

Big Data: bigger than privacy

Dorien Verhulst¹

Wat is de gemeenschappelijke deler van gepersonaliseerde advertenties op Facebook, Google Flu Trends, het gevecht tegen *exploding man holes* in New York City, het monitoren van het rijgedrag van vrachtwagenchauffeurs, *predictive policing*, real-time file-informatie, Prism en de verkiezingscampagne van Obama in 2012? Het zijn allemaal innovatieve toepassingen die gebaseerd zijn op grote datasets en de verbeterde mogelijkheden om die te analyseren: Big Data.

In Washington DC bezocht ik de conferentie [Big Data and Privacy: Making Ends Meet](#), georganiseerd door Stanford's [Center for Internet and Society](#) (CIS) en het [Future of Privacy Forum](#) (FPF), een denktank in Washington, met Microsoft als *host*. Onder de sprekers waren vooraanstaande Amerikaanse academici, het publiek bestond overwegend uit lobbyisten en beleidsmakers uit Washington. De inzichten die gedurende de dag langs kwamen waren voor een groot deel echter grensoverschrijdend en daarom ook voor een Nederlandse advocate en lezers van ITenRecht.nl interessant.

Alle sprekers op de conferentie zijn het erover eens dat het blootleggen van verborgen verbanden en het doen van voorspellingen op basis van grote datasets tot enorme innovatie kan leiden. Niet alleen in de commerciële sector (denk aan gerichtere en persoonlijkere advertenties en dienstverlening), maar juist ook in de publieke sector. Een [prachtig voorbeeld](#) is er in New York City, waar Con Edison sinds 1882 het elektriciteitsnetwerk beheert en zo'n 50.000 *man holes* toegang verschaffen tot ruim 30.000 kilometer elektriciteitskabel onder de grond. Tot de gevaren in de stad behoren [ontploffingen in deze putten](#), waarbij zware putdeksels meters de lucht in worden geblazen. Onderzoekers van Columbia University ontwikkelden op basis van historische data sinds 1882 een lerend en voorspellend algoritme dat helpt te bepalen welke delen van het netwerk meest risicovol zijn en als eerste moeten worden geïnspecteerd of onderhouden.

Maar soms is Big Data meer controversieel. Data engineer [Rayid Ghani](#) steelt tijdens de conferentie de show met een inkijkje in het gebruik van *analytics* in de verkiezingscampagne van Obama in 2012. Tijdens die campagne werd een, niet eens zo grote, dataset samengesteld met daarin (openbare) informatie uit [voting registration forms](#). Met data-analyse werd voor alle stemgerechtigde Amerikanen in de *swing states* voorspeld of ze zouden gaan stemmen, of ze voor Obama waren, en zo niet, of ze overtuigd zouden kunnen worden om Obama te steunen. De matrix met uitkomsten laat zich eenvoudig uittekenen en bepaalde de inzet van beschikbare middelen. Campagnevrijwilligers ontvingen lijsten met kiezers, gerangschikt naar hun *pursuadability*. Voor supporters van Obama werd een [Facebook App](#) ontwikkeld die het mogelijk maakte om hun contacten op het sociale netwerk te matchen met Obama's dataset en gerichte aansporingen te ontvangen: *Dorien, tell Alice to go vote!*

¹ Mr. D. Verhulst is advocaat in Amsterdam (Brinkhof)

De bijdrage van Ghani laat zien dat eventuele zorgen rondom Big Data niet beperkt zijn tot privacy. De vraag is niet alleen of een kiezer het erg vindt dat informatie uit zijn of haar registratieformulier openbaar is. De vraag is ook of een kiezer het gepast vindt dat een campagne team voorspelt of, en op wie, hij zal stemmen. Ook de classificatie die kenmerkend is voor Big Data roept vragen op: uiteindelijk worden putten of mensen in hokjes gestopt en verschillend behandeld. De meest risicovolle *man holes* worden geïnspecteerd; de groep kiezers die het makkelijkst overtuigd kan worden op Obama te stemmen ontvangt specifieke informatie. Met de opkomst van Big Data breidt die classificatie zich uit naar een steeds groter aantal terreinen. Een bank kan jonge, hoogopgeleide rekeninghouders identificeren als beste klanten en haar telefonische helpdesk zo inrichten dat deze groep met minimale wachttijd wordt doorverbonden met de beste medewerkers; voorspellingen op basis van data-analyse kunnen bepalen onder hoeveel toezicht voorwaardelijk in vrijheid gestelde gedetineerden worden geplaatst; en, *last but not least*, de onthullingen van Snowden laten zien dat Big Data massa-surveillance door inlichtingen- en veiligheidsdiensten mogelijk maakt.

Als de conferentie iets duidelijk maakt dan is het dat een adequaat juridisch beoordelingskader voor Big Data nog ontbreekt. De voordelen zijn vaak concreet en overtuigend: we willen wel gepersonaliseerde filmsuggesties van Netflix en geen ontplofende putdeksels op straat. Maar hoe weeg je de voordelen zorgvuldig en consistent af tegen eventuele, vaak nogal abstracte, risico's? In het artikel [It's Not Privacy, and It's Not Fair](#) zetten [Cynthia Dwork](#) (Microsoft Research) en [Deirdre Mulligan](#) (Berkeley Law) helder uiteen dat het bestaande privacy-instrumentarium – dat heel kort gezegd draait om controle over 'eigen' data – niet volstaat. De inschatting hoeveel privacy wordt prijsgegeven met een bepaalde handeling wordt bij moderne datatoepassingen te moeilijk. Controle over eigen data heeft ook geen, of slechts indirect, effect op mogelijke negatieve gevolgen, zoals een gebrekkige data-analyse; de onterechte in- of uitsluiting van mensen bij classificatie; of blikvernauwing als gevolg van langdurig geïndividualiseerde informatie. Tegelijkertijd zijn er talloze Big Data-toepassingen met enorme voordelen en lage risico's, denk aan spamfilters of vertaaldiensten. In die gevallen mag onnodige regulering innovatie niet de nek omdraaien.

Intussen staan technologie en economie niet stil: eerder deze maand [voorspelde](#) ABI Research dat de uitgaven aan Big Data tegen het jaar 2018 zo'n 114 miljard dollar bedragen en volgens [recent onderzoek](#) van Gartner is 64% van alle bedrijven van plan te investeren in Big Data, of heeft dat al gedaan.

Ingezonden artikelen van vrijwel alle sprekers op de conferentie zijn gepubliceerd in een speciale editie van de [Stanford Law Review Online](#).

Dorien Verhulst