



ADVERTORIAL

# 'CBS loopt voorop als het gaat om inzet big data'

Piet Daas is sinds 2000 als methodoloog in dienst van het CBS. Zijn specialisatie ligt op het gebied van big data. Sinds januari van dit jaar is hij ook bijzonder hoogleraar Big Data in Official Statistics bij de Technische Universiteit Eindhoven (TU/e). 'Big databronnen bieden veel kansen voor de statistiek. Maar we moeten daarbij wel een aantal fundamentele vragen beantwoorden.'



**H**et CBS wil voor zijn statistieken zoveel mogelijk gebruik maken van data die er al zijn, waaronder de grote hoeveelheden data die afkomstig zijn van het internet, verkeerslussen, sensoren en mobiele telefoons. Prof. dr. Piet Daas: 'Daarmee kunnen we - naast de huidige wettelijk verplichte statistieken - nieuwe of meer gedetailleerde statistieken maken. Dankzij big data kunnen we ook sneller cijfers produceren én verminderen we de enquêtedruk voor burgers en bedrijven. Dat zijn grote voordelen. Maar het gebruik van big data brengt ook een aantal fundamentele vragen met zich mee.'

### Vragen oplossen

Statistieken maken met big data is heel anders dan statistieken maken op basis van enquêtes. Daas: 'Bij een statistiek op basis van enquêtes maak je eerst een ontwerp voor je onderzoek. Dan stel je de doelgroep vast, trek je een steekproef en ga je de data verzamelen. Bij een statistiek op basis van big data draai je het helemaal om: je begint met een set data. Maar bevat die set data alle informatie die je nodig hebt voor je statistiek? Het CBS loopt als statistiekbureau voorop als het gaat om de inzet van big data. Dat betekent dat we zelf de vragen moeten oplossen die daarbij ontstaan. Als hoogleraar krijg ik ruimte om die vragen te onderzoeken.'

### Representativiteit

Daas richt zich onder andere op de representativiteit van big data. Zeggen de data voldoende over de populatie die je wilt meten? Hij legt uit: 'Big databronnen bevatten vaak gegevens over een heel grote groep mensen of bedrijven. Maar komt deze populatie overeen met de doelgroep van je statistiek? Een voorbeeld: wij beschikken over bestanden met data over scheepsverkeer. Die

bevatten alle scheepsbewegingen van alle schepen. Daar zit de representativiteit dus goed. Maar als je op basis van internetdata een statistiek wilt maken over bedrijven, bijvoorbeeld over de vraag hoe innovatief of duurzaam zij zijn, dan mis je informatie over bedrijven die geen website hebben. Dat kan een probleem zijn en als dat het geval is, moet je een oplossing vinden om de ontbrekende data te corrigeren.'

### Trend of toeval?

Een tweede belangrijke vraag betreft de relatie tussen het fenomeen dat je meet in een big databron en datgene waarvoor het CBS de data wil gebruiken. Hierbij speelt causaliteit een belangrijke rol. Laten de data die je gebruikt een trend zien of is er sprake van toeval? Daas: 'We maken officiële statistiek. Wat we publiceren moet kloppen. Het is heel belangrijk dat je onderzoekt of de gegevens uit je databronnen een antwoord geven op de statistische vraag die je stelt. Zeggen bijvoorbeeld de data die afkomstig zijn van de websites van bedrijven écht iets over hoe innovatief en duurzaam ze zijn?'

### Ideale omgeving

Daas heeft veel zin in zijn hooglerarschap. 'Mijn werkweek bij het Center for Big Data Statistics (CBDS) van het CBS is hectisch. Bij de universiteit heb ik nu een dag per week de tijd om me met de vele experts die daar zijn volledig te richten op de theorie. In Eindhoven komen statistische kennis en computer science samen. Dat is een ideale omgeving om big data-onderzoek te doen. We beginnen nu met een literatuuronderzoek, bedoeld om een goede basis te leggen voor het beantwoorden van de fundamentele vragen. Daarna gaan we met de data aan de slag.' ●

## Curriculum vitae Piet Daas

Dr. Piet J.H. Daas is biochemicus en bio-informaticus. In 1996 promoveerde hij cum laude aan de Radboud Universiteit Nijmegen. Daarna werkte hij bij de Universiteit van Wageningen, waar de datasets die gebruikt werden voor het onderzoek alsmaar groeiden. Sinds 2000 is hij methodoloog bij het CBS en legde hij de basis voor het onderzoek naar big data-statistieken. In 2016 werd het Center for Big Data Statistics (CBDS) bij het CBS opgericht om het gebruik van big data voor officiële statistieken verder te stimuleren. Daas wordt internationaal erkend als expert op het gebied van big data en geeft wereldwijd presentaties en trainingen op het vakgebied.

## Expertisecentrum CBS

Het CBS heeft een expertisecentrum CBS en Hoger Onderwijs (ECHO) in het leven geroepen om de strategische relaties met universiteiten en hogescholen te versterken. Het expertisecentrum heeft de taak om uitwisseling van kennis en expertise op het gebied van data, onderzoek, onderwijs en innovatie te realiseren. Dit gebeurt onder andere in de vorm van promotieonderzoek, zoals van Alfred Teye aan de TU Delft naar 'Hoe ontwikkelt de huizenprijs zich en hoeveel risico lopen huizenbezitters op een huis dat sterk in waarde daalt?'. Het CBS innoveert en is permanent op zoek naar nieuwe databronnen en andere manieren van data verzamelen. Daarvoor werkt het CBS onder andere ook samen met TU/e en Wageningen UR op het gebied van sensordata uit de landbouw. Om kennis en ervaring op het gebied van big data op te doen en uit te wisselen, heeft het CBS diverse keren samen met de Universiteit Twente (UT) datacamps georganiseerd, waarbij onderzoekers van CBS en van UT samen met promovendi aan de slag zijn gegaan. Zo was het doel van het datavisualisatiekamp in 2018: visualisaties ontwerpen die de voortgang van de SDG duurzaamheidsdoelen tastbaar maken.

Het CBS heeft regelmatig ruimte voor stages en afstudeeronderzoek.

Ben je geïnteresseerd in een loopbaan bij het CBS? Kijk dan voor meer informatie op: [www.werkenbijhetCBS.nl](http://www.werkenbijhetCBS.nl). Ook voor (afstudeer)stages kun je bij ons terecht.