

## Big Data: A Revolution That Will Transform How We Live, Work, and Think

**Menno Weij en Yentl van den Winkel**

Vandaag de dag hebben we 2.8 ZetaByte aan data verzameld, dat zijn (jawell!) 21 nullen en vier keer zoveel zandkorrels als dat er op de aarde zijn. Elke tien minuten komen daar nog eens 5 biljoen GigaByte aan data bij. De big data revolutie is in volle gang en the best is yet to come. Professor Mayer-Schönberger, hoogleraar aan het Oxford Internet Institute, en Kenneth Cukier, dé Big Data-journalist van The Economist, concluderen dan ook *“big data is everywhere.”* Samen schreven zij het boek *‘Big Data: A Revolution That Will Transform How We Live, Work, and Think.’* Het is hét big data boek van nu. Een optimistisch, praktisch en prettig geschreven boek over een onderwerp dat volgens de schrijvers de wereld, van bedrijven en de natuurwetenschappen tot de gezondheidszorg, de overheid, het onderwijs, de economie, de alfawetenschappen en alle andere aspecten van de samenleving gaat veranderen.

Het koppelen van beschikbare data met andere databronnen om te komen tot voorspellingen, analyses en inzichten met betrekking tot een bepaald verschijnsel is een revolutionaire ontwikkeling. De tien hoofdstukken, die onderling verband met elkaar houden, beschrijven allen een relevant aan big data verwant onderwerp en worden toegelicht met tal van een leuke, opmerkelijke en fascinerende voorbeelden, van Google books tot sumoworstelaars, datingsites, Facebook, autocorrectie op smartphones en van de genetische configuratie tussen de genen van Steve Jobs en medicatie tot Amazon, Twitter en de ploffende putten in New York. Ook voor degene die juist nog niet veel van big data weten en er alles over te weten willen komen beschrijft het boek dit revolutionaire onderwerp op begrijpelijke wijze. Naast het verhelderen van het big data onderwerp wijzen de schrijvers ons op de mogelijkheden en voordelen van big data, maar ook op de ‘darkside’ ervan.

Met de rijzende big data ster gaan drie verschuivingen gepaard in de manier waarop we informatie analyseren. Daarmee nemen we afscheid van small-data, hypotheses en (a) selecteesteekproeven of zoals Mayer- Schönberger en Cukier schrijven: vasthouden aan steekproeven in een big data tijdperk is net zoiets als met paard en wagen blijven rijden in het tijdperk van de auto.

De eerste verschuiving noemen de schrijvers ‘more’. Deze verschuiving ziet op de mogelijkheid om gigantische hoeveelheden gegevens over een onderwerp te bewaren en te analyseren in plaats van uitsluitend met kleinere verzamelingen te werken. Wanneer we alle gegevens met betrekking tot een bepaald verschijnsel bekijken ( $N=\text{alles}$ ) is het mogelijk verbanden, patronen en details te ontdekken die we anders niet zouden opmerken. De dataonderzoeker heeft ook meer vrijheid om de gegevens vanuit verschillende invalshoeken te onderzoeken of om bepaalde aspecten nauwkeuriger te bekijken. In de small data wereld werden gegevens nauwkeurig geselecteerd om een zo hoog mogelijke gegevenskwaliteit te garanderen en om fouten zoveel mogelijk te reduceren. In veel nieuwe situaties waarmee we tegenwoordig te maken krijgen, kan het toestaan van onnauwkeurigheid, rommeligheid, juist positief zijn in plaats van een tekortkoming. Daarop ziet dan ook de tweede verschuiving die de schrijvers ‘messy’ noemen. Er geldt dus niet alleen dat ‘meer’ de voorkeur boven ‘weinig’ verdient, maar ook dat ‘meer’ de voorkeur boven ‘beter’ verdient. We geven in de big data wereld daarnaast de voorkeur aan ‘sordigheid’ boven ‘exactheid’ om dichterbij de werkelijkheid te kunnen komen. De derde verschuiving ziet op het begrijpen van die werkelijkheid en die noemen Mayer- Schönberger en Cukier ‘correlation’. We hoeven niet meer te weten *waarom*, maar *wat* is genoeg. Hoewel het

mensen eigen is om te willen verklaren waarom iets op die manier werkt of zo is, kan aannemen dát het zo is zeer efficiënt zijn.

In het volgende hoofdstuk wordt uitvoerig het begrip 'dataficatie' beschreven. Dataficieren houdt in dat je een verschijnsel in gekwantificeerde vorm brengt waarin het kan worden vastgelegd en geanalyseerd. Om kwantificeerbare informatie vast te leggen moeten we weten hoe we kunnen meten en hoe we onze metingen vast kunnen leggen. Daarvoor moeten we volgens de schrijvers over de juiste gereedschappen beschikken en daarbij moeten we ook de wens hebben om te kwantificeren en vast te leggen. Dataficatie moet los worden gezien van digitalisering. Digitalisering heeft dataficatie wel vleugels gegeven maar vervangt deze niet, aldus Mayer-Schönberger en Cukier. Digitaliseren op zich –het omzetten van analoge informatie in een door de computer leesbare vorm – is nog geen dataficieren. Met de komst van computers is dataficieren wel een stuk eenvoudiger geworden. Dataficatie zal de samenleving ingrijpend veranderen, we dataficieren namelijk alles om ons heen.

De economische waarde van big data zit in het hergebruik van gegevens. De waarde van gegevens verschuift van het primaire gebruik naar de potentiële toekomstige toepassingen (optiewaarde). Er bestaat volgens de schrijvers nog geen voor de hand liggende manier om de geldelijke waarde van gegevens ('onstoffelijke activa') voor een bedrijf te bepalen. Zij vermoeden wel dat de waarde van gegevens op den duur als nieuwe activa op de balans zal verschijnen zodra boekhoudobstakels en zorgen over wettelijke aansprakelijkheid zijn weggenomen.

