

Meer zicht op regionaal reisgedrag: combinatie van OV-chipkaartgegevens bij de Vervoerregio Amsterdam

Machiel Kouwenberg – Vervoerregio Amsterdam – m.kouwenberg@vervoerregio.nl

Suzanne Kieft – Vervoerregio Amsterdam – s.kieft@vervoerregio.nl

Natalie in 't Veld – Piece of Data – natalie.intveld@pieceofdata.nl

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

Samenvatting

Al vanaf de eerste verkennende studies naar de Noord/Zuidlijn, diep in de jaren negentig, keken geïnteresseerde en bezorgde burgers kritisch mee naar de netwerkscenario's die de gemeente Amsterdam en de Vervoerregio onderzochten. Steeds was het belangrijk om alle afwegingen en keuzes goed uit te leggen. Immers: als je wilt voorkomen dat burgers boos worden, moet je als overheid hierover helder zijn, ook als de uiteindelijke uitkomsten niet iedereen tevreden zullen stellen. Het gebruik van objectieve, gemeten gegevens helpt bij het uitleggen en verklaren van de verschillende opties en de gevolgen daarvan.

De Noord/Zuidlijn is de nieuwe ruggengraat van het OV-netwerk in Amsterdam, maar heeft ook grote invloed op het netwerk in de concessiegebieden Zaanstreek-Waterland en Amstelland-Meerlanden. Een van de uitgangspunten was: meer aanbod tegen dezelfde kosten, dus de Vervoerregio heeft gestuurd op het niet dubbel aanbieden van dezelfde verbindingen. Vanaf de noordkant moest (deels) op de Noord/Zuidlijn worden aangetakt, en ook vanaf de zuidkant rijden de bussen niet meer parallel naar station Amsterdam Centraal. Dit heeft veel invloed op reizigers: meer of minder overstappen, meer of minder reistijd. Uiteraard bevat de OV-chipkaart veel informatie over concessiegrens-overschrijdende reisbewegingen. Tot voor kort konden we echter geen gebruik maken van deze bron. De afgelopen jaren is door decentrale overheden, vervoerders en het ministerie van Infrastructuur en Waterstaat samengewerkt aan de Informatiehuishouding. Deze maakt het mogelijk om onder voorwaarden en met een goede omschrijving van doelen inzicht te krijgen in het totale verplaatsingsgedrag met het OV in een bepaald gebied.

De Vervoerregio Amsterdam heeft als een van de eerste partijen het hele proces van informatieverzoek tot levering doorlopen. De geleverde informatie biedt inzicht in hele OV-ketens, dus inclusief de overstap tussen concessies en de relatie met het NS-net. Wij nemen u graag mee in onze ervaringen.

Machiel is beleidsadviseur bij de Vervoerregio Amsterdam, in het team Kennis en Onderzoek, onder leiding van Suzanne Kieft. Machiel en Suzanne hebben Natalie in 't Veld, data scientist bij Piece of Data, ingehuurd voor de dataverwerking en -analyses.

1. Inleiding

Op 22 juli 2018 ging na vele jaren onderzoek en aanleg de Noord/Zuidlijn rijden. Tegelijk namen de Vervoerregio Amsterdam – als opdrachtgever voor het regionale openbaar vervoer - en de vervoerders Connexxion, EBS en GVB ook het omliggende lijnennet flink op de schop. Dit met als doel om vervoerkundige samenhang met de nieuwe metrolijn te creëren, de reiziger goed te bedienen en het beschikbare exploitatiebudget efficiënt in te zetten.

Al vanaf de eerste verkennende studies naar de Noord/Zuidlijn, diep in de jaren negentig, keken geïnteresseerde en bezorgde burgers kritisch mee naar de netwerkscenario's die de gemeente Amsterdam en de Vervoerregio¹ onderzochten. Steeds was het belangrijk om alle afwegingen en keuzes goed uit te leggen. Immers: als je wilt voorkomen dat burgers boos worden, moet je als overheid hierover helder zijn, ook als de uiteindelijke uitkomsten niet iedereen tevreden zullen stellen. De Vervoerregio, de gemeente en de vervoerders hebben in de aanloop naar 22 juli 2018 naast onderzoek dan ook veel aandacht besteed aan communicatie over de veranderingen in het netwerk en de gedachten daarachter. Dit gebeurde vooral via de website www.wijnemenjemee.nl

Er stond voor de OV-reiziger in de Amsterdamse regio heel wat op het spel. De Noord/Zuidlijn kwam parallel te lopen aan bestaande bus- en tramlijnen die al decennia in het collectief geheugen van de reiziger zaten, maar biedt een snellere verbinding met meer capaciteit. Dus het was vervoerkundig enerzijds logisch om goed te kijken naar de onderlinge rolverdeling en aansluiting tussen bus, tram en metro op die trajecten, en ook of de metro een deel van de bus- en tramlijnen kon vervangen. Anderzijds bood het bestaande bus- en tramnetwerk directe verbindingen die veel mensen wilden behouden. Maar de Noord/Zuidlijn zou ook voor nieuwe directe en snelle reisroutes zorgen. Het was kortom zaak om de juiste mix te vinden tussen bestaande en nieuwe mogelijkheden. Met als financiële randvoorwaarde een fors teruglopende exploitatiebijdrage van het Rijk, en daarmee ook de noodzaak om de kostendekkingsgraad van het netwerk als geheel te verhogen.

