



Grote data, kleine data

BIG DATA



Marc Teunis

'Ja Big Data he!' Ik hoor het naklinken terwijl ik wegzap van RTL Boulevard. Het thema ging over 'App de vakantiedokter'. Een nieuwe service van Zilveren Kruis om tijdens de vakantie in het buitenland medische hulp te krijgen. De gasthuisarts van het programma voegt er nog aan toe dat het wellicht makkelijker is om via 'App de vakantiedokter' je vragen te stellen dan om naar een lokale huisarts of ziekenhuis te gaan. Zij zouden, naar zijn zeggen, je toch alleen maar willen afzetten en het communiceren in het Frans op je vakantie is toch al lastig genoeg, laat staan dat je een arts moet uitleggen wat je mankeert.

Dat 'Big Data he', waar komt die opmerking nou vandaan? Zilveren Kruis kan met deze App informatie verzamelen over medische zaken die relevant zijn voor mensen die op vakantie zijn. Een mooi initiatief voor Nederlanders op reis denk ik voor vragen van 'niet-zo-acute' aard. Voor Zilveren Kruis ook mooi want zij krijgen via deze App allerlei informatie. Maar er zijn ook schaduwkanten. De zorgverzekering die meekijkt met jou wat je op vakantie doet en mankeert?

Big Data...Het lijkt het buzzwoord van 2015. Te pas en te onpas hoor je er over. Volgens de *Big Data* Goeroes van deze wereld alleen maar voordelen. Je kunt er meer geld mee verdienen, de kosten reduceren, de veiligheid van je bedrijf, je medewerkers en je producten mee verhogen en innovatie mee versterken. De kwaliteit van de gezondheidszorg kun je er mee verbeteren en ook in het wetenschappelijke onderzoek is *Big Data* niet meer weg te denken. Maar wat is nu eigenlijk *Big Data*?

Het hangt af van je referentiekader. Op het moment dat de dataset nog in te laden is op je eigen laptop of pc is het waarschijnlijk nog *Small Data*. Het wordt pas echt *Big* als voor het inlezen en analyseren van de dataset(s) meerdere, zo niet duizenden, processoren tegelijkertijd noodzakelijk zijn. Om met dergelijke hoeveelheden data te kunnen werken is specialistische kennis nodig waar ICT giganten zoals IBM groots op inzetten. Hun bedrijfsmodel verandert van het produceren van hardware en onderhouden van infrastructuur naar het managen van data en het ontsluiten van waardevolle informatie uit die enorme hoeveelheden data.

Is *Big Data* hetzelfde als de *Cloud*? Nee, maar het heeft veel met elkaar te maken. Softwarefa- »

brikant Tableau zegt over deze relatie: *Big Data* gaat gelijk op met dataopslag in de Cloud (1). De toename in hoeveelheid digitale data stuwt de vraag naar grotere datadragers en daarmee naar flexibelere dataopslag. Bekende grote spelers zoals Google en Microsoft spelen daar met hun Clouddiensten handig op in. Een gratis account bij een deze aanbieders geeft je de mogelijkheid om 15 Gigabyte aan data op te slaan. Ter referentie voor wie ze nog kent: dat staat gelijk aan 10.417 standaard 3½-inch floppy disks! En dat is dan weer 1/1000-miljardste deel van de totale hoeveelheid data die er per dag zou worden gegenereerd (2). In pakweg twintig jaar is de capaciteit om je persoonlijke data op te slaan een factor 10.000 toegenomen. Tegenstanders van *Big Data* zien vooral nadelen m.b.t. bescherming van persoonlijke informatie en de mogelijkheid dat data in de verkeerde handen kan komen. Het eigenaarschap van *Big Data* is een ander lastig onderwerp. Wanneer je iets gebruikt dat van iedereen is, naar wie gaat dan de gegenereerde meerwaarde? Het debat over *Big Data* wordt gevoerd, door overheden, besturen, management van bedrijven, wetenschappers en het publiek: over een ding zijn ze het eens: ‘*Big Data is here to stay.*’ Het is nu aan ons de keuze: verzetten of meegaan? Ik kies zelf voor het laatste en heb inmiddels mijn eerste code in R geschreven. ‘Kom maar op met die *Big Data!*’

Referenties

- 1 Whitepaper door Tableau ‘Top 7 Trends in Big Data for 2015’, <http://www.tableau.com/learn/whitepapers> (29-06-2015).
- 2 ‘IBM What is big data? — Bringing big data to the enterprise’. www.ibm.com (29-06-2015).

