

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

info@boomamsterdam.nl
www.boomuitgeversamsterdam.nl

Big Data: kleine data worden groot

Wat is dat Big Data eigenlijk en waarom doet iedereen er zo opgewonden over? Om de ontwikkelingen die nu onder de noemer 'Big Data' op ons afkomen goed te kunnen plaatsen, is het nuttig terug te kijken op de ontwikkeling van elektronische informatiedeling op het web, en om vooruit te kijken naar wat web 4.0 genoemd kan worden. De vragen die daarbij centraal staan, zijn niet primair technologisch of juridisch. Cruciaal is of gebruikers de weerbarstige mogelijkheden van Big Data tot verstandige volwassen praktijken kunnen omvormen.

door *Freek Bomhof*

De auteur is werkzaam bij TNO als programmamanager van het Big Data-onderzoek.

HET INTERNET IS EEN VERZAMELING van allerlei technologieën die het mogelijk maken om informatie op computers gemakkelijk te verspreiden – een soort basisinfrastructuur waar veel verschillende dingen mee gedaan kunnen worden. De toepassing die in de jaren negentig het bekendst geworden is, is het world wide web: een manier om documenten die op computers staan gemakkelijk voor de hele wereld inzichtelijk te maken. Het world wide web is door Tim Berners-Lee op CERN ontwikkeld en was in eerste instantie vooral in academische en technische kringen bekend. Pas midden jaren negentig werd een computeraansluiting ook voor consumenten haalbaar, waren zogenoemde 'grafische browsers' (waarvan *Netscape* een van de eerste was) gemeengoed, én waren er inmiddels handige diensten, zoals zoekmachines. Wie medio jaren negentig zelf een website had, hoorde bij een technische voorhoede: internettechnologie was

technologiegedreven. Achteraf kunnen we deze fase web 1.0 noemen: het *web of information*.¹ Het is nu amper meer voor te stellen, maar voor eind 1997 gebruikten we geen Google, maar Yahoo, Ilse of AltaVista: het icoon van web 1.0 is vrij laat in beeld gekomen.

Vanaf circa 2000 nam de ontwikkeling van sociale media een grote vlucht: denk aan LinkedIn, Hyves, Facebook. Het gebruik van het web werd gemeengoed en vooral niet-technische gebruikers bepaalden in hoge mate wat er gebeurde. Dit wordt ook wel web 2.0 genoemd: het *web of people*. Opmerkelijk is dat de technologische vernieuwing in deze fase veel minder ingrijpend was: het was vooral gebruikmaken van de technieken van de jaren negentig, maar dan sneller, mooier en vooral toegankelijker. Web 2.0 is vooral aangedreven door de gebruikers: ook ingrijpende vernieuwingen, zoals breedband en smartphones, deden zich in deze periode vooral door de enorme marktvraag voor.

Vanaf ongeveer 2010 krijgen technologische ontwikkelingen weer meer aandacht. Eén ervan is het *Internet of Things*: de gedachte dat ook 'dingen' online kunnen zijn en gegevens kunnen delen met hun omgeving. Niet alleen de spreekwoordelijke koelkast (die volgens sommigen nog steeds zelf de supermarkt zou moeten informeren over het feit dat de melk op is), maar ook de thermostaat, beveiligingscamera's, auto's, sluizen, parkeermeters en verkeerslichten. Sensoren worden goedkoper en kleiner, en gevoegd bij de goedkopere en alomtegenwoordige (draadloze) verbindingen leidt dat tot een stortvloed aan gegevens: Big Data. Een minder in het oog springende ontwikkeling is het 'semantisch web': een verzameling technologieën die computers in staat stelt om niet alleen 'dom' met gegevens zelf te werken maar ook met de betekenis en onderlinge samenhang ervan. Dit alles kan ook wel 'web 3.0' genoemd worden: het *web of data*. Uit de beschrijving ervan blijkt wel dat deze fase weer vooral technologiegedreven is.

Als we nu deze ontwikkeling doortrekken, is het niet vreemd om te veronderstellen dat het huidige technologisch gedreven web 3.0 gevolgd gaat worden door een web 4.0 waarin individuen, organisaties en de maatschappij de nieuwe mogelijkheden omarmen op een manier die we ons nu nog niet helemaal kunnen voorstellen. (Als de mooie indeling in decennia zich doorzet, zou dit rond 2020 duidelijk moeten gaan worden.)