1.1 Behoeftte aan meer informatie over reisgedrag

Het hele proces van de eerste verkenningen tot aan de definitieve netwerkkeuzes duurde circa 20 jaar. Telkens hadden de gemeente, de Vervoerregio en de vervoerders een zo scherp mogelijk beeld nodig van de huidige en toekomstige vervoervraag van reizigers in Amsterdam en de regio. Dat was niet eenvoudig. Tot de geleidelijke overschakeling op de OV-chipkaart in de periode 2006 - 2011 reisden we allemaal nog met papieren abonnementen en strippenkaarten. Vooral de laatste gaven nauwelijks inzicht in het reisgedrag van H(erkomst) naar B(estemming), laat staan in tussentijdse overstappen. Wel was er aanvullend onderzoek via enquêtes die reizigers bij de aankoop van hun

¹ In de jaren negentig verkende gemeente Amsterdam als initiatiefnemer voor de Noord/Zuidlijn samen met de vervoerders de mogelijke toekomstige lijnvoering. Rond 2000 werd de Vervoerregio (eerst nog ROA en daarna Stadsregio) opdrachtgever voor het regionale OV en was vanaf toen verantwoordelijk voor de kaders van het toekomstige netwerk. Voor de leesbaarheid van het paper hebben we overal de term 'Vervoerregio' gebruikt, ongeacht de periode.

strippenkaart kregen uitgereikt². Daarnaast boden gemeentelijke en landelijke reisonderzoeken zoals OViN, de Amsterdamse variant OViA en hun voorgangers nauwkeuriger informatie over de H/B-relaties in de regio. Deze informatie vormde samen met tellingen een belangrijke basis voor de verkeersmodellen die zijn gebruikt bij de netwerkstudies. En tenslotte wisten de vervoerders natuurlijk ook uit hun eigen tellingen en ervaringen 'in het veld' hoe reizigers binnen hun eigen netwerk onderweg waren en wat de grote overstaprelaties waren.

Toch bleef er behoefte aan een meer gedetailleerd en samenhangend totaalbeeld van reispatronen in de regio, op basis van grote aantallen waarnemingen. Onze verwachting was dan ook dat dit totaalbeeld spoedig beschikbaar zou komen na de volledige omschakeling op de OV-chipkaart in 2011, vanwege de grote potentiële meerwaarde voor het onderbouwen van OV-beleid en investeringen in infrastructuur. Het duurde echter langer dan gedacht, onder andere vanwege vraagstukken rond privacy, mededinging en bedrijfsvertrouwelijkheid (zie verder in paragraaf 4.2).

In 2015 zette de Vervoerregio samen met Translink een eerste stap om zelf beter gebruik te maken van de OV-chipkaartdata, door binnen de afzonderlijke concessiegebieden losse ritten te laten koppelen tot complete reizen van herkomst naar bestemming (bijvoorbeeld van A naar B in Amsterdam, maar met een overstap op C). Dit vergde destijds enige afstemming met de vervoerders, onder andere over het eigendom van de data, maar leidde wel tot een werkbaar compromis: per concessiegebied kregen we de beschikking over een dataset met tot reizen gekoppelde ritten. Deze informatie is gebruikt voor de definitieve netwerkkeuzes in de [Lijnennetvisie](#) van de Vervoerregio uit 2015 en de daarop gebaseerde vervoerplannen van Connexion, EBS en GVB uit 2016. Maar de overstappen tussen de concessiegebieden en ook tussen bus/tram/metro en het NS-net bleven tot nu toe buiten beeld. Wel heeft de Vervoerregio in het kader van de Lijnennetvisie een inschatting hiervan laten maken door middel van een statistische toetsing.

1.2 Nieuwe mogelijkheden door samenwerking

De laatste jaren hebben decentrale OV-autoriteiten, vervoerders en het ministerie van Infrastructuur en Waterstaat zich via hun samenwerking in het Nationaal Openbaar Vervoer Beraad (NOVB; www.nationaalovberaad.nl) ingespannen voor een oplossing om ondanks de genoemde vraagstukken toch meer inzicht te krijgen in het totale verplaatsingsgedrag in de regio's. Dit met als doel om deze kennis te kunnen toepassen voor verbeteringen voor de reiziger. Ook Trans Link Systems (Translink) is als beheerder van de OV-chipkaartdata nauw hierbij betrokken. In 2017 leidde dit proces tot de Intentieverklaring open reisdata, die overheden en vervoerders hebben ondertekend. Geleidelijk is er vervolgens onder regie van het Samenwerkingsverband Decentrale OV-autoriteiten (SV DOVA) en Translink schot in de zaak gekomen. Een grote stap was de introductie van de nieuwe Informatiehuishouding in 2018 ([Kompas Informatiehuishouding](#)). Deze maakt het mogelijk om onder voorwaarden en met een goede omschrijving van doelen inzicht te krijgen in het totale verplaatsingsgedrag met

² Deze enquêtes werden uitgevoerd door de Werkgroep Reizigers Omvang en Omvang Verkopen (WROOV)

het OV in een bepaald gebied. In hoofdstuk 4 gaan we verder in op de Informatiehuishouding.

De Vervoerregio heeft in het voorjaar van 2018 via de Informatiehuishouding een verzoek ingediend en in juni 2019 de gevraagde informatie over reisgedrag in de regio ontvangen. De directe aanleiding hiervoor is de evaluatie van de lijnennetvisie. In hoofdstuk 3 komen de doelen en beoogde toepassingen van het verzoek verder aan de orde. Het informatieverzoek van de Vervoerregio is een van de eerste dat het hele proces – van vraagstelling tot uitlevering van de informatiebestanden – succesvol heeft doorlopen. Wij nemen u graag mee in onze ervaringen. Daarbij presenteren we wel gelijk een 'spoiler': we laten in dit artikel inhoudelijk niets los over onze inzichten in het reisgedrag. Dit heeft te maken met de leveringsvoorwaarden van ons informatieverzoek: de Vervoerregio mag de informatie alleen zelf gebruiken en dus niet zo maar delen, ook niet met de betrokken vervoerders. Wanneer een andere partij dergelijke informatie nodig heeft, dan moet zij een eigen informatieverzoek indienen.

2. De Vervoerregio Amsterdam

Voordat we verder ingaan op het proces en de inhoud rond het informatieverzoek, introduceren we eerst kort de Vervoerregio Amsterdam. Dit is een regionaal samenwerkingsverband van vijftien gemeenten, met de gemeente Amsterdam als grootste gemeente en geografisch middelpunt. Tot de Vervoerregio behoort een aantal grotere, stedelijke gemeenten zoals Amstelveen, Haarlemmermeer, Purmerend en Zaanstad, maar ook een aantal wat meer landelijke gemeenten. In totaal wonen er circa 1.500.000 mensen in de Vervoerregio en er zijn circa 1.000.000 arbeidsplaatsen (gemeten in werkzame personen). Grote vervoerkundige dynamiek zien we bij de belangrijke treinstations zoals Amsterdam Centraal en Zuid en bij de diverse concentraties van werkgelegenheid, onderwijs, voorzieningen en toeristische attracties in en direct rond Amsterdam. Daarnaast natuurlijk bij Schiphol, maar bijvoorbeeld ook bij de bloemenveiling Flora Holland in Aalsmeer en de Zaanse Schans in Zaanwijk.

