



*

* Wat is dit? Het wordt vanzelf duidelijk als je dit boek leest.

INHOUD

Voorwoord 9

Door Viktor Mayer-Schönberger

Inleiding: Alles kan 15

Extreme verwachtingen zijn het nieuwe normaal

1. Big, bigger, biggest data 25

*De nieuwe mogelijkheden die ontstaan door
een overvloed aan data*

2. Angst wegnemen en weerstand overwinnen 40

*Bedrijven moeten inzetten op de maatschappelijke waarde
van Big Data*

3. You ain't seen nothing yet 64

*Hoe alles op zijn kop gaat in de nieuwe
informatiesamenleving*

4. Met scherp schieten op patiënten 75

*Zeg me welke tumor u hebt en ik heb het juiste
medicijn*

5. Dataprikkels voor een betere wereld 104
Er ontstaan nieuwe evenwichten nu de wereld tegen grenzen op loopt
 6. Data-analyse is de samenleving 126
Ethiek in de nieuwe informatiesamenleving
 7. Wanted: duizenden Sherlock Holmes-kloons 141
Data-analyse vraagt om het uitschakelen van vooroordelen en verwachtingen en het inschakelen van het gezond verstand
 8. De vraag is belangrijker dan het antwoord 150
Onderwijsinstellingen moeten een andere rol krijgen
 9. Transparantie als wapen voor de burger 173
Burgers dwingen zelf verandering af en hebben daar geen wetgeving voor nodig
 10. Systemen bepalen ons gedrag 192
Maar de wetgever is niet machteloos
 - II. Een nieuw ecosysteem 205
De nieuwe informatiesamenleving vraagt om nieuwe waarborgen
- Epiloog 225
The art of looking sideways
- Dankwoord 233
- Noten 235
- Register 247

VOORWOORD

Big Data is belangrijk – heel belangrijk zelfs. Maar op een subtielere en meer fundamentele manier dan tegenwoordig vaak wordt voorgesteld. Marketeers van consultancy- en ICT-bedrijven schreeuwen de voordelen van Big Data van de daken, oorverdovend hard, met weinig variatie en op steeds hogere snelheid. Hun mantra is dat met nieuwe ICT-technologieën en een consultant als partner ieder bedrijf een Big Data-supermacht kan worden. Dat is pertinent onwaar, en het gedrag van de marketeers is zowel voorspelbaar als onverantwoordelijk.

Tegelijkertijd schilderen commentatoren en zelfbenoemde cyberexperts grimmige, monochromatische visioenen van een toekomst met Big Data, waarin al het menselijke geëlimineerd is. Als je sommigen van hen moet geloven, veranderen we met Big Data onze computers in massavernietigingswapens. Zulke doemscenario's zijn misschien goed voor de boekenverkoop, maar net als de Big Data-hype helpen ze ons niet erg bij het begrijpen van wat Big Data nu eigenlijk is – en hoe fundamenteel het onze economie en onze samenleving zal veranderen.

Big Data is in wezen niets anders dan een proces, een

uniek mechanisme dat wij mensen gebruiken om de werkelijkheid om ons heen te duiden, en op basis waarvan we voorspellingen doen over onze toekomst die veel accurater zijn dan het koffiedikkijken uit het verleden,

In de hele geschiedenis van de mensheid hebben we de wereld om ons heen geduid door ernaar te kijken en erover na te denken – door te theoretiseren over hoe de werkelijkheid is opgebouwd. We verzamelden gegevens over de wereld om onze theorieën te bewijzen (of te ontkrachten). In de loop van de tijd werden we ons bewust van de noodzaak de wereld rationeel te beschouwen en te proberen haar op systematische wijze te begrijpen. Dat is de kiem die tot bloei kwam tijdens de Verlichting en die werd geplant door onder anderen Spinoza, die onmisbare intellectuele gift van Nederland aan de wereld. Uiteindelijk leverde dit de ‘wetenschappelijke methode’ op, en de grote vooruitgang in het begrijpen en benutten van de wereld die plaatsvond in de negentiende en twintigste eeuw. Als gevolg daarvan leven veel meer mensen langer en beter.

