

De weg naar data-gedreven mobiliteitsbeleid

Simone Schiebergen – Provincie Noord-Brabant – sschiebergen@brabant.nl
Miranda Driessen – ALTEN Nederland – miranda.driessen@alten.nl

**Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk
21 en 22 november 2019, Leuven**

Samenvatting

Data en informatie speelt een steeds belangrijkere rol in onze samenleving. Eén van de belangrijkste taken van de overheid is het maken van beleid. Hoe kunnen we de vele data efficiënt en zorgvuldig gebruiken om ons beleid te bepalen? Welke keuzes maak je als overheid en waarop zijn deze keuzes gebaseerd? Ook de burger eist steeds vaker transparantie over gemaakte keuzes.

Deze paper gaat in op de weg naar meer data-gedreven mobiliteitsbeleid en de keuzes en ervaringen die de provincie Noord-Brabant hiermee heeft opgedaan. Het opzetten van een dataplatform om grote hoeveelheden data te kunnen opslaan, analyseren, interpreteren en combineren met andere data. Of we het nu hebben over actueel verkeersmanagement, assetmanagement, knelpuntenanalyses of het combineren van mobiliteitsdata met data vanuit de beleidsterreinen energie of ruimte ontwikkeling. Hoe maken we slim gebruik van bestaande landelijke dataknooppunten, zoals het Centraal Bureau voor Statistiek of databank wegverkeer?

Het geven van inzichten binnen de organisatie, maar ook buiten de organisatie aan burgers, marktpartijen en regionale overheden. Data opwerken tot informatie en vervolgens visualiseren in dashboards of delen met externe partners.

Klinkt mooi! Maar zoals bij de meeste projecten zijn er ook uitdagingen en leerpunten welke we graag met jullie delen. Met een mooie technische oplossing alleen ben je er nog niet. Net zo belangrijk zijn de governance en goede procesafspraken.

1. Inleiding

1.1 Achtergrond

Het maken van beleid is één van de belangrijkste taken van de overheid. Om te komen tot onderbouwde beleidskeuzes zijn data en informatie cruciaal. Waar men in het verleden handmatig data verzamelde en analyseerde is er tegenwoordig data in overvloed. Dit brengt nieuwe kansen en uitdagingen met zich mee. Immers wat is precies het doel en de vraag? Welke databronnen gebruik je? Wat is de kwaliteit van de data? Hebben we te maken met privacygevoelige data en hoe gaan we hier mee om? En zo zijn er nog vele vragen rondom het gebruik van data anno 2019.

Daarnaast zien we dat de omgeving, zoals burgers en regionale overheden, een belangrijkere rol speelt bij het maken van beleid en zouden vraagstukken integraal opgepakt moeten worden. Voor het maken van mobiliteitsbeleid spelen namelijk ook milieuaspecten, ruimtelijke ontwikkeling en energie een rol. Hoe kunnen we tot onderbouwde keuzes komen op basis van (gecombineerde een eenduidige) data en informatie?

Om meer grip te krijgen op de vele data en het kunnen combineren van data tot informatie en om zo te komen tot meer data-gedreven beleid, heeft de provincie Noord-Brabant een dataplatform ontwikkeld. We vertellen de lezer hier graag meer over en delen onze ervaringen in deze paper.

1.2 Structuur paper

De paper gaat in op het meer data-gedreven maken van beleidsanalyses en –keuzes, met een focus op mobiliteitsbeleid en de ervaringen die tijdens dit proces zijn opgedaan.

In hoofdstuk 2 wordt ingegaan op de aanleiding van het project en de knelpunten die we binnen mobiliteit ervaarden.

Hoofdstuk 3 vertelt meer over de technische keuzes en het dataplatform dat gebruikt wordt om het data-gedreven werken te faciliteren.

Hoofdstuk 4 gaat in op het proces en de uitdagingen en geeft een doorkijk in toekomstige ontwikkelingen.

In het laatste hoofdstuk komen aanbevelingen aan bod.