Bestuurlijk is de Vervoerregio een van de twee gemeentelijke samenwerkingsverbanden met wettelijke taken op het gebied van verkeer en vervoer, samen met de Metropoolregio Rotterdam Den Haag (MRDH). De Vervoerregio ontvangt van het Rijk rechtstreeks een jaarlijks budget voor de exploitatie van haar OV-concessies en voor investeringen in het regionale mobiliteitssysteem zoals fietspaden, vrije OV-banen, betere benutting van weginfrastructuur, vergroting van verkeersveiligheid en verduurzaming van voertuigen. De politieke besluitvorming vindt plaats in een dagelijks bestuur van drie wethouders uit de gemeenten en een regioraad met 50 raadsleden en bestuurders uit de vijftien gemeenten.

Uitgangspunt voor het beleid en de investeringen van de Vervoerregio is het [Beleidskader Mobiliteit](#). Doelstellingen hierin zijn onder andere:

- Acceptabele en betrouwbare reistijden in de regio.
- Voldoende capaciteit en een logische opbouw van het OV-netwerk zodat we de grote vervoerstromen goed kunnen bedienen.

- Beleving van de reiziger: gemakkelijk en comfortabel reizen van A naar B met de snelste en handigste combinaties tussen fiets, OV, auto en lopen.
- Knooppunten waar je gemakkelijk en veilig je weg kunt vinden, o.a. via verbetering van looproutes.

Voortbordurend op het Beleidskader is in 2019 de nieuwe Investeringsagenda Mobiliteit opgesteld die het beleid vertaalt in activiteiten en investeringen.

3. Doelen en toepassingen van de informatie uit ons verzoek

De opgevraagde informatie willen we gebruiken voor verschillende doelen en toepassingen die passen bij de beleidsdoelen uit het bovenstaande hoofdstuk 2:

- Algemeen doel: samen met de vervoerders het regionale OV-netwerk in de toekomst verder verbeteren door aan te sluiten bij de vervoervraag van reizigers en door het teruglopende exploitatiebudget effectiever in te zetten, waarbij we negatieve uitwerkingen voor de reizigers zo veel mogelijk beperken.
- Evalueren van de Lijnennetvisie uit 2015 en de netwerkwijzigingen die bij de ingebruikname van de Noord/Zuidlijn zijn toegepast.
- Toetsen van vervoerkundige uitgangspunten bij de aanbesteding van de concessie Zaanstreek-Waterland, een 'fusie' van twee afzonderlijke concessies.
- Onderbouwen van investeringen in de verbetering van knooppunten, overstappunten en trajecten.

Hieronder gaan we nog wat uitgebreider in op twee doelen en toepassingen.

3.1 Evalueren van de Lijnennetvisie uit 2015

Vanaf de eerste onderzoeken naar de aanleg van de Noord/Zuidlijn in de jaren negentig is er altijd gewerkt met scenario's voor de uiteindelijke lijnvoering en frequenties van het aansluitende bus- en tramnetwerk. In deze lijnvoeringsscenario's stonden vanaf het begin twee belangrijke afwegingen centraal:

- De verhouding tussen het laten doorrijden van bussen uit Amsterdam Noord, Waterland/Zaanstreek en Amstelland-Meerlanden naar Amsterdam Centraal/busstation Elandsgracht versus het 'aantakken' van bussen op de Noord/Zuidlijn op de metrostations Noord, Noorderpark of Zuid.
- De inrichting van het tramnetwerk van Amsterdam

Tegelijk bezuinigde het Rijk vanaf 2008 circa 180 miljoen euro op het exploitatiebudget voor de vier OV-concessies van de Vervoerregio, waardoor de noodzaak voor een efficiënte lijnvoering groter werd en de 'bandbreedte' voor de bovenstaande afwegingen afnam: we moesten kritischer kijken naar paralleliteit met de Noord/Zuidlijn. Ook werd het vanuit gemeentelijk beleid voor luchtkwaliteit, verkeersveiligheid en inrichting van de openbare ruimte minder vanzelfsprekend dat er grote aantallen bussen door het centrum van Amsterdam bleven rijden.

De laatste jaren tot aan de opening zijn de scenario's uitgekristalliseerd tot de Lijnennetvisie die de Vervoerregio in 2015 bestuurlijk vaststelde. Het centrale doel in de Lijnennetvisie is om een kwaliteitssprong in het netwerk te realiseren: met sneller en

beter OV meer reizigers vervoeren, meer inkomsten genereren, en daardoor het OV in de toekomst betaalbaar houden, ondanks een teruglopend exploitatiebudget. De Vervoerregio heeft er in de lijnennetvisie voor gekozen om met de ingebruikname van de Noord/Zuidlijn het metronet nog meer dan voorheen te benutten als de ruggengraat van het OV-systeem, waarbij aansluitende tram- en buslijnen de stad en de regio verder bedienen. Deze netwerkopbouw versterkt de aantrekkelijke en 'zware' lijnen, en verbetert de kostendekkingsgraad van het OV-netwerk. Hierdoor wordt ook de sociale functie van het openbaar vervoer in stand gehouden.

De Lijnennetvisie vormt een kader met doelstellingen en randvoorwaarden voor het netwerk. Vervolgens hebben de vervoerders Connexxion, EBS en GVB binnen dit kader hun vervoerplannen met de precieze lijnvoering, frequenties en bedieningsperioden in detail uitgewerkt.