Eind jaren negentig werden er congressen, boeken en cursussen aan 'virtueel winkelen' gewijd: men zag de mogelijkheden, maar hoe die effectief te benutten was de grote vraag. Inmiddels is het starten van een webshop met een uurtje muisklikken wel geregeld en vragen we ons af waarom we daar twintig jaar geleden zo ingewikkeld over deden. Natuurlijk, de technologie heeft het allemaal veel gemakkelijker gemaakt, maar er zijn ook ontwikkelingen geweest – zoals iDeal (veilig en makkelijk betalen),

het keurmerk Thuiswinkel Waarborg en het pionierswerk van Amazon, Marktplaats en bol.com – die dit vroeger zo moeizame fenomeen tot iets heel gewoons hebben gemaakt. En zoals er eind jaren negentig grote (maar niet-technische) ingewikkeldheden rondom het internet hingen, is dat nu rondom Big Data, en eigenlijk het hele web 3.0, ook weer het geval.

DE TECHNOLOGIEGEDREVEN MOGELIJKHEDEN VAN WEB 3.0

Net zoals de eerste internettoepassingen vooral gericht waren op efficiency, zijn veel toepassingen van Big Data ook te vinden in nogal ‘technologisch-industriële’ omgevingen. Zo is een van de eerste Big Data-toepas-

Net zoals de eerste internettoepassingen vooral gericht waren op efficiency, zijn veel eerste toepassingen van Big Data ook te vinden in ‘technologisch-industriële’ omgevingen

singen binnen het TNO-onderzoek gevormd door het IJkdijkprogramma. De gedachte die bij deze toepassing is uitgewerkt, komt erop neer dat het nuttig is om zo precies mogelijk de sterkte van een dijk te kunnen bepalen. Immers, een dijkdoorbraak leidt tot zeer kostbare gevolgschade die je wilt voorkomen, bijvoorbeeld door preventief onder-

houd te plegen. Als daarentegen een dijk nog in prima staat verkeert, kan het regulier geplande onderhoud uitgesteld worden, waarmee ook geld bespaard wordt. In een reeks experimenten heeft TNO met een groot aantal partners aangetoond dat de data die uit grote hoeveelheden metingen op basis van allerhande sensoren afkomstig is, tot nauwkeurige modellen kan leiden om de sterkte van een dijk goed te kunnen voorspellen.

Deze gedachte is op meer plaatsen toegepast: de sterkte van bruggen, de betrouwbaarheid van ondergrondse gasleidingen, en recent nog de sterkte van huizen in Groningen in de gaswinningsgebieden. In feite kan het bij alle kostbare infrastructuren nuttig zijn om zo goed mogelijk in te schatten of er al dan niet onderhoud nodig is.

Kenniswerkers kunnen tot op zekere hoogte óók als zo’n kostbare infrastructuur gezien worden: de Nederlandse kenniseconomie. En inderdaad loopt er een verkennend onderzoek om na te gaan of op basis van allerhande sensorgegevens kan worden voorspeld of iemand risico loopt om binnenkort burn-outklachten te ontwikkelen.² Dat blijkt overigens vrij goed te kunnen. Er wordt gebruikgemaakt van sensoren die iets zeggen over de fysieke leefstijl van de kenniswerker (bijvoorbeeld de hoeveelheid beweging die hij krijgt), maar ook van de tijdstippen waarop mail verstuurd wordt en van de toon waarin die berichten dan gesteld zijn. Hoe nuttig dat

ook kan zijn, dit roept natuurlijk direct vragen op: van wie zijn die gegevens eigenlijk? Wie mag ze inzien? Wat mag ermee gebeuren? En het roept ook vragen op over de wenselijkheid van zo'n mogelijkheid. Kan een werkgever bijvoorbeeld zijn verantwoordelijkheid voor de gezondheid van zijn mensen van zich afschuiven omdat hij hun een app ter beschikking heeft gesteld waarmee zij zelf kunnen zien of ze gevaar lopen?