2. Aanleiding data-gedreven mobiliteitsbeleid

Al enige jaren staat het inwinnen, borgen en gebruik van data in Noord-Brabant op de ‘overheidsagenda’ en is onderdeel van diverse dialogen, maar het blijkt nog onvoldoende georganiseerd. Dit leidt tot praktische problemen, niet alleen binnen de provincie, maar ook bij onze stakeholders, zoals regionale overheden:

- Provinciale en gemeentelijke overheden worden vanuit verschillende invalshoeken gefragmenteerd verzocht om het organiseren en beschikbaar stellen van data. Dit is een extra taak waar de meeste regionale overheden qua kennis en capaciteit niet op zijn ingericht.
- Overheden geven onvoldoende of ieder op eigen wijze invulling aan de datavragen en werken langs elkaar heen.

- De inefficiënte werkwijze resulteert in een wisselende kwaliteit en er ontstaat geen compleet beeld van heel de provincie Noord-Brabant.
- Inefficiëntie leidt ook tot hogere kosten, geen kwaliteitsborging, geen kennisdeling en onvoldoende innovatie.
 - Dezelfde data wordt soms meerdere keren ingewonnen voor verschillende doeleinden.
 - Door verschillen in data en definities is het vergelijken soms niet of nauwelijks mogelijk.
 - Informatie wordt voor specifieke ad-hoc vraagstukken ingewonnen en door diverse marktpartijen geleverd. Onderliggende data ontbreekt vaak, wat herbruikbaarheid, maar ook trendanalyses lastig maakt.
 - Er gebeurde nog veel handmatig bij het inwinnen en verwerken van data.

Er is de laatste jaren ook veel vooruitgang geboekt op het gebied van mobiliteitsdata, zo zijn er regionaal vele pilots gestart en diverse “datalabs” opgezet. Dit vraagt echter ook om een betere verbinding, opschaling en het delen van data, informatie en kennis. Meer resultaat met minder inspanning!

2.1 Context en stakeholders

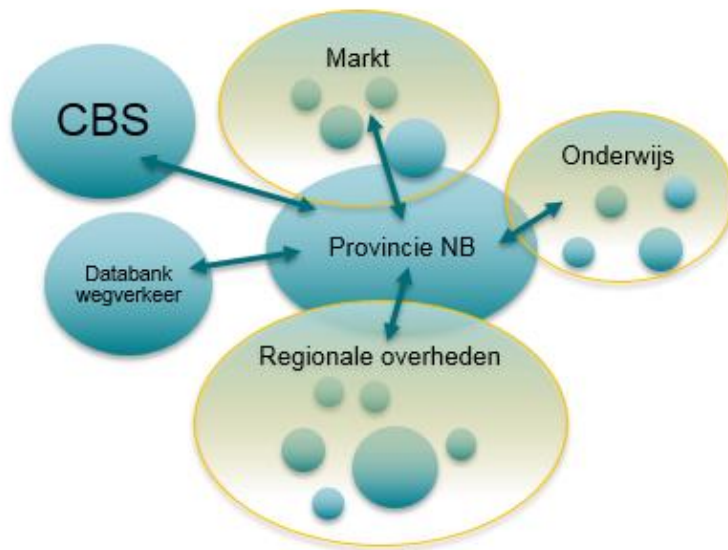
Als provincie zijnde hebben we een breed speelveld met diverse stakeholders; Rijk, regionale overheden, maar ook burgers en marktpartijen.

Er is nagedacht hoe we als provincie een rol kunnen spelen in de verschillende digitaliseringsopgaven binnen dit brede speelveld. Onder andere gebaseerd op afspraken met het Rijk om;

“In 2023 90% van de publieke mobiliteitsdata in Zuid-Nederland ingewonnen, ontsloten en beschikbaar gemaakt wordt en daarbij actueel, betrouwbaar en correct is en onderdeel is van de ‘staande’ organisaties van de overheden. Uitwerking van dit doel leidt tot de volgende prestaties:

- *Inwinning/ verzameling:*
- *Beschikbaar stellen:*
- *Tooling:*
- *Kwaliteit en standaarden:*
- *Veilig en betrouwbaar:*
- *Organisatie: “*

(Derksen & de Jong, 2019)



Afbeelding 1: Provincie Noord-Brabant en diverse externe stakeholders op het gebied van datasamenwerking

Om bovenstaande doelstellingen te kunnen realiseren diende er een omgeving te komen die hierin zou kunnen faciliteren. De provincie Noord-Brabant heeft gekozen om een provinciaal dataplatform te ontwikkelen dat enerzijds helpt bij het bereiken van bovenstaande gezamenlijke mobiliteitsopgave. En anderzijds kan ondersteunen bij een efficiëntere bedrijfsvoering binnen de provincie en het maken van eenduidige beleidskeuzes.