Het is inmiddels ruim een jaar na de ingebruikname van de Noord/Zuidlijn op 22 juli 2018. De lijn functioneerde vanaf de eerste dag goed, de reizigersaantallen namen snel toe en groeien nog steeds. In juni 2019 vervoerde GVB gemiddeld 105.000 reizigers per werkdag en op Koningsdag 2019 zelfs 130.000 (www.gvb.nl). We zien dit als een succes, maar we willen ook graag evalueren of het netwerk in de Vervoerregio voldoet aan de doelstellingen die we in 2015 in de Lijnennetvisie hebben vastgelegd. Wat is het effect van de gemaakte keuzes voor de reiziger? Hoeveel reizigers moeten bijvoorbeeld meer of minder vaak overstappen, en hoe heeft de reistijd zich ontwikkeld? Ook kijken we naar de heroriëntatie van NS-reizigers tussen de stations Amsterdam Centraal en Zuid. Als Vervoerregio leggen wij bij de evaluatie de nadruk op de werking van het regionale netwerk als geheel. De evaluatie van specifieke lijnvoeringskwesaties ligt met name op het terrein van de vervoerders, dit past bij de rolverdeling tussen de Vervoerregio en de vervoerders. Overigens is wel afgesproken dat het netwerk na 22 juli 2018 de eerste tijd niet sterk zou wijzigen zodat reizigers de gelegenheid kregen om er goed aan te wennen en de voordelen te gaan benutten.

3.2 Aanbesteding concessie Zaanstreek-Waterland

Een half jaar na de ingebruikname van de Noord/Zuidlijn begon de Vervoerregio met de voorbereidingen voor de aanbesteding van de nieuwe concessie Zaanstreek-Waterland, die in december 2021 van start gaat. Dit is een samenvoeging van de huidige concessiegebieden Zaanstreek en Waterland. Deze gebieden aan de noordkant van de regio hebben een sterke relatie met de Noord/Zuidlijn, en per 22 juli 2018 rijden verschillende lijnen uit met name Waterland tot aan metrostation Noord. Voor het toetsen van het vervoerkundig programma van eisen voor de nieuwe concessie was het nodig dat er op tijd (zomer 2019) actuele informatie beschikbaar was over het reisgedrag tussen Zaanstreek-Waterland en diverse gebieden in en rond Amsterdam.

3.3 Andere vraagstukken en soorten informatie

We verwachten in de komende jaren zeker meer vraagstukken waarvoor we informatie nodig hebben over reisgedrag tussen concessiegebieden en over de combinatie bus/tram/metro-trein. Het kan bijvoorbeeld gaan om infrastructurele projecten zoals de herinrichting van een OV-knooppunt waarbij we zo goed mogelijk willen anticiperen op de

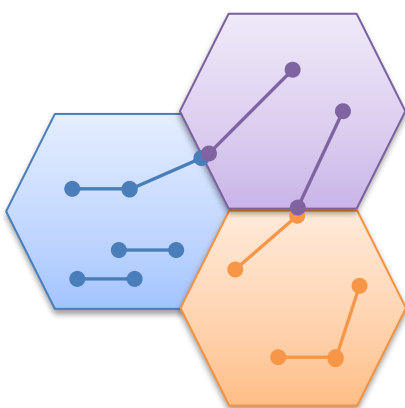
overstapbewegingen en loopstromen van reizigers. Daarnaast zijn er (inter)regionale netwerk- en corridorstudies met NS en de andere vervoerders, zoals voor het OV-Toekomstbeeld 2040. Ook bij verbetering van verkeersmodellen en bij de ontwikkeling van nieuwe mobiliteitsconcepten als Mobility as a Service (MaaS) is er behoefte aan adequate inzichten in het daadwerkelijke reisgedrag van mensen, onder andere om situaties voor en na de invoering van een dergelijk concept met elkaar te kunnen vergelijken. Het is in zulke gevallen logisch als de betrokken partijen voor dergelijke maatwerkvragen nieuwe informatieverzoeken indienen, zodat de gevraagde informatie actueel en is toegesneden op het doel van de vraag.

OV-chipkaartdata zijn een hele belangrijke, maar zeker niet de enige bron van informatie die een beter inzicht kunnen bieden in reisgedrag van mensen en het functioneren van mobiliteitsnetwerken. De laatste jaren komen er – als open data of tegen betaling - steeds meer en nauwkeuriger informatiebronnen over reisgedrag beschikbaar, zoals de verplaatsingsdata van de Nationale Databank Wegverkeersgegevens (NDW), de voertuigdata van Nationale Data Openbaar Vervoer (NDOV), data van mobiele telefoons en van smartphoneapplicaties die reisbewegingen op basis van GPS en vrijwillige toestemming inzichtelijk maken. En er zijn diverse longitudinale mobiliteitsonderzoeken zoals OnDerweg in Nederland (ODiN, vroeger OViN) en het Mobiliteitspanel Nederland (MPN). Het is de kunst hoe we met al deze data op een verantwoorde manier (privacy, transparantie, economische belangen) inzichten verwerven die ons ook echt verder helpen om het reizen aangenamer, efficiënter en duurzamer te maken.

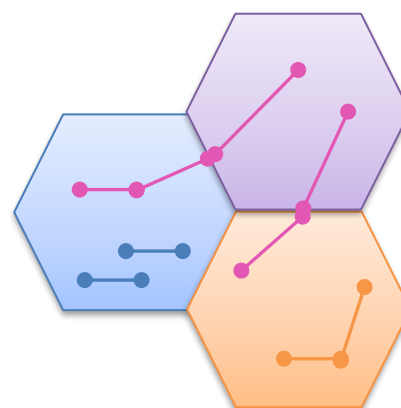
4. Informatiehuishouding bij Translink

4.1 Achtergrond

Het overgrote deel van de reizigers met het openbaar vervoer in Nederland reist met de OV-chipkaart. Ze checken in en uit, en op basis van de afgelegde afstand wordt de ritprijs bepaald en afgerekend. Bijkomend voordeel van dit systeem is dat er ook vervoerstromen in beeld kunnen worden gebracht. Tot voor kort was het alleen mogelijk om inzicht te krijgen in deze stromen per vervoerder. De overstappen tussen vervoerders/concessies bleven buiten beeld. Binnen de Informatiehuishouding is het mogelijk om wel de hele (OV-reis)keten samen te stellen:



Figuur 1: Zonder Informatiehuishouding: ketens per concessie



Figuur 2: Informatiehuishouding: ketens over de concessiegrens heen

4.2 Inrichting van de Informatiehuishouding

In een langdurig traject waarbij zowel overheden als vervoerders betrokken waren, is het mogelijk gemaakt om de chipkaartgegevens óver de vervoerders heen aan elkaar te koppelen. Aangezien Translink alle chipkaarttransacties ontvangt en verwerkt, is de Informatiehuishouding bij Translink ingericht. Bij het maken van keuzes bij het inrichten van de Informatiehuishouding speelden vier elementen een grote rol: privacy, mededinging, bedrijfsvertrouwelijkheid en interpretatie. Deze vier elementen komen ook terug bij het formuleren en toetsen van een informatieverzoek.