Maar tot voor kort was het verzamelen, analyseren en opslaan van data tijdrovend en kostbaar. Daarom verzamelden we de minimale hoeveelheid data die nodig was om onze vragen te beantwoorden. Sterker nog, in het tijdperk van ‘small data’ was zelfs de manier waarop we inzicht verwierven uit data ingegeven door de schaarste ervan. We maakten gebruik van steekproeven, niet van alle data – of het nu ging om kwaliteitscontroles in fabrieken of om peilingen in verkiezingstijd. Die *shortcut* werkte als hij correct werd toegepast (wat niet altijd het geval was), maar miste details. We konden zelden dieper ingaan op een interessant resultaat, omdat we gewoon niet genoeg gegevens hadden.

Nog ongunstiger was het dat we met zo'n benadering alleen vragen konden beantwoorden die we al vanaf het begin hadden, en de data niet effectief konden gebruiken om nieuwe vragen te stellen die konden leiden tot werkelijk vernieuwende inzichten. We gaven daarnaast voorrang aan de kwaliteit van data en streefden ernaar de accuraatheid te verbeteren van de weinige gegevens die we verzamelden, zelfs als dat hoge kosten met zich meebracht.

Tegenwoordig zijn veel van de grenzen aan het verzamelen, analyseren en opslaan van grote hoeveelheden data sterk vervaagd. Dat maakt het veel eenvoudiger om alle of bijna alle data te verzamelen, en zo in veel meer gevallen de complexiteit van onze wereld te vangen in plaats van genoeg te nemen met een kleine steekproef van de werkelijkheid. Het stelt ons ook in staat er veel vaker voor te kiezen meer, maar 'messy' data te verzamelen in plaats van heel kleine hoeveelheden.

Het resultaat is dat we de wereld nu in ongekend detail kunnen bekijken, en Big Data-analyse levert nieuwe vragen op, die ons op het spoor zetten van waardevolle inzichten die we nog niet hadden. Zo kan Google de verspreiding van griep voorspellen door de miljarden zoekopdrachten te analyseren die binnenkomen. Zo kan Rolls Royce voorzien wanneer een onderdeel in een van zijn vliegtuigen kapot zal gaan voordat dat daadwerkelijk gebeurt, wat niet alleen geld bespaart, maar ook levens kan redden. En zo kan een start-up als Inrix honderd miljoen klanten op weg van en naar hun werk de files laten omzeilen, en tegelijkertijd de lokale overheid adviseren over ontstane verkeerspatronen. Zo kan de Watson-computer van IBM de beste menselijke deelnemers verslaan in de tv-quiz *Jeopardy* en artsen helpen bij het stellen van diagnoses. In het

Hospital for Sick Children in Toronto is zelfs een manier gevonden om levens van premature baby's te redden door gebruik te maken van Big Data. In het tijdperk van Big Data hebben meer en 'messy' data het overgenomen van schone, maar kleine data.

Maar er is een derde belangrijke verschuiving in hoe we de wereld duiden. Wij mensen geven er de voorkeur aan de wereld te zien als een aaneenschakeling van oorzaken en gevolgen. We zoeken dus onwillekeurig naar de oorzaken van alles wat we opmerken – de geluiden die we horen, hoe onze auto's remmen, hoe consumenten winkelen en zelfs hoe we ons voelen. Maar het vinden van die oorzaken blijkt lastig te zijn. Zoals Nobelprijswinnaar Daniel Kahneman liet zien, wijzen we ze vaak intuïtief aan, zonder doorslaggevend bewijs. Dat stelt ons gerust en geeft ons ten onrechte het gevoel dat we de wereld begrijpen, ook al doen we dat niet.

Veel Big Data-analyse kan geen oorzaken destilleren uit de analyse van effecten; vaak geeft ze geen aanwijzingen over het 'waarom', alleen over 'wat', datgene wat er gebeurt. Maar verrassend genoeg is dat vaak al voldoende om ons betere beslissingen te laten nemen. Het helpt Google razendsnel te vertalen van en naar tientallen talen en is verantwoordelijk voor ongeveer een derde van de verkoop van Amazon. Het zorgt er zelfs voor dat auto's zichzelf kunnen besturen.

