

# De datapartner van Nederland

*“De kerntaak van het CBS is het feitelijk informeren van de maatschappij”*

Jeroen van Velzen is sinds september 2020 waarnemend hoofddirecteur dataservices, research en innovatie bij het Centraal Bureau voor de Statistiek (CBS), waar hij in 1997 als trainee terechtkwam. Het CBS, opgericht op 9 januari 1899, hanteert in haar strategisch meerjarenprogramma 2019-2023 het credo: maatschappijgericht, betrouwbaar en innovatief. (Rob van Kan, afkomstig van de Nationale Politie Limburg, is sinds 1 februari 2021 de nieuwe hoofddirecteur).



**De regering, gemeenten, de Algemene Rekenkamer, het Sociaal Cultureel Planbureau (SCP), de Sociaal Economische Raad (SER), wie maakt geen gebruik van de informatiebronnen van het CBS voor beleid en besluitvorming? In de 122 jaar dat het bureau bestaat, heeft het zich ontwikkeld tot één van de koplopers waar het gaat om het gebruik van data. “De kerntaak van het CBS is het feitelijk informeren van de maatschappij”, vat Jeroen van Velzen krachtig samen.**

“De toegevoegde waarde die wij hebben is dat wij data kunnen vertalen in maatschappelijke fenomenen, in werkgelegenheid, in economische groei, in sociale cohesie, kortom: allerhande informatie die nodig is voor het maken van beleid of het nemen van besluiten.”

Doordat de informatiebehoefte de afgelopen jaren zo enorm is toegenomen, vertelt Van Velzen, is het CBS meer vanuit de vraag gaan formuleren. “Beleidsmatig zijn veel meer beslissingen op gemeentelijk niveau beledigd en om de verbinding met landelijke ontwikkelingen te leggen, hebben wij ons ook toegelegd op statistische informatie voor gemeenten en provincies. Daarmee helpen we hen data gedreven en *evidence based* te werken.”

## Dialog

Op dit moment zijn er twintig van die centra operationeel, “en beginnen wij steeds meer gevoel te krijgen van wat de informatiebehoefte op dat niveau is. Het goede is dat wij die nu ook generiek kunnen gaan invullen. Natuurlijk, iedere gemeente is uniek en heeft zijn eigen eisen en wensen, maar er is ook iets gemeenschappelijks aan behoefte. We gaan nu proberen om daar een veel breder dienstenpakket voor te maken. Dat is een heel mooie stap, vind ik, want het maakt ook dat we verbindingen kunnen leggen tussen

landelijke en regionale politiek, werelden die lastig op elkaar zijn te leggen. Nu hoeft daar geen licht tussen te zitten, in die zin dat de uitspraken die wij op landelijk niveau doen, gebaseerd zijn op dezelfde methode, dezelfde data, dezelfde inzichten, dezelfde aannames, definities en tijdigheid als op lokaal niveau. Dat maakt dat je een veel objectievere dialoog kunt voeren over hoe het nou eigenlijk gaat.”

Van Velzen vertelt dat buurgemeenten inmiddels aansluiten bij de verschillende centra waardoor er regionale organisaties ontstaan. Een ontwikkeling die hij toejuicht, “want je kunt uiteindelijk niet 300 centers starten, dat moeten we ook niet willen. We gaan naar een nieuwe fase waarin het een meer generieke dienst wordt.” Hij zegt dat er samen met de Vereniging van Nederlandse Gemeenten (VNG) gekeken wordt naar een goede rolverdeling in wat hij noemt “het kader van de grotere maatschappelijke opgave.”

## Rol

Het CBS geniet in de statistische wereld internationaal faam om de manier waarop zij, lang geleden al, de volkstellingen virtueel heeft gemaakt, vertelt Van Velzen. “Het CBS is ook beroemd omdat Nederland als eerste land heel veel van de prijswaarneming gebaseerd heeft op de betalingsgegevens van supermarkten en *web scraping*<sup>1</sup>-achtige technieken.”

Ook biedt het CBS een *remote access* voorziening aan kennisinstututen en universiteiten. “Dat betekent als je onderzoek wilt doen, dat je de data op een veilige manier kunt gebruiken. Let wel: alleen de data die je nodig hebt, zonder identificerende kenmerken en alleen de uitkomsten kunnen, gecontroleerd door het CBS om onthulling te allen tijde te voorkomen, naar buiten. Dat is een faciliteit, ik ken geen land dat dit net zo goed geregeld heeft als wij.”

Het CBS mag trots zijn op de positie die zij heeft bereikt, maar *noblesse oblige* vindt Van Velzen: “Het betekent dat je voorop moet blijven lopen op het gebied van privacy, voorop moet blijven lopen met innovatie. Dat is soms lastig, want innovatie impliceert risico nemen, tot op zekere hoogte, en daar zijn we best zoekende naar. Je kunt niet experimenteren met je producten. Het betrouwbare imago dat we hebben - heilig is een groot woord - weegt voor ons zwaar, en op het gebied van privacy is risico nemen uitgesloten.

Wel ontwikkelen de technieken om privacy te borgen zich in rap tempo en dat biedt ook mogelijkheden.”