Privacy

Chipkaarttransacties zijn persoonsgegevens: uit reispatronen van individuele reizigers is veel informatie over die reiziger af te leiden. Er moet dus rekening worden gehouden met de Algemene Verordening Gegevensverwerking (AVG). Belangrijk is dat het doel van de verwerking strookt met het doel waarvoor de gegevens verkregen zijn, het moet mogelijk zijn om aan reizigers uit te leggen dat zijn gegevens worden gebruikt voor het informatieverzoek. Elke vervoerder heeft in zijn privacy statement staan waarvoor de chipkaartgegevens worden gebruikt. Toch is het niet zomaar logisch om die gegevens ook vervoerder-overstijgend te verwerken. Om reispatronen in beeld te brengen, moeten de gegevens op transactieniveau worden gecombineerd. Dit is op persoonsniveau (met wat moeite) ook zonder OV-chipkaartnummer tot individuele reizigers herleidbaar. Het volgen van individuele reizigers is bij beleidsvragen nooit het doel, geaggregeerde gegevens voldoen prima. Bij Translink zijn verschillende waarborgen ingebouwd om directe en indirecte herleidbaarheid te voorkomen. In de verwerking worden de gegevens eerst (extern) gepseudonimiseerd. Hierdoor is directe koppeling met het OV-chipkaartnummer onmogelijk. Vervolgens worden de gegevens geaggregeerd tot het minimaal benodigde niveau om nog conclusies voor het beleidsdoel te kunnen trekken, en de indirecte herleidbaarheid wordt voorkomen. Denk hierbij aan het gebruik van klassen of het gemiddelde in plaats van absolute waarden bij kleine aantallen.

Bedrijfsvertrouwelijkheid en mededinging

Een bepaalde gedetailleerde combinatie van gegevens kan heel veel zeggen over bedrijfsvoering, operatie en strategie. Uit het oogpunt van mededinging en/of bedrijfsvertrouwelijkheid mogen dat soort gegevens niet met alle partijen worden gedeeld of openbaar gemaakt. De autoriteit Consument en Markt houdt dit aspect nauwlettend in de gaten.

Interpretatie en juist gebruik

Het is belangrijk dat de ontvangende partij met verstand van zaken naar de verkregen informatie kijkt. Al lijken de cijfers soms voor zich te spreken, achtergrondkennis van openbaar vervoer en de werking van het systeem is wel nodig. Ook de beperkingen van de informatie moeten helder zijn; als voorbeeld: Translink verwerkt alle transacties van de plastic OV-chipkaarten. Bepaalde doelgroepen en betaalwijzen blijven hierdoor buiten beeld. Dit hoeft geen probleem te zijn, maar moet wel worden onderkend.

4.3 Een informatieverzoek

Een informatieverzoek kan bij Translink worden ingediend. Doelbinding voor de verwerking van de persoonsgegevens is hierbij heel belangrijk. Dit betekent dat het doel van de verwerking in lijn moet zijn met het doel waarvoor de gegevens in eerste instantie zijn verzameld. Verder moet het verzoek heel concreet zijn. Het verzoek kan in het Voorportaal voor het eerst beoordeeld worden op toetsbaarheid. Dit gremium bestaat uit experts op het gebied van de data, en denkt mee over de juiste formulering van het informatieverzoek. Vervolgens buigt de Commissie Toetsing, bestaande uit vertegenwoordigers van vervoerders, decentrale overheden en het ministerie van Infrastructuur en Waterstaat, zich over het verzoek om te kijken of het past binnen de eerder genoemde criteria. Als laatste stap vóór het daadwerkelijk uitvoeren van de analyses, moeten de betrokken vervoerders akkoord geven. De transacties zijn bij hen ontstaan, zij zijn daarmee Verwerkingsverantwoordelijke.

5. Van informatieverzoek tot levering van bestanden

De Vervoerregio begon in het voorjaar van 2018 met de voorbereiding van haar informatieverzoek. Het doel was om in elk geval het reisgedrag gedurende het jaar voorafgaand aan de ingebruikname van de Noord/Zuidlijn te kunnen vergelijken met dat van het jaar na deze datum. Dus een periode 22 juli 2017 - 21 juli 2018 met een periode 22 juli 2018 - 21 juli 2019. Voor OV-chipkaartdata geldt een bewaartermijn van achttien maanden. Daarom moest uiterlijk 22 januari 2019 de toestemming geregeld zijn dat Translink de data uit haar database kon verwerken in onze informatiebestanden. Gunstig was dat het informatieverzoek van Vervoerregio als een van de eerste verzoeken gebruik kon maken van de nieuwe Informatiehuishouding. Dit bood voor alle betrokkenen (vervoerders, Samenwerkingsverband DOVA, Translink en Vervoerregio) een gezamenlijk overeengekomen procedure. Desondanks was het ook 'pionieren' voor de betrokkenen. Voor de Vervoerregio was het een leerpunt om de doelen van het informatieverzoek, maar ook de inhoud en vorm van de gevraagde informatie zo eenduidig en overzichtelijk mogelijk te formuleren. Daarbij hebben we de gevraagde informatie ingedeeld in een aantal categorieën, waaronder:

- Aantal reizigers van H(erkomst) naar B(estemming) in verschillende tijdseenheden (zie het einde van dit hoofdstuk)
- Per H/B-relatie de aantallen overstappen die reizigers op overstaplocaties maken
- Per H/B-relatie de gemiddelde reistijd tussen inchecken en uitchecken
- Aantal overstappen van lijn op lijn op overstaplocaties en de overstaptijd daartussen

Het informatieverzoek is besproken in de Commissie Toetsing van het NOVB, waarin alle NOVB-partners (zie hoofdstuk 1) zitting hebben. Na een aantal aanscherpingen heeft deze commissie het verzoek in november 2018 goedgekeurd, waarna de betrokken vervoerders Connexxion, EBS, GVB en NS het verzoek hebben ondertekend. Daarna kon Translink als databeheerder de informatiebestanden gaan samenstellen. Translink en de Vervoerregio moesten hierbij samen nog diverse keuzes maken over de manier waarop we de informatie binnen de privacy spelregels wilden indelen en presenteren in de informatiebestanden. Ook moest Translink onderzoeken of alle informatie te leveren was uit de beschikbare data; gelukkig bleek dat voor de meeste informatie categorieën wel

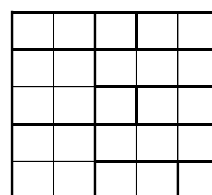
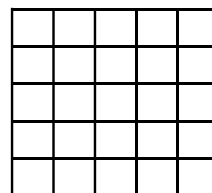
het geval. Een leerpunt was om voor het proces van de uitwerking en de verdere keuzes daarbij ruim voldoende tijd uit te trekken zodat je stap voor stap kunt werken en kunt bijsturen. We merkten dat het lastig is om vooraf precies te voorspellen of bepaalde keuzes goed uitpakken. Soms gold dan de bekende uitspraak van Johan Cruijff: "Je gaat het pas zien als je het doorhebt".

Een belangrijke voorwaarde voor goedkeuring van het informatieverzoek was dat in geen enkel geval herleidbaarheid mag optreden naar een individu. Daarom is onder de drempelwaarde van 300 het aantal waarnemingen altijd geaggregeerd tot klassen (1 – 50; 51 – 100; 101 – 200; 201 – 300). Daarboven wordt het aantal waarnemingen afgerond op twintigtallen, om te voorkomen dat door combinatie van gegevens alsnog klassen tot kleinere aantallen terug te rekenen zijn.

Deze aggregatie dwong de Vervoerregio ook om zorgvuldig een geografische clustering van de herkomst- en bestemmingslocaties te kiezen. Natuurlijk is dit voor goede vervoerkundige analyses sowieso de bedoeling, maar we moesten er ook rekening mee houden dat we bij een te gedetailleerde clustering bijna alleen maar waarnemingen in klassen zouden ontvangen. Als we deze geklasseerde waarnemingen vervolgens verder willen optellen om de aantallen herkomsten en bestemmingen tussen wat grotere gebieden te bepalen, dan krijgen we een tamelijk grote bandbreedte van waarnemingen. Uiteindelijk hebben we gekozen voor twee verschillende clusteringen: een 'fijnmazige' en een 'grofmazige'. In beide clusteringen hebben we alle halte-ID's (afzonderlijke haltepalen) van bus-, tram- en metrolijnen van Connexxion, EBS en GVB in de Vervoerregio verdeeld in groepjes (de clusters). Elk cluster hebben we toegewezen aan één unieke postcode-5. Bij de fijnmazige clustering ging het om kleinere groepjes van haltes dan bij de grofmazige clustering. We hebben er bij de clustering bewust voor gezorgd dat er een vervoerkundig logische indeling van OV-knooppunten en buurten ontstond: bijvoorbeeld alle haltes van Amsterdam Centraal of metrostation Noord in één cluster, of alle haltes in Amstelveen-Zuid in één cluster. Deze werkwijze leidde tot een fijnmazige H/B-matrix van $250 \times 250 = 62.500$ H/B-relaties en een grofmazige matrix van $65 \times 65 = 4.225$ H/B-relaties. Zie figuren 3 en 4 voor een indruk van de clustering(swerkzaamheden).



Figuur 3: clustering van haltes



Figuur 4: fijnmazige en grofmazige clustering; fijnmazige clusters met veel reisbewegingen konden in de grofmazige clustering zelfstandig blijven

Ook hebben we elk treinstation van NS, dat wordt bediend door bus-, tram- of metrolijnen van Vervoerregio-concessies, gekoppeld aan een van de 65 of 250 clusters. Daarmee kunnen we de aantallen overstapbewegingen tussen bus, tram, metro en trein op deze stations zien. Tevens krijgen we inzicht in de aantallen reizen tussen NS-stations en BTM-halteclusters in de Vervoerregio. Daarbij hebben we de NS-stations buiten de regio samengevoegd tot 'windrichtingen': alle stations in Nederland waarvan je op basis van reistijd kunt verwachten dat treinreizigers de regio Amsterdam binnenkomen of verlaten via één bepaald spoortraject. Zo omvat de windrichting 'Utrecht' alle stations in een gebied van Abcoude tot Maastricht, inclusief zijtakken naar bijvoorbeeld Nijmegen of Venlo.

Bij het indelen van de aantallen reizigers naar periode hebben we gekozen voor 'pseudokwartalen', die enerzijds het jaar in enigszins gelijke blokken verdelen en anderzijds samenvallen met veranderingen in de dienstregelingen van de vervoerders (b.v. het einde van de zomervakantie en de gebruikelijke wijziging in de eerste helft van december). Binnen die pseudokwartalen hebben we waar mogelijk weer een fijnere indeling toegepast met gemiddelden per maand, dagsoort en uurblok.