3. Provinciaal dataplatform

In dit hoofdstuk gaan we in op het provinciaal dataplatform dat is ontwikkeld.

“Een architectuur is meer dan code en componenten. Het is een combinatie van drie onderdelen: technologie, data en processen” (Madsen, 2015)

We definiëren het Provinciaal dataplatform als een technische data infrastructuur met daaromheen een organisatie en processen (governance) om efficiënt met de data te kunnen werken (het gebruik ervan).

Het doel van het dataplatform is simpelweg de schakel te zijn tussen het provinciaal beleid en beschikbare data om zo eenieder te kunnen voorzien van gevraagde informatie op een toegankelijke, veilige en efficiënte manier (ontzorgen).

De technische data infrastructuur is opgezet als een data lake:

Eén opslag voor alle data welke de provincie op enig moment zou willen (gaan) gebruiken voor analyses.

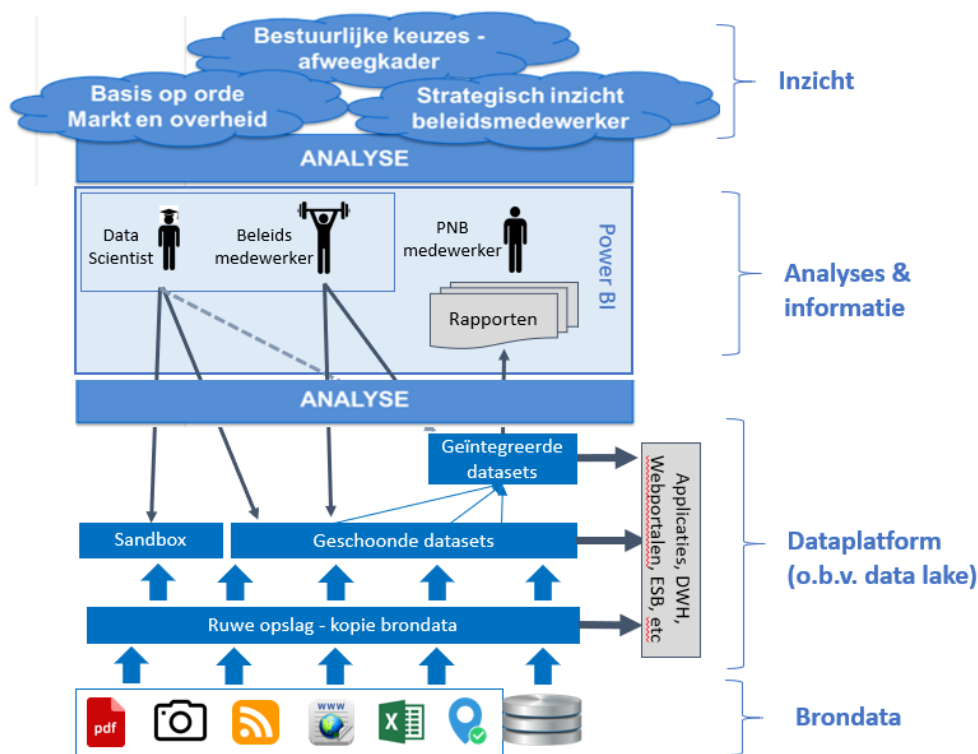
Uitgangspunt voor het data lake is dat het mogelijk moet zijn om alle soorten data, die beschikbaar zijn te verzamelen, ongeacht structuur, omvang, kwaliteit en gebruik.