Dat betekent natuurlijk niet dat onze levenslange zoektocht naar oorzaken ineens tot stilstand zou moeten komen. Het helpt ons wel ons te realiseren dat we moeten leren lopen voor we kunnen rennen, en met Big Data hebben we daar een krachtig instrument voor. Als ze op de juiste manier wordt toegepast, ontdekt Big Data-analyse het 'wat' en

vormt ze een buitengewoon waardevol filter voor de daaropvolgende diepgravende causale analyse.

Dit helpt ons in te zien dat de wereld waarin we leven veel complexer, maar ook veel interessanter en stimulerender is dan we dachten. Als we ons realiseren dat de data die we nu kunnen verzamelen zo veel latente waarde bevatten die we met Big Data kunnen blootleggen, zullen bedrijven hun manier van werken heroverwegen, en daarmee zullen ze onze economie hervormen en onze maatschappij verder ontwikkelen.

Met dit slimme en inzichtelijke boek, geschreven door experts die Big Data werkelijk begrijpen en beheersen, zult u Big Data en het gebruik daarvan voor de juiste doeleinden gaan waarderen. Het neemt u mee op een spannende reis die verrassende inzichten kan opleveren en veel waarde kan creëren. En, niet onbelangrijk, het laat u ook zien hoe de macht van Big Data gekanaliseerd kan worden, zodat we ervan kunnen profiteren zonder dat we te maken krijgen met de negatieve consequenties die het kan hebben.

Big Data is belangrijk – heel belangrijk zelfs. Zorg dat het dat ook voor u wordt!

Viktor Mayer-Schönberger
coauteur van De big data-revolutie

INLEIDING: ALLES KAN

Extreme verwachtingen zijn het nieuwe normaal

De iPad-generatie

YouTube.¹ We zien een klein meisje – zo jong dat ze alleen nog maar kreetjes uitstoot – dat een paar glossy magazines voor zich heeft liggen. Ze begint erdoorheen te bladeren en wil inzoomen of doorklikken. Het gebruikelijke *swipen*, *pinchen* en *tappen* – dat ze kennelijk op een tablet al volledig onder de knie heeft – blijkt niet te werken. Teleurgesteld gooit ze de magazines opzij. Dit filmpje blijkt een opname van haar vader, Jean-Louis Costanza, destijds CEO van Orange Vallée, die in de aftiteling kort en krachtig de boodschap neerzet: ‘Voor mijn dochter van 1 is een tijdschrift een iPad die niet werkt. Dat zal haar hele leven zo blijven. Steve Jobs heeft een deel van haar os [besturingssysteem, red.] geprogrammeerd.’

Nieuwe technologie heeft geweldig veel dingen mogelijk gemaakt. En nieuwe technologie is met ons leven verweven op een manier die we twintig jaar geleden nauwelijks konden dromen. Zo, die clichés zijn eruit.

Wees gerust, dit boek gaat nauwelijks over technologie.

We zouden wat data kunnen rondstrooien over de fenomenale reken capaciteit die er nodig was bij de succesvolle zoektocht naar het higgsboson – het ‘goddeeltje’ – bij CERN, de Europese organisatie die in Genève fundamenteel onderzoek doet naar elementaire deeltjes en daarmee in de Champions League van de data-analyse speelt. Sander was daar betrokken bij het Atlas-experiment, een van de experimenten waarmee de data van de deeltjesversneller worden geanalyseerd. We zouden kunnen uitleggen dat nieuwe open-sourceproducten de basis leggen onder veel baanbrekende nieuwe toepassingen van data. We zouden een poging kunnen doen om te ontrafelen hoe de algoritmen in elkaar steken die ervoor zorgen dat Google zulke goede antwoorden geeft op uw zoekvragen.