6. Verwerking

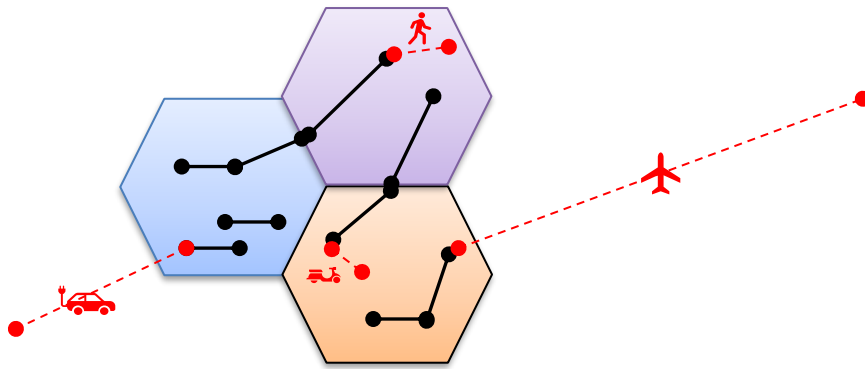
De levering op fijnmazig detailniveau bestond uit 49 bestanden, in zeven bestandsoorten, verdeeld in kwartalen en waar mogelijk kleinere deelperioden. Vanwege de omvang van de bestanden is er gekozen om de verwerking en analyses in R uit te voeren. R is een open-source software pakket voor statistische analyses. Er is een actieve community die zogenaamde packages ontwikkelt voor allerlei doeleinden, van machine learning en neurale netwerken tot visualisaties.

6.1 Wat zit er (niet) in de bestanden?

Het is goed om steeds helder te hebben wat er precies in de bestanden staat: het gaat om aantallen reizen (hele ketens) tussen herkomst en bestemming. Die herkomst en bestemming zijn de eerste instap- en de laatste uitstaphalte. De bestanden kennen uiteraard een aantal beperkingen. Voor de Vervoerregio waren dat de volgende:

- De ketens kunnen alleen worden gemaakt als dezelfde OV-chipkaart wordt gebruikt. De keten van één reiziger met twee chipkaarten kan niet worden gecombineerd en resulteert dus in twee ketens.
- Eenmalige chipkaarten, barcodes en papieren vervoerbewijzen (zoals internationale tickets en dagkaartacties voor de trein) maken geen deel uit van de set. Voor de analyses ten aanzien van de aanbesteding van Zaanstreek-Waterland is dat geen bezwaar, maar we zouden voor de toekomst graag willen dat deze informatie ook beschikbaar komt, omdat het om substantiële aantallen gaat.
- De overstap binnen metro wordt niet gezien, omdat er alleen op het eerste en laatste metrostation wordt in- en uitgecheckt.

Verder zien we alleen het OV-deel van de reis. Er is geen inzicht in de afstand of het gebruikte vervoermiddel vóórdat reizigers instappen in bus, tram, metro of trein.



Figuur 5: alleen het OV-deel van de reis is inzichtelijk

6.2 Verwerkingsstappen

Er zijn maar weinig mensen die direct op basis van postcodes de cijfers kunnen gebruiken. Een eerste stap in de verwerking was het koppelen met CBS-gegevens over wijken, woonplaatsen en de deelgebieden van de Vervoerregio.

Door deze stap is het in principe mogelijk om de vervoersstromen op allerlei niveaus te laten zien: zoals aangeleverd van postcode-5 naar postcode-5, maar ook van wijk naar wijk, bijvoorbeeld vanuit heel Zaanstad naar de verschillende gebiedsindelingen van Amsterdam of per gebiedje in Purmerend naar het gebied rond het Weesperplein in Amsterdam. Een complicerende factor hierbij was de indeling in klassen. Deze indeling zorgt bij het optellen voor een aanzienlijke marge:

Postcode	Plaats	Aantal	Ondergrens	Bovengrens
1234A	CDE	51-100	51	100
1234B	CDE	1-50	1	50
1234C	CDE	101-200	101	200
	Totaal CDE		153	350

De tweede stap in de verwerking was het toevoegen van de kwartaalinformatie. Om te kunnen werken met daggemiddelden is het handig om het aantal dagen in het betreffende kwartaal of de maand te weten.

Vervolgens zijn de afzonderlijke bestanden per type samengevoegd, zodat op een makkelijke manier de periodes met elkaar kunnen worden vergeleken.

6.3 Analyses voor Zaanstreek-Waterland

Een van de aanleidingen om de bestanden te laten samenstellen, was de aanbesteding van de concessie Zaanstreek-Waterland. Momenteel bestaat deze concessie nog uit twee concessies, niet heel verrassend Zaanstreek en Waterland. Beide concessies krijgen hoge cijfers (7,7 en 7,9) in de OV-klantenbarometer 2018, reizigers zijn behoorlijk tevreden met het OV-aanbod. Toch is een aanbesteding ook de kans om nog eens naar de gevraagde opbouw van het OV-netwerk te kijken en daar eventueel wijzigingen in aan te brengen. Om de keuzes in het programma van eisen te toetsen, is een aantal analyses uitgevoerd. We geven hier vanwege het vertrouwelijke karakter van de bestanden en de uitkomsten alleen de vragen en de aanpak weer.

Verdeling reizigers vanuit Zaanstreek-Waterland over Amsterdam

De vervoerkundigen van de Vervoerregio hebben gekeken hoe de bus-/tram-/metroreizen vanuit regio's Zaanstreek en Waterland zich verdelen over Amsterdam. Dit om te kijken of er behoefte is aan nieuwe, directe verbindingen tussen een van de kernen en een specifiek gebied in Amsterdam. Een van de bestandstypes bevat het aantal reizen tussen elke combinatie van postcode-5 gebieden, voor bus/tram/metro. Bij deze analyse was het handig dat zowel de herkomst- als de bestemmingskant kan worden bekeken op verschillende niveaus, van (semi-)postcode-5, buurt-wijk-woonplaats tot gemeente. Voor Amsterdam was het eigenlijk handig geweest om nog een niveau tussen de wijk en de woonplaats te hebben, er zijn in de CBS-indeling eigenlijk net te veel wijken om inzichtelijke kaartjes van te kunnen maken. Postcode-3 niveau heeft hierbij geholpen.