Het verzamelen van de data staat los van databeheer, dus initieel wordt de data niet gestandaardiseerd en gemodelleerd, maar zo eenvoudig mogelijk opgeslagen, zodat deze beschikbaar kan worden gemaakt, wanneer iemand deze nodig heeft.

Alleen maar “ruw” opslaan van de data maakt het nog niet toegankelijk voor de beleidsmedewerkers, die er analyses over willen doen. Beleidsmedewerkers zijn geen data ontwikkelaars en willen niet bezig zijn met het ophalen en prepareren van hun data voor analyses. Data moet meestal gefilterd, geherstructureerd, opgeschoond worden, voordat analyse mogelijk is.

Daarom moet het mogelijk zijn in de datalake infrastructuur om schaalbare parallele processen te laten plaatsvinden, zodat er met de data gewerkt kan worden. Hierbij rekening houdend met de verschillende analysebehoefte (van eenmalige analyse ten behoeve van een beleidsvraagstuk tot dagelijkse rapportages, waarbij gewaarborgd is dat iedereen dezelfde kernregistraties gebruikt). Daarom is het handig om datasets en processen voor te bereiden. Elke benodigde dataset is uniek, met zijn eigen structuur, indeling, kwaliteit en afkomst. Er zijn verschillende technieken nodig om het te verzamelen, te catalogiseren en op te slaan zodat de informatie kan worden gevonden en gebruikt.

Bovenstaande structuur en mogelijkheden is weergegeven in het volgende plaatje:

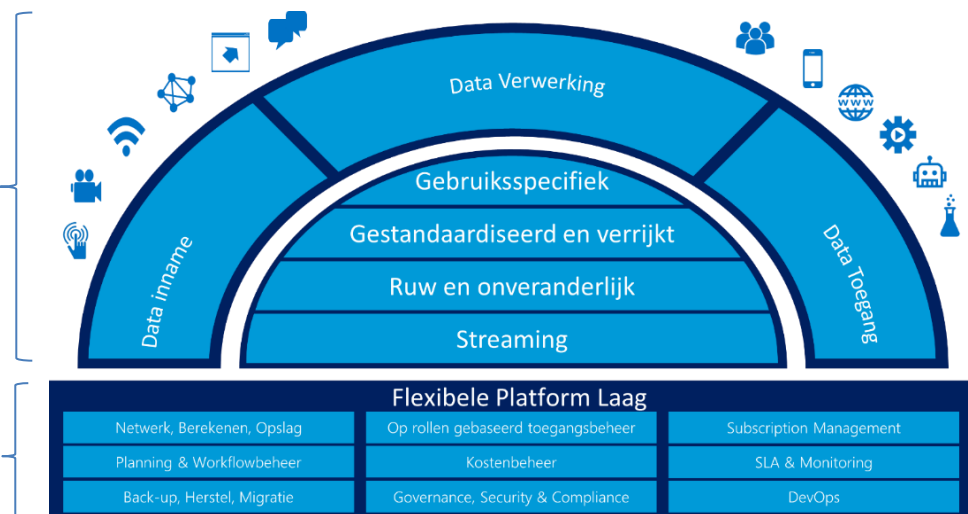


Afbeelding 2: structuur dataplatform provincie Noord-Brabant

Maar alleen verzamelen van data en onafhankelijk databeheer op basis van de informatievraag is niet voldoende. Er zijn standaarden en richtlijnen, wetgeving, security eisen, autorisatie vraagstukken en dergelijke waar aan voldaan moet zijn om juist datagebruik te waarborgen. Leuk dat we alle data verzamelen, maar niet iedereen mag altijd alles zien en combineren volgens de wetgeving. Maar wanneer er doelbinding is, moet het wel mogelijk zijn.

Daarom wordt de volgende standaard inrichting gebruikt:

1. Dit is de data- en applicatie laag van het data platform. Waarin alle originele, gewijzigde en afgeleide data bewaard wordt voor rapportage, analyse of beschikbaar gemaakt voor gebruikers of apps. De ronde vorm geeft aan dat elke toepassing/ apparaat data kan leveren, evenals data ontvangen. Beschikbaarheid, schaalbaarheid en voldoen aan de functionaliteit zijn de belangrijkste architectuur principes voor deze laag.
2. Dit zijn de generieke services die elk data platform moet bevatten. Het zorgt voor een basisniveau van controle en stelt zakelijke projecten in staat om data en functionaliteit op een juiste, beheerde en veilige manier toe te voegen.