In het volgende hoofdstuk richten de schrijvers de aandacht op bedrijven die gegevens gebruiken. Mayer-Schönberger en Cukier onderscheiden drie soorten bedrijven: bedrijven die de gegevens bezitten, bedrijven die de expertise hebben om gegevensanalyses uit te voeren en bedrijven met ideeën ofwel een big data mindset. Op dit moment, aan de beginfase van big data, lijken ideeën en expertise de grootste waarde te hebben. Uiteindelijk zal het grootste deel van de waarde in de gegevens zelf liggen omdat we meer met informatie zullen kunnen doen en de bezitters van gegevens zich de (potentiële) waarde van de gegevens meer zullen realiseren.

Naast de positieve kanten van big data gaan de schrijvers ook uitvoerig in op de darkside ervan. Zij beschrijven drie risico's: privacy, de neiging tot straffen en een gegevensdicatuur. Dat big data de privacygevaaren vergoot is niet nieuw, maar door het secundaire gebruik van gegevens verandert de aard van het probleem. Het principe van 'informer en vragen om toestemming' is niet meer toereikend. Het alternatief, de gebruikers meteen om toestemming voor toekomstig hergebruik vragen, is ook geen optie. Dan kan je net zo goed niks vragen! Ook anonimiseren biedt geen uitkomst omdat door de kwantiteit en variëteit van de gegevensverzameling het makkelijker is om alsnog een persoon te herleiden. De privacybeschermingsmaatregelen zijn in het big data tijdperk aldus niet meer toereikend.

Het risico neiging tot straffen ziet op een soort 'preventieve misdadbestrijding'. Big data analyses kunnen voorspellen of iemand een misdaad zal plegen waarvoor we de potentiële dader willen straffen. Dit is volgens de schrijvers in tegenspraak met grondbeginselen zoals de onschuldpresumptie, vrije wil en rechtvaardigheid. We weten immers nooit zeker of diegene het misdrijf daadwerkelijk zou plegen. De schrijvers illustreren deze situatie met de openingsscene van de film *Minority Report*. Hierbij valt een politieagent een woning binnen en treft daarbij een man aan die in alle staten voornemens is een schaar in de borst van zijn vrouw te steken. In mijn ogen een niet heel gelukkig gekozen voorbeeld omdat dit mij nou een situatie lijkt waarin het voorspellen van misdaad uitstekend kan werken. Want hoe verhoudt de 'preventieve

'misdaadbestrijding' zich tot strafbare poging dan wel voorbereiding? Daarvan lijkt mij in het door Mayer- Schönberger en Cukier aangehaalde voorbeeld toch zeker wel sprake van. Het risico ziet volgens mij nu juist op de situatie dat de politieagent naar aanleiding van bijvoorbeeld verontrustende zoekgegevens in Google, zoals "hoe vermoord ik mijn vrouw" en "is een schaar een effectief moordwapen", de woning zou binnenvallen en het stel zit koffiedrinkend het nieuws te kijken. In dat geval wordt wel onrecht aan de hiervoor genoemde grondbeginselen gedaan.

Het derde risico dat de schrijvers benoemen ziet op een heel oud probleem: het vertrouwen op getallen die in werkelijkheid veel feilbaarder zijn dan we denken. We raken zozeer gefixeerd op de gegevens en zo geobsedeerd door de macht en wijsheid die ze beloven, de we de beperkingen ervan (o.a. slechte kwaliteit, vertekenend, verkeerd geanalyseerd) niet meer zien. Als big data voorspellingen beter worden, zal die verleiding volgens de schrijvers alleen maar toenemen.

De schrijvers stippen niet alleen de risico's aan, maar komen ook met oplossingen. Wij moeten de gegevens controleren zodat zij niet ons gaan controleren. Een paar kleine aanpassingen zijn hiertoe niet voldoende. Er dient een verschuiving in de regelgeving plaats te vinden van privacy door het vragen van toestemming naar privacy door aansprakelijkheid. In dit alternatieve privacykader is de gegevensgebruiker aansprakelijk. De oplossing voor de neiging tot straffen wordt gezien in een fundamentele pijler van big data beheer die de garantie moet zijn dat we mensen zullen blijven beoordelen op grond van hun persoonlijke verantwoordelijkheden en hun daadwerkelijke gedrag en dat we niet door 'objectieve' gegevensanalyse gaan bepalen of ze mogelijk misdadigers zijn. Voor het gegevensdicatatuur risico menen de schrijvers dat we deskundigen en instellingen, 'algoritmisten', nodig hebben zodat controle en transparantie van big data verplicht kunnen worden. Zij onderscheiden hierbij interne algoritmisten zoals een bepaalde afdeling binnen een bedrijf van externe algoritmisten zoals onafhankelijke controleurs.

De grootste verandering met de komst van big data is volgens Mayer- Schönberger en Cukier dat de gegevens gaan spreken. Niet meer de verzameling gegevens is van belang, maar hoe het wordt toegepast. Dat zal ons leven, denken, werken eigenlijk *alles* om ons heen veranderen. De mens verdwijnt –gelukkig– niet helemaal achter de coulissen van het big data toneel. In de big data wereld moeten we volgens de schrijvers juist onze aller-menselijkste waarden koesteren, onze creativiteit, intuïtie en intellectuele ambitie. De vindingrijkheid van de mens blijft namelijk de bron van vooruitgang.