Dit voorbeeld is dan nog een toepassing die redelijk rechttoe rechtaan is. De steeds bredere beschikbaarheid van data, de aanwezigheid van sensoren en de ontwikkeling van slimme manieren om dat allemaal te combineren, leiden tot onverwachte relaties. Israëliische onderzoekers hebben gevonden dat de straalverbindingen tussen de masten van het mobiele telefoonnetwerk bij nauwkeurige analyse gegevens over neerslag blijken te bevatten: een buienradar op basis van het gsm-netwerk.³ In Afrika is dat een nuttige toepassing. Ander hergebruik van gegevens is veel enger: het blijkt mogelijk om uit camerabeelden het geluid te destilleren dat op dat beeld aanwezig is.⁴ Een gesprek afluisteren door 'gewoon' met een gevoelige camera een zak chips te filmen blijkt mogelijk. Dat roept nog veel meer vragen op: van wie is de data eigenlijk? Mag je ook alles waarnemen, opslaan en analyseren wat je kúnt waarnemen?

De antwoorden op deze vragen zullen de komende jaren, in de aanloop naar web 4.0, gegeven gaan worden. Dat de gedachten daarover nogal grillig kunnen uitpakken, illustreert het volgende voorbeeld.

Toen de slimme meter geïntroduceerd werd, kwam er al snel een discussie op gang over de privacyaspecten daarvan: 'ze' zouden aan het energieverbruik bijvoorbeeld kunnen zien of de bewoners afwezig of zelfs op vakantie waren. Dit konden 'ze' natuurlijk al jaren zien aan pinbetalingen, reisgegevens en locatiegegevens van de mobiele telefoon, maar de slimme meter was duidelijk de druppel die de emmer deed overlopen.

In Duitsland is daarentegen de slimme meter met veel minder discussie geaccepteerd. Een van de redenen die een Duitser mij daar ooit eens voor gaf was dat de slimme meter in Duitsland voor veel mensen ook een soort geldboompje is: lange tijd was het in Duitsland wettelijk zo dat teruggeleverde energie voor een hoge prijs werd afgerekend, en voor het terugleveren van energie (uit zonnepanelen of windmolens) is zo'n slimme meter noodzakelijk.

Vergelijk dit met de manier waarop de dienst Google Street View ontvangen is. In Duitsland ontstond hierover een vrij sterk debat, wat ertoe leidde dat Google verplicht werd om foto's van woonhuizen te vervagen (blurren) als bewoners daarom vroegen; inmiddels heeft een kwart miljoen Duitsers dit gedaan. In Nederland is de ontvangst van Google Street View veel minder kritisch geweest.

We zien dus dat de ontvangst van dezelfde dienst in min of meer verge-

lijkbare landen diametraal van elkaar kan verschillen. Het zou goed zijn om na te denken over hoe we willen omgaan met dit soort ontwikkelingen die onze informatiesamenleving vormgeven.

HOE WILLEN WIJ DAT WEB 4.0 ERUIT GAAT ZIEN?

We zullen de komende jaren met elkaar gaan bepalen hoe web 4.0 eruit gaat zien. De technici zijn daar vandaag de dag mee bezig. Juristen zijn er ook mee bezig, maar worstelen met de snelheid van de technologische ontwikkelingen: een wet moet relevant blijven, ook als de onderliggende technologie verandert. Dat leidt soms tot het achter de feiten aan rennen, zoals bijvoorbeeld ‘the right to be forgotten’⁵ pas nu voor het eerst enige praktische betekenis begint te krijgen, terwijl het probleem eigenlijk al met de komst van zoekmachines, maar zéker met de opkomst van sociale media actueel werd.

Maar: de belangrijkste discussie over wat we met web 4.0 willen, zal niet technologisch zijn en ook niet juridisch. Die discussie wordt misschien ook helemaal niet expliciet gevoerd, maar vormt zich door het dagelijks gebruik van informatietechnologie: het gedrag van de internetgebruiker. Daarmee is dit eerder een ethische discussie: sommige dingen dóe je gewoon niet.

Basisschoolkinderen wordt tegenwoordig geleerd hoe om te gaan met informatie op het internet: niet alles wat je met Google vindt, is betrouwbaar. Je wachtwoorden geef je niet aan anderen. Je maakt niet zomaar geld over naar een onbekende webshop. Het gaat helaas nog geregeld mis, maar hoe je op internet om zou moeten gaan met het vertrouwen van anderen is inmiddels toch redelijk gemeengoed geworden.