NIEUWE INFORMATIESAMENLEVING

Stuk voor stuk zijn het belangwekkende technologische onderwerpen. Toch gaan we er in dit boek maar beperkt op in. Liever praten we over de impact van die technologie op de maatschappij, en specifieker over hoe technologie die gerelateerd is aan data, dataverwerking, informatie en communicatie onze maatschappij kan verbeteren. Er ontstaat in rap tempo een nieuwe wereld, die we in dit boek aanduiden als de nieuwe informatiesamenleving. Het is een wereld waarin alles meetbaar is en waarin mensen en vrijwel alle denkbare apparaten 24/7 via internet met elkaar verbonden zijn. Dat netwerk van verbindingen en sensoren levert een fenomenale hoeveelheid data op en biedt fascinerende nieuwe mogelijkheden, die vaak onder de noemer Big Data worden geschaard. Het lijkt geen twijfel dat die mogelijkheden groot

zijn en dat ze zullen worden benut. Sterker nog: we zitten er al middenin, al hebben we dat misschien niet altijd in de gaten.

Dat besef grijpt om zich heen bij bedrijfsleven en overheid, die zich afvragen wat ze moeten met deze ontwikkeling. Het gaat hierbij verder dan het lanceren van wat nieuwe producten of diensten: het gaat om het inspelen op een radicaal andere wereld. In dat kader duikt de term *the age of disruption* de afgelopen jaren steeds vaker op. De term beschrijft de ontwikkeling dat snel opeenvolgende technologische doorbraken een ontregelend effect hebben op vrijwel alle onderdelen van de samenleving. Dat is zonder meer uitdagend, maar biedt ook enorme kansen.

Macro-economisch gezien is er eigenlijk reden voor een geweldig optimisme. We zitten midden in een klassiek economisch fenomeen, namelijk de creatieve destructie.* Dat betekent dat we pas tot vernieuwing kunnen komen als we het oude afbreken. En dat is wat er momenteel in vrijwel alle sectoren gebeurt. Big data fungeert als het groeihormoon voor deze creatieve destructie. Die groei valt niet te ontkennen: al ons geklets op sociale media, het streamen van muziek en video, onze e-mails, onze transacties in webshops et cetera leiden tot nauwelijks te bevatten hoeveelheden dataverkeer. En die databrij groeit in rap tempo nu het *internet of things* miljarden apparaten – tv's, koelkasten, beveiligingsapparatuur, thermostaten, rookmelders – aan elkaar knoopt, die allemaal data gaan produceren en die met elkaar gaan delen.

De verleiding is groot om hier met indrukwekkende cij-

* In de jaren vijftig van de vorige eeuw werd dit al beschreven door de Oostenrijkse econoom Joseph Schumpeter. Zie http://en.wikipedia.org/wiki/Creative_destruction.

fers te gaan smijten, maar dat is niet het doel van dit boek. Eén cijfer volstaat eigenlijk: er zijn in de afgelopen twee jaar tien keer zoveel gegevens vastgelegd als in de hele geschiedenis daarvoor.² Dat lijkt indrukwekkend, maar eigenlijk is wat je met die data kunt doen pas écht indrukwekkend. En daar zullen we in dit boek dan ook verder op ingaan.

De hoeveelheid data groeit dan wel exponentieel tot astronomische hoogte, toch is het woord ‘big’ in Big Data eigenlijk wat verwarrend. Want allerlei nieuwe toepassingen gaan helemaal niet om het bewerken of interpreteren van enorme hoeveelheden data – de zogenaamde *big messy data* –, maar veel meer om het *slim combineren* van data.

Neem Google Now. Deze dienst monitort verschillende data rondom een gebruiker – je internetverkeer, je mailverkeer, je sociale-mediagebruik, je agenda, je locatie en andere gegevens – en geeft je op basis daarvan realtime-adviezen. Het algoritme ‘weet’ bijvoorbeeld dat je vandaag om 16.00 uur in Zaltbommel moet zijn en laat je om 14.15 uur weten dat je wegens een kapotte bovenleiding aan het spoor maar beter wat eerder kunt vertrekken. Zo’n persoonlijke assistent maakt vooral gebruik van *small clean data* en loodst je wat gemakkelijker door het leven door data uit verschillende bronnen te combineren. Google Now doet dat niet op basis van enorme hoeveelheden data, maar op basis van een slim algoritme dat chocola kan maken van data uit verschillende bronnen.

