

De Amerikaanse inlichtingendienst NSA verzamelt bergen informatie van burgers over de hele wereld. Waarom zijn ze geïnteresseerd in al die data?

TEKST **Kristel van Teeffelen**

Uw gedrag is goud waard

“Verzamel het, label het en sla het op”. Met dit motto geeft generaal Keith Alexander leiding aan de Amerikaanse inlichtingendienst NSA. In plaats van op zoek te gaan naar de speld, vergaart hij liever de hele hooiberg, zo werd Alexander onlangs in Amerikaanse media getypeerd.

Dat het de NSA-baas lukt om naar zijn motto te leven, bewijzen de documenten van klokkenluider Edward Snowden. De NSA verzamelt een enorme berg aan informatie van internetgebruikers over de hele wereld. Met een groot net slaapt de dienst alles naar binnen wat medewerkers tegenkomen op internetkabels of op de servers van internetbedrijven. Pas later wordt er gekeken wat relevant is. Zo weten we tot nu toe dat de afgelopen jaren onder meer 1,8 miljoen telefoongesprekken uit Nederland, zo'n 250 miljoen adresboeken van e-mailaccounts, en ontelbare hoeveelheden profielgegevens van sociale media in de datacentra van de NSA zijn opgeslagen.

Wat moeten Amerikanen met al die informatie van die miljoenen burgers? Het tegengaan van terrorisme, luidt de eigen verklaring. “Het is voor de NSA van belang om de relaties tussen mensen in kaart te brengen”, zegt Sander Duivestein van het Sogeti Verkenningeninstituut Nieuwe Technologie, dat volgende maand het boek ‘No More Secrets’ over big data uitbrengt. “Door gegevens van telefoongesprekken en e-mailadressen te analyseren, krijgen ze een beeld van het netwerk van mensen. Ook die van potentiële terroristen.”

Maar de mogelijkheden met al die opgeslagen data zijn veel groter. Vooral commerciële bedrijven maken al volop gebruik van big data-onderzoek. Zoveel, dat we volgens deskundigen zelfs midden in een big data-revolutie zitten. Daar bent u ook onderdeel van. Of liever: uw gedrag.

Het idee: alles wat u doet – door de stad struinen, een e-mail sturen, een AH-bonusaanbieding kopen, een foto op Facebook delen – levert bedrijven en overheden informatie op. Informatie die, mits op de juiste manier geanalyseerd, veel geld waard is. Duivestein: “In de internetsector wordt gepraat over data als de nieuwe olie. Ruw is het niet veel waard, maar als je het weet te destilleren, en het weet te combineren, dan is de waarde enorm.”

Opgeslagen data

Die big data-revolutie komt niet uit de lucht vallen. De hoeveelheid opgeslagen informatie in de wereld is enorm en nog altijd groeiende.

Om hoeveel informatie het precies gaat, is alleen maar te schatten. Viktor Mayer-Schönberger, hoogleraar aan het Oxford Internet Institute, en *Economist*-journalist Kenneth Cukier hebben het in hun boek ‘Big Data Revolutie’ over 1200 exabyte aan opgeslagen data in 2012. Om een beeld te geven hoeveel dat is: als al die informatie in gedrukte boeken stond, dan zou de hele VS met een laag ter dikte van 52 boeken zijn bedekt. Als het op opgestapelde cd-roms zou worden gezet, dan zou je vijf stapels hebben van de aarde naar de maan.

Maar wat kun je met die grote berg data? Misschien wel het beroemdste voorbeeld komt van de Amerikaanse winkelketen Target. Een vader klopte boos aan bij de winkel omdat zijn tienerdochter voordeelbonnen van babyproducten kreeg toegestuurd. De winkel wilde daarmee zijn dochter op ideeën brengen, dacht de vader. Wat hij toen nog niet wist, maar Target al wel, was dat het tienermeisje inderdaad zwanger was. Door uit de enorme informatiebak van klantgegevens het aankoopgedrag van zwangere vrouwen te analyseren, kwam Target tot de conclusie dat het meisje hetzelfde patroon vertoonde als die zwangere vrouwen. De winkelketen wist bijvoorbeeld dat geurloze bodylotions populair zijn bij vrouwen in de eerste drie maanden van hun zwangerschap.

Gerichte marketing

Het voorbeeld van Target toont hoe bedrijven big data kunnen inzetten om winsten te vergroten. Maar de techniek wordt niet alleen gebruikt voor gerichte marketing. Eigenlijk gaat het verder dan dat: door een grote hoeveelheid gegevens te analyseren, daar patronen in te herkennen, kan big data gedrag voorspellen.

Ook de overheid heeft daar baat bij. Nu al wordt er in Amerika getest met big data-onderzoek dat crimineel gedrag voorspelt. Zo maakt de politie in de Californische plaats Santa Cruz gebruik van een systeem dat berekent waar op dat moment de grootste kans op een misdrijf is. De voorspellingen zijn gebaseerd op informatie over criminaliteit in het verleden. Surveillerende agenten krijgen door welke wijk het hoogste risicogetal heeft en kunnen er een extra rondje rijden.