Ontwikkeling van de overstaplocaties

Een ander bestandstype bevat ook het aantal reizen met bus/tram/metro, maar dan uitgesplitst naar aantallen overstappen en de overstaplocaties. Voor die overstaplocaties is dezelfde clustering van haltes gebruikt als voor de herkomsten en bestemmingen. In het oude lijnennet reden (bijna) alle lijnen vanuit Zaanstreek-Waterland naar Amsterdam Centraal. In het nieuwe lijnennet haakt een substantieel deel van de bussen aan op de Noord/Zuidlijn op metrostation Noord. Vanwege de indeling in de pseudokwartalen was het mogelijk om de verschillen tussen het oude en het nieuwe lijnennet in beeld te brengen.

7. Conclusies en vervolgstappen

7.1 Conclusies

Sinds de invoering van de OV-chipkaart heeft het enige tijd geduurd, maar het is inmiddels mogelijk om (onder strikte voorwaarden) de beschikking te krijgen over concessiegrens-overschrijdende reisbewegingen uit de OV-chipkaart. Dit is wel de belangrijkste conclusie.

Het proces om tot een goed informatieverzoek te komen, heeft aardig wat (doorloop)tijd gekost. Enerzijds komt dit doordat de Informatiehuishouding nog vrij nieuw is, en er door de verschillende partijen samen moest worden nagedacht over de verschillende aspecten van het verzoek. Anderzijds was de gevraagde informatie heel uitgebreid, en zijn de ruwe OV-chipkaartbestanden in de database van Translink zodanig groot dat de verwerking tot onze op maat gesneden informatiebestanden an sich ook veel doorlooptijd kost. Door dit soort trajecten wordt het voor alle betrokkenen wel steeds makkelijker om aan te geven waar een informatieverzoek aan moet voldoen en hoe het geformuleerd moet worden.

De hoeveelheid nuttige informatie die in de bestanden zit, is bijna niet te overzien. Bij de uitvraag is veel tijd genomen voor de details: welke gegevens in welke combinaties van kenmerken. Dit zorgt ervoor dat zo'n beetje elke denkbare beleidsvraag (over het reisgedrag in het OV dan) hiermee beantwoord kan worden.

De verwerking van de geleverde bestanden zelf is op een klein aantal punten wat lastig: de weergave in klassen voor de lagere aantallen zorgt voor behoorlijke bandbreedtes bij het aggregeren. De bestanden met de grofmazige indeling kwamen pas later beschikbaar en lossen dit probleem wellicht op.

7.2 Vervolgstappen

Ongeveer wekelijks komen collega's van de Vervoerregio met nieuwe vragen die wellicht beantwoord kunnen worden met de chipkaartbestanden. Er wordt dan steeds eerst gekeken of de vraag past binnen de leveringsafspraken. Een greep uit de vervolgvragen:

- Hoe is het reisgedrag binnen Zaanstad, van reizigers die daar aankomen per trein?
- Hoe verdelen treinreizigers uit een bepaalde windrichting zich over de Amsterdamse treinstations, en waar gaan ze vervolgens naar toe? En zien we wijzigingen in stationskeuze voor dezelfde H/B-relatie, voor en na invoering van de Noord/Zuidlijn?
- In hoeverre is reisgedrag symmetrisch: zien we in de avondspits precies het omgekeerde als in de ochtendspits, of worden er dan andere keuzes gemaakt?
- Zijn de bestanden geschikt om een complete H/B-matrix samen te stellen? Kan die gebruikt worden bij de actualisatie van het regionale verkeersmodel VENOM?
- De bestanden zullen uiteraard een rol gaan spelen bij de evaluatie van de lijnennetvisie.

Wat betreft de verwerking van de bestanden zou het interessant zijn om (in elk geval voor Amsterdam) een geografisch niveau te construeren dat tussen de CBS-wijken/buurtten en postcode-5 niveaus in ligt. Dit om te gebruiken voor visualisaties op kaarten en diagrammen.

De Vervoerregio wil graag het reisgedrag in het regionale netwerk blijven monitoren zodat zij in samenwerking met de vervoerders kan inspelen op veranderingen in reisgedrag en reiswensen. Het voornemen is om het informatieverzoek van afgelopen jaar te herhalen zodat er op termijn meerjarige veranderingen zichtbaar worden in het gebruik van het netwerk en mogelijk ook in H/B-relaties.

8. Leerpunten en tips

Als Vervoerregio hebben we als een van de eerste partijen zo'n omvangrijk en complex informatieverzoek gedaan. Een aantal zaken zal voor elk informatieverzoek verschillen, overeenkomsten zullen er ook zijn. Vandaar een aantal leerpunten en tips vanuit de ervaring die wij inmiddels hebben opgedaan.

- Neem de tijd om de uitvraag van tevoren goed door te denken: wat is precies het doel, waarvoor gaan we de informatie gebruiken en wat is daarvoor het goede aggregatieniveau?
- Ruim ook voldoende tijd in voor het proces bij de Informatiehuishouding. Er zijn veel partijen bij betrokken en de uitvraag moet over een aantal schijven.
- Houd daarbij wel de 18-maanden bewaartermijn voor OV-chipkaartdata in het oog.

- Betrek in een vroeg stadium Translink bij het vraagstuk, zodat zij kunnen meedenken over de aanpak.
- Betrek ook de vervoerders op tijd zodat zij op de hoogte zijn: zowel de vervoerkundigen als de juristen en privacyspecialisten.
- De daadwerkelijke samenstelling van de informatiebestanden – na goedkeuring van het informatieverzoek – kan tijdrovend zijn.
- Begin op tijd met Translink en de vervoerders aan de leveringsvoorwaarden. Juridische aspecten kunnen nog tijd voor nadere afstemming vergen.
- Wees je bewust van de beperkingen van de informatie: sommige betaalwijzen zijn niet opgenomen en alleen ritten met dezelfde OV-chipkaart kunnen worden gecombineerd tot reizen.
- Zorg dat er voldoende kennis is over het (OV-)netwerk en de specifieke kenmerken daarbinnen. Denk bijvoorbeeld aan het ontbreken van inzicht in de metro-metro overstap.
- Beleidsvragen gaan vaak over een voor- en na-situatie. Daar kun je in de uitvraag al rekening mee houden, ook met periodes die in de toekomst liggen.