BIG DATA IN DIT BOEK

We hebben ervoor gekozen om Big Data te gebruiken als een paraplubegrip. Big Data omvat alle nieuwe kansen, mo-

gelijkheden, bedreigingen en technieken die samenhangen met het feit dat we op een andere manier kunnen omgaan met data. De positieve en negatieve kanten van de ‘dataficatie’ van de samenleving dus, inclusief maatschappelijke elementen als privacy en sociale invloeden. Dat varieert van commerciële precisiebombardementen op klanten en verregaande monitoring van persoonlijke gedragingen door inlichtingendiensten tot en met het oplossen van files, het stoppen van epidemieën en het aansluiten van een koelkast of thermostaat op het internet. Een belangrijk onderdeel van Big Data is data-analyse: het zoeken naar patronen in (grote hoeveelheden) data om daar relevante conclusies uit te trekken.

RELATIE MET TECHNOLOGIE

Onze overtuiging is dat Big Data een enorme verrijking kan zijn voor de wereld. Die mooie kanten kunnen we (voor een deel) nu al elke dag ervaren, want technologie is aan alle kanten ons leven binnengedrongen. Iedereen herkent waarschijnlijk het beeld dat we overal ter wereld communiceren met onze vrienden en familie. We kunnen op de smartphone tekstberichten sturen, op onze tablet online games met hen spelen of via sociale media contact onderhouden.

We kunnen ook praktische zaken regelen door de zegeningen van de technologie. Zo kunnen we de bonnetjes van onze fysiotherapeut via internet indienen bij onze zorgverzekeraar. We slaan foto’s en video’s ergens op in de cloud – lekker handig als we ze willen delen – en willen ook de administratie van ons eigen bedrijf vooral niet op onze eigen harde schijf. Want online is wel zo gemakkelijk en beperkt

het risico dat we het kwijtraken. We regelen onze bankzaken met een appje. Willen we informatie over de plots opkomende ziektesymptomen bij onze cockerspaniël, dan hebben we die zo gevonden op internet. Ontstaat aan de bar van een café een discussie of acteur Peter O'Toole nu wel of niet ooit een Oscar won, dan kost het met een smartphone in de hand hooguit tien seconden om daarachter te komen.

Peter Hinssen zei het al: digitaal is het nieuwe normaal geworden in alles wat we doen. Digitaal leven doen we niet alleen volstrekt routineus, we zijn er ook aan verslaafd geraakt en onze verwachtingen over de mogelijkheden van dat digitale leven zijn tot grote hoogten opgestuwd.³

VERWACHTINGEN

Waar eindigt het? Hoe zal het digitale leven er in de toekomst uitzien? Gaan we massaal een Google Glass dragen – of een alternatief van concurrenten – als vervanging van onze smartphone? Zou kunnen. Gaat de werkgelegenheid de komende vijftig jaar structureel met de helft omlaag doordat steeds meer taken, zoals het schrijven van een boek als dit, door computerkracht worden overgenomen? Zou kunnen. Kunnen we onze gemiddelde levensduur met twintig jaar verlengen, onder andere omdat we de reken capaciteit hebben om ons menselijk genoom in kaart te brengen? Het is niet onwaarschijnlijk. Maar het eerlijke antwoord is dat niemand met zekerheid kan zeggen hoe het gebruik van technologie zich zal ontwikkelen.

Dat technologie – en daarmee ook Big Data – de komende jaren nog meer ons leven binnendringt, daar twijfelen we echter niet aan. Niet alleen omdat verzet ertegen zinloos is,

maar vooral omdat wij mensen het zelf willen en er zelf voor kiezen. Omdat we het gemak, het comfort of de toegevoegde waarde ervan niet meer willen missen.