“Ik durf hardop te zeggen dat er enorm veel kennis bij het CBS zit en die kennis wordt ook actief onderhouden en ook actief gedeeld. Het CBS heeft maximale meerwaarde als zij geen concessies doet in autonomie en in de manier waarop ze met privacy omgaat. Het expliciet en ook bewust beperken tot de statistische rol binnen het stelsel, die is belangrijk en dat is een soort ijkpunt. Het betekent dat je als CBS ook niet de regie moet voeren over de operationele datastromen die bij de grote uitvoeringspartijen gebruikt worden. Wij gaan niet over de uitvoering, de handhaving of de opsporing. De statistieken bieden wel inzicht voor het maken van goed beleid. Wij willen bij onze rol blijven. Wij zijn als onafhankelijk statistiekbureau feitelijke degene die bepaalt welke statistiek we maken, welk onderwerp voldoende maatschappelijke waarde heeft. Maar sterk gebaseerd op Europese verplichtingen en vooral ook beleidsvragen die aan ons gesteld worden.”

## Innovatie

Het CBS werkt continue aan verbetering van de statistische processen en producten. Het bureau doet onderzoek naar nieuwe databronnen, methoden zoals de inzet van Artificial Intelligence (AI) om nieuwe statistieken te ontwikkelen of om bestaande statistieken te verbeteren. Ook onderzoekt het nieuwe technieken om data te combineren met data van andere partijen op basis van *Privacy preserving techniques* (PPT)<sup>2</sup>.

“Het traject om nieuwe informatiestromen goed te ontsluiten is best complex om goed te doen, want je moet vanuit privacy standpunt eigenlijk de noodzaak kunnen aantonen voor je de pilot hebt gedaan. Er werken bij ons onderzoekers op het gebied van statistische beveiliging en de definitie die wij hanteren is dat je geen individuele gegevens kunt herleiden uit onze statistische output.”

“Traditioneel gezien betekent statistiek maken de methode volledig uitleggen, uittekenen hoe die werkt. Voor een deel is dat bij AI zó complex dat dat bijna niet meer gaat. Dus, hoe zorg je dat een algoritme eerlijk is, geen ongewenste selectiviteit heeft? Hoe borg je dat het systeem robuust is tegen modelaanne? Dat zijn onderwerpen waar wij volop mee bezig zijn. En dat is wel

steeds de slag die we moeten maken vanuit die pilots: wat je ook in je model doet, je krijgt altijd een uitkomst, dat is het punt niet. Maar hoe borg je nú dat het over vijf jaar en over tien jaar nog steeds diezelfde werkelijkheid beschrijft?”

“We hebben een model gemaakt waarin we relaties tussen Nederlanders kunnen beschrijven; de werkrelaties, de gezinsrelaties, de woonrelaties, dat soort relaties kunnen we leggen. Dat wordt in onze ogen een heel belangrijk instrument om de ‘werking van de samenleving als systeem’ te kunnen begrijpen. Dat gaat om grote vragen zoals ‘wat verergert of doorbreekt segregatie en polarisatie tussen bevolkingsgroepen?’, of ‘welke mechanismen verkleinen of vergroten ongelijkheid in Nederland? Het is nog in de pilotfase maar we zijn wel aan het nadenken hoe je dit zou kunnen delen. En dan wel op een manier dat die individuele gegevens gewoon veilig blijven, maar de kracht van het model natuurlijk wel beschikbaar is.”

### Verdieping

“Wat we ook beogen, is om de cijfers die we op lokaal niveau aangeven, als een verdieping te zien van de landelijke cijfers. Als je lokaal een cijfer over werkgelegenheid beschrijft, moet je dat natuurlijk op dezelfde manier op landelijk niveau doen, want anders gaat die vergelijking niet op. Wij hebben een heel sterk argument om ook die definities zo expliciet te maken omdat je dan die vertaling kan maken naar het landelijk cijfer. Dus wij zijn daar, van nature, heel erg op gericht om dat op een zelfde manier te doen. Het idee is: wij maken een raamwerk van de beschrijving van bijvoorbeeld de economie, of de maatschappij, we kunnen verdichten en verdiepen richting een bepaalde doelgroep, een bepaalde regio. En juist de kracht dat we dit vanuit dat raamwerk opbouwen, maakt dat het veel meer samenhang heeft.”

“Onze primaire waarneming bestaat er ook uit dat we enquêtes afnemen bij bedrijven en bij personen. Het bedrijfsleven ervaart dat nog wel eens als een administratieve last en dus zijn we bezig om andere bronnen in te zetten, slim gebruik te maken van situaties. Bijvoorbeeld door samen met bedrijven een toegang tot een administratiesysteem te bewerkstelligen zodat we daaruit zelf informatie kunnen halen. Wij hebben gemerkt dat de bereidheid om mee te werken veel groter werd als we het bedrijf de regie gaven, een meerderheid vond dat heel handig. Dat is ook wat ik met innovatie bedoel: ik denk dat we op zoek moeten naar de technieken en methoden waarin mensen zelf regie kunnen hebben over de data die ze beschikbaar stellen en dat betekent dat je moet weten: wat zijn het voor data, waar is het voor? Daar moet wel een soort van principe inzitten van: dat soort data sta ik toe, dat soort niet en wel voor dat doel maar niet voor dat doel. Dat moet je kunnen wijzigen en daar moet je inzicht in kunnen hebben en in dat begrip zijn we met elkaar nu wel volwassen aan het worden.”

<sup>1</sup> Web scraping is een computertechniek om meta-informatie in de vorm van tekst en foto's van webpagina's te verzamelen.

<sup>2</sup> Privacy preserving techniques (PPT) voorzien erin dat wanneer de gegevens worden overgedragen tussen verschillende partijen, het verplicht is om die gegevens te beveiligen zodat andere partijen niet weten welke gegevens tussen oorspronkelijke partijen zijn overgedragen.