De nieuwe mogelijkheden van web 4.0 zullen echter veel verder gaan dan vertrouwen. Het gaat over de vraag of een verzekeraar (of werkgever) iemands persoonlijke data mag gebruiken voor een risico-inschatting. Of mensen hun persoonlijke gegevens voor zichzelf mogen houden. Op welke manier een dienstverlener ter verantwoording kan worden geroepen als een beslissing op basis van een Big Data-computeranalyse fout blijkt te zijn. Wie aansprakelijk is als zelfrijdende auto’s (die zeer data-intensief zijn) ongelukken veroorzaken. Hoe ver politie kan gaan bij het op basis van Big Data-analyses voorkomen van misdaden (*predictive policing*).

Eén ding is duidelijk: als iets technologisch mogelijk is, zal het gemaakt worden ook. De vraag is dan niet *of*, maar *hoe* we willen gebruikmaken van die mogelijkheden.⁶

NIET VERBIEDEN, MAAR BIJSTUREN

Van nieuw ontwikkelde technologie is zelden in één klap duidelijk wat we ermee aan moeten. De grammofoon werd door Edison gezien als een uitstekende manier om de laatste woorden van mensen op hun sterfbed vast

*Van nieuw ontwikkelde
technologie is zelden in één
klap duidelijk wat we ermee
aan moeten*

te leggen, om de tijd weer te geven, of om blinden de gelegenheid te geven naar boeken te luisteren. Muziek vastleggen en verspreiden voegde hij onder aan zijn lijstje toe; dat zou ook kunnen – tegenwoordig zouden we juist die toepassing de ‘killer app’ van de grammofoon

noemen. Het is dus zoeken naar de manier waarop we er het meest voordeel van hebben, en daarbij vermijden dat de nadelen de overhand krijgen.

De ontwikkelde technologie zou een aantal kenmerken moeten hebben: de nieuwe technologie zou transparant moeten zijn zodat duidelijk is waar ze vandaan komt en hoe ze werkt; ze zou mensen moeten verbinden in plaats van isoleren; de productie, het beheer en het eigendom ervan zouden decentraal moeten zijn; men moet de keus hebben om de technologie *niet* te gebruiken, of op zijn minst de mogelijkheid om deze aan te passen; er moeten dus ook alternatieven voor zijn; en ze moet een minimale impact op ecosystemen hebben.⁷

Is dit alleen iets wat relevant is voor technologen? Nee, zeker niet. Introductie van nieuwe technologie vergt anticipatie (hoe kan ze gebruikt gaan worden?), continue beoordeling (verandert er iets in een verkeerde richting?), het inschatten van risico's (wat kan er misgaan?), het snel herstellen van schade (ingrijpen waar nodig) en ombuigen in plaats van verbieden. Dat is een taak voor ontwerpers, wetgevers, beleidsmakers, financiers en dergelijke, maar ook voor gebruikers, burgers, belangengroeperingen; zoals Peter-Paul Verbeek ons voorhoudt beïnvloeden techniek en moraal elkaar wederzijds.⁸

TECHNIEK BEÏNVLOEDT ONZE NORMEN

Technologische ontwikkelingen kunnen dus niet sec bekeken worden en langs de meetlat gelegd worden van wat wij acceptabel vinden; die meetlat wordt namelijk door de technologie beïnvloed. Zoals prenatale diagnostiek het debat over abortus heeft beïnvloed, zo zal een datagedreven innovatie als de zelfrijdende auto het debat over verkeersveiligheid beïnvloe-

den. In de voorbeelden hierna zal ik vooral naar persoonlijke data kijken: dat staat momenteel het meest in de belangstelling.

Een voorbeeld uit web 1.0/web 2.0: privacy vinden we zeer belangrijk en iedereen heeft het recht om zijn persoonlijke gegevens buiten de openbare sfeer te houden. In de zakelijke wereld echter is het hebben van een (publiek) LinkedInprofiel inmiddels dermate normaal geworden dat het daar op zijn minst een lichte achterdocht wekt als een zakenpartner *niet* op LinkedIn te vinden is.