Afbeelding 3: architectuurplaat dataplatform PNB (ALTEN Nederland; Provincie Noord-Brabant, 2019)

Hier is gekozen voor de Azure Cloud van Microsoft. Dit past tevens in de Cloud ambitie van de provincie, waarin een Cloud first beleid gekozen is.

De bovenste laag wordt vormgegeven door de verschillende beleidsprojecten. Deze paper gaat voornamelijk over het mobiliteitsbeleid, maar ook de andere onderdelen van de provincie maken gebruik van het dataplatform. Hierdoor ontstaat de situatie dat data gedeeld kan worden. In plaats van dat verschillende provinciale onderdelen, met hun eigen bril en business rules, dezelfde bron ontsluiten, zal dit nu eenmalig gebeuren. Het inzichtelijk maken van de data, middels Power BI of geografische kaarten (ESRI), leidt nu vaak tot de vraag hier nog een andere databron aan te koppelen.

Zo sluit het dataplatform ook aan bij het concept Common Ground, ten behoeve van het uitwisselen van gegevens tussen provincies en gemeenten. Hierbij kan voldaan worden aan privacywetgeving en tegelijk efficiënt worden omgegaan met data governance.

De onderste laag van 9 vlakken garandeert een basisniveau waarop alle data en dataprocessen uit de bovenste laag, goed, beheerd en veilig kunnen plaatsvinden. Ook zorgt het dat er op een slimme en schaalbare manier omgegaan kan worden met de functionaliteit. Hieronder een korte omschrijving per service:

1. Netwerk, berekenen en opslag: Het verplaatsen, verwerken en opslaan van data moet op een veilige manier. Door het toepassen en up-to-date houden van de best practices van Microsoft, blijft het platform optimaal ingericht.
2. Scheduling & Workflow management: Ook binnen de datalake infrastructuur, blijft scheduling en van de processen erg belangrijk. Daarbij zijn in de Cloud oplossingen ook schedules toegepast om kosten te besparen. Denk hierbij aan bijvoorbeeld het 's avonds terugschalen van databases, die dan toch niet gebruikt worden (Azure Automation).

3. Back-up en herstel: Binnen de Azure standaard, waarbij gekozen is voor back-up van data binnen Europa, waarbij altijd recovery van data (35 dagen) kan plaatsvinden.
4. Toegangsbeheer: Standaard wordt niemand rechtstreeks geautoriseerd op het platform. Personen kunnen uitsluitend rechten krijgen door een rol binnen de organisatie of bedrijfsproces toegewezen te krijgen. Ook de permissies op objecten/services binnen Azure kunnen worden toegekend op rollen.
5. Kostenbeheer: Het provinciale dataplatform is gebaseerd op “pay per use”. Per object/service worden kosten afgerekend. Wanneer hier niet gedegen naar gekeken wordt, kunnen de kosten onnodig hoog oplopen. Kostenbeheer op dag basis is dan ook van belang. Ook is door gebruik van tags & labeling, kosten per afnemende afdeling inzichtelijk te maken.
6. Governance, Security & Compliance: Zeker met de GDPR-wetgeving, maar ook met de reeds bestaande governance, is waarborgen dat het data platform blijft voldoen aan de door PNB gestelde richtlijnen en eisen een zeer belangrijke basisservice. Hier wordt o.a. gebruik gemaakt van de door Microsoft gegarandeerde compliance op het trustcenter: <https://www.microsoft.com/en-us/trustcenter/default.aspx>.
7. Subscription management: De subscription is de (administratieve) beveiligingsgrens binnen Azure, welke voor het provinciale dataplatform is aangemaakt en waarin de resource groups komen te staan. De subscription en de resources worden opgezet volgens het Infrastructure as Code (afgekort: IaC) principe. De voordelen zijn: consistentie, snelheid, herbruikbaarheid en uitbreidbaarheid.
8. SLA & Monitoring: Het dataplatform maakt gebruik van de PaaS diensten van Azure. Microsoft en garandeert een minimale beschikbaarheid van 99,9%. Daarnaast worden middels Log Analytics standaard loggegevens gemonitord en zijn proactieve alerts ingesteld.
9. DevOps: Voor de provincie best wennen, maar OTAP is voor het data platform losgelaten. Wanneer er in de Cloud gewerkt wordt, is het een onnodig dure kostenpost om structureel vier omgevingen in de lucht te hebben. In het data platform kan met één druk op de knop een nieuwe omgeving opgesteld worden, waarbij alleen de nodige onderdelen opgehaald worden om wijzigingen door te voeren. De wijzigingen worden na de acceptatie teruggezet naar de productieomgeving, en worden daarmee live gebracht.

