de data te kunnen gebruiken om de opgaven op basis van data aan te pakken.

6. Besturingsmodel

Hier wordt beschreven wat 1) doelen: wat willen we bereiken? 2) projecten/processen: Wat gaan we doen om deze doelen te realiseren? 3) verantwoordelijkheid: wie gaat wat doen? 4) indicatoren: hoe gaan we dat kwalitatief en/of kwantitatief volgen? 5) overlegstructuur: wanneer, waarover, hoe en met wie?



In Zaanstad bepalen we met behulp van de informatieketen welke verbeteringen er allemaal nodig zijn om data (bijvoorbeeld in dashboards) goed te kunnen gebruiken. We doen dit als volgt:

1. samen de analyse per laag van informatieketen te maken (en jaarlijks te actualiseren)
2. samen verbeteringen te bedenken per laag van de informatieketen op basis van de analyse
3. samen de verbeteringen te prioriteren voor het komende jaar
4. samen de aanpak (op activiteitniveau) vorm te geven om de verbeteringen te realiseren
5. samen de benodigde capaciteit (expertise en tijd) op activiteitniveau te organiseren
6. samen te bepalen wanneer de realisatie van de verbetering gaat starten
7. samen 2-wekelijks sturen op activiteitniveau met de spelers die de verbetering realiseren

Het ontwikkelen van een dashboard kan al onderdeel zijn van de eerste ronde geprioriteerde verbeteringen, maar vaak zijn eerst andere verbeteringen noodzakelijk, zoals het versterken van de besturing, het valideren van data, het ontwikkelen van een administratieprotocol, het verbeteren van registraties, het organiseren van capaciteit van key-users, het ontwikkelen van een datafundament, het beleggen van het eigenaarschap, het ontwikkelen van een werkwijze om tot goed gebruikte dashboards te komen, etc.

Les 14: Doe het leren & inspireren niet alleen, maar altijd over grenzen heen

Bij de start zijn we begonnen met kennissessies om inspiratie van buiten te halen om de mogelijkheden te ontdekken van werken met data. De sessies waren alleen voor de ambtenaren van Zaanstad. Op een gegeven moment stelden we onszelf de vraag: waarom nodigen we de omliggende gemeenten en regionale partners eigenlijk niet uit? We kunnen dan het leren en inspireren samendoen. Er worden in de Data Academie trainingen, masterclasses en inspiratiesessies georganiseerd voor de gemeenten in Zaanstreek Waterland en maatschappelijke partners in de regio. In de Data Academie staan de volgende uitgangspunten centraal: (1) Inspiratie & leren. Lokale successen van binnen en buiten de organisaties en mensen leren werken met data. (2) Kansen & mogelijkheden. Op basis van de verhalen van de lokale successen proberen we interesse te kweken om te gaan experimenteren met datagestuurde werken. Het doel is een leergemeenschap die die leert en ontwikkelt over de grenzen van de organisaties heen.

Les 15: Stapel lokale verbeteringen

Datagestuurde werken betekent in onze organisatie het gebruiken van beschrijvende, diagnosticerende, voorspellende analyses per inhoudelijk thema. De hype zit op het voorspellen, én de gemeente Zaanstad, andere gemeenten en de publieke sector in zijn geheel hebben een wereld te winnen op het goed gebruiken van beschrijvende en diagnosticerende analyses. Als je naar een gemeente kijkt dan bestaat deze uit grofweg 50 inhoudelijke thema's van vroegtijdig schoolverlaten, funderingen, armoede, werkgelegenheid, jeugdzorg, woningbouw, etc. Het is de uitdaging om per inhoudelijk thema scherp te krijgen welke analyse door het team nu gebruikt wordt en dan een analyse toe te voegen.

- Bij het team woningbouw bleek de beschrijvende analyse niet op orde, dus zijn we bij deze analyse (een woningbouw monitor) gestart.
- Bij het team huiselijk geweld lag het zwaartepunt op diagnosticerende analyse (lokaal en wetenschappelijk onderzoek).
- Bij de teams jeugd, WMO en participatie hebben we beschrijvende analyses (operationele dashboards) ontwikkeld.
- Etc.

Het idee hierachter is dat door de stapeling van lokale verbeteringen, waarbij inhoudelijke teams data-analyses (in wat voor vorm dan ook) beter toepassen, we als organisatie in zijn geheel beter zijn in het gebruiken van data bij het aanpakken van maatschappelijke vraagstukken.

Datagestuurde werken van idee tot concrete resultaten

Hieronder een overzicht van de belangrijkste lessen die wij hebben geleerd op onze weg naar het verbeteren van maatschappelijke vraagstukken op basis van data:

1. Creëer een gemeenschappelijke taal
2. Start met de juiste analyse
3. Machine learning vraagt om afgebakende context en data
4. Stop met het voorspellen van menselijk gedrag
5. Creëer een paar successen en bouw daar een krachtig verhaal omheen
6. Creëer ruimte door de macht van de organisatie te laten kiezen voor data

7. Start een dataproject altijd bij de doelen: als je niet weet wat je wil bereiken, dan weet je zeker niet wat je wil weten en is de kans op een goed gebruikte analyse nihil
8. Ga op zoek naar kansrijke dataprojecten
9. Werken met data betekent je onder dompelen in de data en samen inzichten uit de analyses halen, samen tot betere oplossingen komen en deze samen realiseren
10. Ontwikkel een middel om complexe vraagstukken samen met partners aan te pakken op basis van data
11. Organiseer capaciteit om data te ontsluiten (datacluster) en analyseren (data-analisten) en stimuleer de samenwerking in het datateam
12. Ontsluit in samenhang alle beschikbare data (bijv. Datapakhuis) en visualiseer die in dashboards
13. Creëer de kans om informatieketens te mogen verbeteren
14. Doe het leren & inspireren niet alleen, maar altijd over grenzen heen (bijv. Data Academy)
15. Stapel lokale verbeteringen

In de komende jaren gaan we in Zaanstad verder om alle inhoudelijke teams die een gemeente rijk is, van jeugdzorg tot woningbouw, te leren werken met beschrijvende (monitors/dashboards), diagnosticerende (lokaal/wetenschappelijk onderzoek) en voorspellende analyses.