Nog een voorbeeld: de Amerikaanse hoogleraar statistiek en criminologie aan de Universiteit van Pennsylvania, Richard Berk, beweert dat hij in 75 procent van de gevallen goed kan voorspellen of iemand die voorwaardelijk is vrijgelaten uit de gevangenis, betrokken zal raken bij een moord. Hij gebruikt daarvoor een

enorme berg aan data: zoals leeftijd, geslacht, reden van gevangenisstraf en de datum van het eerste gepleegde criminele feit. Daar laat hij een ingewikkelde analyse op los en zo schat hij de kans op een moordpartij.

En niet alleen op het gebied van criminaliteit is big data voor de overheid inzetbaar. Zo bleek zoekgigant Google tijdens een griepiepidemie in 2009 beter in staat het verloop van het virus te voorspellen dan de officiële overheidsinstelling dat kon. Door gigantische hoeveelheden zoekopdrachten van mensen te analyseren, wist Google virushaarden aan te wijzen, terwijl de instantie moest wachten op de meldingen en daardoor als het ware achter de feiten aanliep.

Wie denkt dat dat al revolutionair is, staat als je Viktor Mayer-Schönberger en Kenneth Cukier mag geloven, nog heel wat te wachten. We staan nog maar aan het begin van de big data-revolutie, zo betogen zij in hun boek. Het opslaan en vergaren van grote hoeveelheden informatie wordt steeds goedkoper en eenvoudiger. Steeds meer bedrijven en overheden zien daarom hun kans schoon en investeren in dit onderzoek.

Maar gevaren kent de revolutie ook, betogen de twee. Neem de praktijk dat gegevens worden gebruikt om crimineel gedrag te voorspellen. Gaan we iemand veroordelen om een misdrijf dat hij nooit gepleegd heeft, maar volgens patronen in het big data-onderzoek mogelijk wel gaat plegen? Mag een crimineel die volgens onderzoek grote kans heeft een moord te plegen, straks niet met verlof?

Ook Sander Duivestein ziet nadelen aan de big data-revolutie: “Er zijn straks geen geheimen meer, alles wat je ooit hebt gedaan, valt altijd te achterhalen. Laat je zelf niets achter,

Dat zijn dochter zwanger was, wist de winkel via haar klantgegevens eerder dan hijzelf

dan is het wel je omgeving, of zijn het andere instrumenten, als camera's op straat, die je registreren. De vraag is wie er bij die informatie kan en wat hij daarmee gaat doen.”

Terug naar de NSA. Dat de organisatie grote hoeveelheden gegevens van miljoenen burgers over de hele wereld heeft opgeslagen, weten we sinds Edward Snowden met zijn onthullingen kwam. Maar wát ze precies met die informatie doen, weten we volgens Viktor Mayer-Schönberger nog altijd niet. “De NSA weigert tot nu toe te vertellen waarom ze al die informatie precies verzamelen. Wel is er in de Amerikaanse media gesuggereerd dat veel data-analyse wordt gebruikt om relatief gewone misdrijven op te lossen. Als dat klopt, dan is de VS een gigantische surveillancestaat aan het bouwen, die burgers over de hele wereld in de gaten houdt. Of dit een evenredige toepassing van een krachtig hulpmiddel is, of gevaar oplevert voor de fundamenten van onze samenleving, is een vraag waarover we zouden moeten debatteren.”

Bonusaanbieding

De potentie van de NSA is enorm, zegt Sander Duivestein. “Bedrijven als Google en Facebook bezitten heel veel informatie van hun gebruikers, maar de NSA kan verschillende databanken combineren en dat geeft ze veel macht.”

Inmiddels worden er van verschillende kanten bovendien vraagtakens gezet bij de stelling dat het de NSA puur om terrorismebestrijding gaat. Om maar een voorbeeld te noemen: daarmee is het afluisteren van de Duitse bondskanselier Angela Merkel nog niet verklaard.

Ondertussen gaat de opslag van informatie over uw gedrag gewoon door. Daar ziet u, afgezien van een gerichte bonusaanbieding, momenteel nog weinig voor terug. Daarom verwachten deskundigen dat burgers in de toekomst het heft meer in eigen hand nemen. Zo bestaan er al bedrijfjes die de gegevens namens mensen te gelde maken.

Maar hoeveel zijn al die data eigenlijk waard? Dat is moeilijk te zeggen. De waarde zit hem niet per se in de data zelf, maar in de manier waarop de gegevens gebruikt worden. Zo was er een Amerikaans bedrijf dat nog voordat de kwartaalcijfers van winkelketens bekend waren, kon voorspellen of het een positief bericht ging worden of niet. Daarvoor analyseerden zij de gegevens van autoverkeer: was het druk geweest op de parkeerplaats? Met de verkeerspatronen konden mensen op die manier opeens geld verdienen op de beurs.