Alleen een technische infrastructuur is niet voldoende om een dataplatform te laten slagen. Er is veel meer nodig. Er moet een visie en een strategie zijn vanuit de organisatie om te garanderen dat iedereen weet dat dit de manier is om data te verzamelen, gebruiken en delen. Invulling van de visie moet afgestemd worden met alle geldende uitgangspunten, kaders en richtlijnen van de organisatie. Daarnaast zijn vele processen nodig om de infrastructuur te vullen en gebruiken. Denk hierbij aan; hoe bepalen we welke bronnen ontsloten worden, met welke prioriteit en hoe komen de beleidsvraagstukken vanuit de organisatie terecht bij het dataplatform en zorgen we dat er geen andere paden bewandeld worden? En wanneer er teveel vragen om rapporten zijn, wie bepaald dan de prioriteit? Het is noodzakelijk om hier een governance voor in te richten.

4. Proces, uitdagingen en ontwikkelingen

In dit hoofdstuk gaan we in op het doorlopen proces en de uitdagingen die we tijdens de realisatie zijn tegengekomen en soms nog steeds mee te maken hebben.

4.1 Proces; van klein naar groot

Binnen het mobiliteit werkveld was de behoefte om data en informatie beter te organiseren erg groot. De angst bestond dat grote projecten en programma's om dit binnen de gehele provincie op te pakken te groot zouden worden en vertragend zouden werken.

In de zomer van 2018 is dan ook besloten om klein te beginnen met het opzetten van het dataplatform als pilot en een enkele databron te gaan ontsluiten voor mobiliteit.

De insteek is weliswaar wel geweest om van pilot naar een structurele oplossing te gaan die toekomstbestendig en uit te breiden zou zijn.

Het idee was door klein te beginnen en op korte termijn de eerste resultaten te kunnen tonen de rest van de organisatie te enthousiasmeren en aan te laten haken.

Met een vast ontwikkelteam, van in eerste instantie 5 personen, is begonnen met de ontwikkeling van het dataplatform.

Vanaf de start zijn belangrijke stakeholders uit de organisatie betrokken en zijn keuzes bijvoorbeeld rondom architectuur besproken en wel afgewogen genomen. Door middel van kort cyclische ontwikkeling en een intensieve samenwerking tussen de klant, de vragende afdeling mobiliteit, en de leverancier zijn al snel concrete stappen gemaakt.

De scope van het dataplatform was in eerste instantie intern binnen de provincie om goed onderbouwde beleidskeuzes te kunnen maken door inzicht te geven. Wel met het oog op de toekomst om data en informatie ook extern te kunnen delen met regionale overheden, burgers en marktpartijen. Ook hier was het idee “van klein naar groot”.

Door het ontsluiten van één enkele databron en het opwerken van de data tot informatie kon al vrij snel een eerste dashboard worden gerealiseerd.

Dit dashboard werd gebruikt ter inspiratie in vele interne, maar ook externe overleggen. Het gewenste effect volgde inderdaad al snel en vanaf begin 2019 kwamen vragen vanuit collega's binnen mobiliteit om meer databronnen te ontsluiten en dashboards voor visualisatie van informatie te realiseren. Andere afdelingen binnen de provincie, zoals energie en ruimtelijke ontwikkeling wilde ook aanhaken.