De technologie schept verwachtingen, net als bij de dremes die verwacht een papieren magazine te kunnen gebruiken als een tablet. Alles kan; voor alles moet er een app zijn. De uptime van alle diensten moet 100 procent zijn. Bedrijven moeten maatwerk leveren en liefst tegen een kostprijs van nul. Alles moet bovendien ook nog leuk en duurzaam zijn.

De nieuwe uitdaging voor bedrijven, overheden en andere organisaties is dan ook simpel samen te vatten: *onmogelijkheden moeten direct worden gerealiseerd, op wonderen mag een levertijd zitten van maximaal twee weken.*

MOGELIJKHEDEN

Big Data is meer dan een nieuwe technologie of concept. Het biedt in alle denkbare sectoren en branches mogelijkheden om ons op een andere manier te organiseren, vooruitgang te boeken en dingen te doen die tot voor kort simpelweg niet mogelijk waren. Veel mensen denken bij de term Big Data direct aan bedrijven die nog meer aan hun klanten willen verkopen door alles over hen te weten te komen. Dat is echter maar één kant van de medaille; een kant waar we in dit boek zeker ook op terugkomen. We willen ook laten zien dat Big Data maatschappelijke problemen op kan lossen en ons leven kan veraangenamen. We kunnen er mensenlevens mee redden. We kunnen onderhoud beter en goedkoper plannen. We kunnen de verkeersveiligheid verhogen. We kunnen de opbrengsten in de landbouw verhogen. We kunnen hulpbe-

hoevende ouderen een beter leven bieden. We kunnen de wereld verduurzamen. We kunnen een revolutie ontketenen in medische diagnoses en behandelingen. We kunnen de betaalbaarheid verhogen. We kunnen de sportieve prestaties van een team verbeteren. Of we kunnen de filedruk verminderen.

OPBOUW

We beginnen het boek met een overzicht van wat Big Data ons kan brengen. Daarbij gaan we uitgebreid in op zowel de voordelen als de nadelen. We zullen laten zien dat we als burgers en consumenten extreme verwachtingen hebben, maar dat we tegelijkertijd weerstand voelen tegen de datahonger van organisaties en vrezen voor onze privacy. We komen daarbij als vanzelf tot de conclusie dat we Big Data niet in isolatione kunnen behandelen, omdat er sprake is van een zeer fundamentele wisselwerking met maatschappelijke ontwikkelingen.

Daarom gaan we in het tweede deel van het boek in op de nieuwe informatiesamenleving die aan het ontstaan is en waar Big Data een essentieel onderdeel van is. We laten zien hoe burgers gaan eisen dat organisaties maatschappelijk de juiste dingen doen en niet alleen gaan voor eigen gewin; dat we als collectief van burgers steeds machtiger worden en daarmee het doen en laten van organisaties ook sterk kunnen beïnvloeden; en dat we in die uitdagende omgeving een manier moeten vinden om Big Data op de juiste manier in te bedden in onze samenleving.

De verbreding van ons blikveld naar de informatiesamenleving leidt uiteindelijk in het derde deel van het boek

tot een discussie over de manier waarop Big Data succesvol kan worden ingebed in die samenleving. Zaken die aan bod komen zijn onder meer de ethische uitgangspunten, het onderwijs en de rol van de overheid.

Big Data kan – op een beheersbare manier – bijdragen aan een betere samenleving. En dat is ook precies onze drijfveer om dit boek te schrijven. We willen laten zien dat bedrijven en overheden nu nog niet goed weten hoe ze met Big Data moeten omgaan. Er is onbegrip, onkunde en onwetendheid. Big Data-projecten zijn vaak fragmentarisch en ongestructureerd. Dat alles leidt tot veel teleurstelling, tot incidenten en tot weerstand. Dat moet anders. Dat kan anders. Wij proberen met dit boek een idee te geven van de manier waarop het anders kan. In het volgende hoofdstuk zullen we eerst eens bekijken welke voordelen Big Data ons te bieden heeft.