Het idee om meerdere databronnen te ontsluiten en informatie met elkaar te combineren begon realiteit te worden. Zo ging de status van het project van pilot naar structurele toekomstige oplossing voor het organiseren van data en informatie binnen de provincie. Het gewenste vliegwiel was bereikt!

4.2 Uitdagingen

Zoals in elk project zijn we ook hier diverse uitdagingen tegen gekomen, welke we in deze paragraaf toelichten.

1. Alle data extern

Vrijwel alle te ontsluiten databronnen zijn extern belegd. Soms bij marktpartijen, maar ook bij koepelorganisaties of samenwerkingsverbanden. Op zich niet vreemd, echter de wens was nu om data met elkaar te gaan combineren en samen te laten komen in het dataplatform. Het bleek in de meeste gevallen toch complexer dan gedacht om de data te ontsluiten. Om

technische redenen, zoals het ontbreken van goede standaarden aan de kant van de leverancier en API's, maar ook door afgesloten contracten waarin het terug leveren van data niet voorzien was. Dit laatste bracht soms extra kosten mee of geringe motivatie van partijen om hieraan mee te werken.

2. Datakwaliteit en wijzigingen

De kwaliteit van de data blijkt vaak matig te zijn of slecht te gebruiken voor het combineren van meerdere databronnen. Dit kan te maken hebben met slechte inwinning van de data of dat data met een oorspronkelijk ander doel is ingewonnen. Een ander probleem is dat structuur en kwaliteit regelmatig bleek te wijzigen bij volgende data leveringen. Om de data goed inzichtelijk te maken zijn vervolgens herstelacties nodig wat vaak een tijdrovende klus is. Toch levert dit wel veel inzicht. Voorheen werd gedacht dat er veel data was en dat alleen de techniek ontbrak om de data te gebruiken. Door de data in het dataplatform op te slaan en vervolgens op basis van een informatievraag de data te prepareren voor gebruik, werd duidelijk dat er geen standaarden waren voor de data. De ene bron gebruikt wegnummers met hectometers en de volgende weer lat/long coördinaten. Dit inzichtelijk maken leidt tot het kunnen stellen van de juiste vragen om datakwaliteit te kunnen verbeteren. Dit was en is wel een traag proces, dat bewijst dat dataopslag en toegang alleen niet voldoende is voor inzicht en analyses.

3. Ontbreken van documentatie of goede definities

Het ontbreken van goede definities of documentatie maakte de verwerking van data tot informatie lastig. Veel kennis en definities zitten in hoofden van collega's of bij externe marktpartijen. We zien dat er matig is gedocumenteerd en het is een tijdrovende klus om duidelijkheid te krijgen over definities en betekenis van de data en deze vervolgens alsnog te documenteren.

4. Afstemming stakeholders

Omdat het dataplatform klein is opgestart was er minimale afstemming nodig in de organisatie. Nu het dataplatform groter wordt en er meer stakeholders bijkomen is de besluitvorming complexer geworden. Er zijn meer partijen waar keuzes mee moeten worden afgestemd. Volle agenda's dragen hier niet in bij en werkt vertragend.

5. Minimale standaardisatie en uniformering

Binnen het werkveld mobiliteit is er nog weinig sprake van standaardisatie en uniformering. Zo is er bijvoorbeeld geen gestandaardiseerd geografisch netwerk dat als basis voor alle kaartdata wordt gebruikt. Er wordt gebruik gemaakt van verschillende netwerken, zoals Open Street Map, Nationaal Wegenbestand of andere netwerken. Er lopen landelijk trajecten voor het ontwikkelen van een basiskaart of basisregistratie, echter is dit nog niet op korte termijn gerealiseerd en toepasbaar. Dit maakt het combineren van diverse geografische data ingewikkeld. Een ander voorbeeld is dat contracten en business rules niet geuniformeerd zijn, bijvoorbeeld door wegbeheerders voor het inwinnen en verwerken van telgegevens. Het combineren van data en informatie over wegbeheerders heen is dan appels met peren vergelijken.

6. Privacy

Het niveau waarop data soms verkrijgbaar is leidt tot privacyvraagstukken. Zo mag er vooralsnog geen ov-chipcard data worden gedeeld op bronniveau. Dit dient eerst door vervoerders geaggregeerd en geanonimiseerd te worden.

Het combineren van databronnen kan leiden tot privacygevoelige informatie, waardoor steeds de afweging moet worden gemaakt welke bronnen kunnen worden gecombineerd, voor welk doel en welke maatregelen moeten we nemen om de privacy te waarborgen. Dit proces leidt met enige regelmaat tot interne en externe discussies.

4.3 Doorkijk naar de toekomst

Op dit moment wordt gekeken hoe we het dataplatform extern beschikbaar kunnen maken om inzichten en informatie niet alleen intern binnen de provincie, maar ook te kunnen delen met externe partijen, zoals regionale overheden, burgers en marktpartijen.

Op landelijk niveau is het gesprek op gang gekomen voor meer standaardisatie en uniformering, denk aan gezamenlijke inwinning van bepaalde data of een basis geografische netwerkkaart.

Interne doorontwikkeling van het dataplatform door het combineren van data over beleidsterreinen heen en het ontsluiten van meer databronnen, zodat een integraal inzicht kan worden gegeven aan management en bestuur.

5. Aanbevelingen

De ervaringen die wij als provincie Noord-Brabant hebben opgedaan wilde wij graag met jullie delen, zodat anderen met soortgelijke ontwikkelingen hun voordeel er mee kunnen doen. Vanuit onze ervaringen hebben we ook een aantal aanbevelingen.

- Begin klein en niet te groot.
- Denk wel goed na over basale (architectuur) keuzes en zorg dat de zaken die je doet ook toekomstbestendig zijn.
- Zorg voor een vast en volledig inzetbaar team medewerkers dat zich volop kan bezighouden met de ontwikkeling.
- Denk bij nieuwe contracten voor data-inwinning goed na over afspraken rondom de terug levering van data. Wees hier niet te vrijblijvend in en zorg dat zaken als formaat, frequentie, kosten e.d. duidelijk zijn.
- Hou rekening met uitloop in tijd (en geld) door vertragende factoren, zoals eerder in deze paper beschreven.
- Zet niet alle ontwikkelingen in een organisatie stop vanwege een te ontwikkelen dataplatform. Het is immers niet van de een op andere dag gerealiseerd en wees reëel in verwachtingen wat je kan leveren op welke termijn.
- Denk goed na over connecten of collecten; sluit zoveel als mogelijk aan op bestaande dataplatforms of dataknooppunten middels API's. Alleen collecten en opslaan waar nodig. Zorg wel dat je flexibel blijft, soms kan het wel handig zijn om bepaalde data zelf (tijdelijk) op te slaan, bijvoorbeeld voor een goede performance van dashboards.
- Zorg voor goede meta data, wees duidelijk waar men naar kijkt in dashboards, van wanneer is de data? wat is de status ervan? wat is de bron?
- Maak goede afspraken met leverende partijen, waar mogelijk stel een GLO (gegevens leveringsovereenkomst) op. Dit geeft wat meer houvast op de structuur en kwaliteit van de aan te leveren data.

- Sluit zoveel als mogelijk aan bij bestaande dataknooppunten of koepelorganisaties. Vaak hebben zij al data verzameld op landelijk niveau en de nodige databewerkingen gedaan, waardoor al een bepaald informatieniveau beschikbaar is.
- Met alleen het opzetten van een technische oplossing ben je er niet, een heel groot deel draait ook om governance en procesafspraken!

6. Bibliografie

ALTEN Nederland; Provincie Noord-Brabant, 2019. *Architectuur Dataplatform Provincie Noord-Brabant*, 's-Hertogenbosch: Provincie Noord-Brabant.

Derksen, R. & de Jong, K., 2019. *Implementatieplan digitalisering Zuid-Nederland*, 's-Hertogenbosch: Provincie Noord-Brabant.

Madsen, M., 2015. *A third nature whitepaper How to build an Enterprise datalake*, sl: sn