Web 3.0, met al zijn meet- en analysemogelijkheden, kan ook aanleiding geven tot zulke veranderingen in houding. Neem het eerder genoemde onderzoek waarin werd nagegaan of kenniswerkers risico lopen om binnenkort een burn-out te ontwikkelen. Een eerste belangrijke keus is natuurlijk wie eigenaar van de data is; de consensus lijkt te gaan worden dat die data van de werknemer zelf is en blijft, inclusief de uitkomst van een analyse. Werkgevers zouden zo'n burn-outrisicoapp echter wel aan kunnen bieden aan hun werknemers. Deelname is natuurlijk vrijwillig, maar hoe zou de werkgever reageren als iemand zich dan *tóch* met burn-outklachten ziek meldt? De middelen die aangereikt worden houden een impliciete opdracht in: 'Maak er gebruik van! Het is in je eigen belang!' Zo bekeken zou zelfs een verplichting een optie kunnen zijn: het gebruik van fysieke beveiligingsmiddelen kan immers ook door werkgevers verplicht gesteld worden, en de overheid doet hetzelfde met bijvoorbeeld veiligheidsgordels. De norm kan dus worden: als er mogelijkheden zijn om je eigen gezondheid in de gaten te houden, ben je eigenlijk (moreel of wettelijk) verplicht om dat ook te doen.

BIJSTUREN DOOR KEUZES TE MAKEN EN DOOR BIJ TE LEREN

Zoals eerder gesteld: niet verbieden, maar ombuigen. Wat voor soort keuzes zijn dat dan? Hierna volgen enkele voorbeelden van keuzes die de overheid kan maken voor de manier waarop zij Big Data inzet.⁹

Ten eerste de keus op welke waarden nadruk gelegd wordt: privacy & autonomie aan de ene kant, versus effectiviteit, efficiency & veiligheid aan de andere kant. Als er een tegenstelling is (dat hoeft niet altijd), wat geeft dan de doorslag?

Ten tweede de keus van de mate van toegankelijkheid van data: een lage mate van toegang (dus data alléén gebruiken voor het doel waarvoor deze verzameld is, en deze ook niet makkelijk uitwisselbaar laten zijn), versus een hoge mate van toegang (data kan en mag vrij gedeeld en gekoppeld worden).

Door deze keuzes maximaal door te trekken en uit te werken in toekomstscenario's is nagegaan hoe dat dan uit kan pakken. Bij de toetsing (welke kant bewegen we nu op?) blijkt dat er momenteel helemaal niet zulke duidelijke

keuzes gemaakt worden. Soms ligt de nadruk op privacy, soms op efficiency. Soms mag data helemaal niet gekoppeld en gedeeld worden, soms verwacht de burger dat juist. Kennelijk is hier nog geen duidelijke visie op. Het is goed ons te realiseren dat slecht doordachte of impliciet gemaakte keuzes hier tot ongewenste resultaten kunnen leiden: het zal niet veel moeite kosten om een voorstelling te maken van doemscenario's die een combinatie van *1984* en *Minority Report* zijn; of waarin datamonopolisten alle macht hebben gekregen; of waarin er nieuwe tweedelingen ontstaan tussen haves en havenots.

De overheid is natuurlijk niet de enige die keuzes maakt – bedrijven en burgers doen dat ook. Maar de overheid maakt wel wetten waar de maatschappij mee te maken heeft en die soms anders uitpakken dan bedoeld: de cookiewet bijvoorbeeld leidt vooral tot ergernis bij degenen wiens privacy verondersteld werd erdoor beschermd te worden.

De komende tijd zullen we ook als maatschappij duidelijker keuzes moeten gaan maken over hoe we met persoonlijke data om wensen te gaan. Uit het voorbeeld van Google Street View versus de slimme meter in Duitsland en Nederland blijkt al dat we onze normen momenteel nog aan het vaststellen zijn. Om die normen te kunnen bepalen, moeten we wel weten waar we het over hebben. Techniek, of deze nu door overheid of bedrijfsleven gebruikt wordt, zal daarom transparant moeten zijn. Transparantie is echter wel een noodzakelijke, maar geen voldoende voorwaarde.¹⁰ We hoeven heus niet allemaal een universitaire graad *computer science* te halen, maar we moeten wel een basisbegrip ontwikkelen van wat data en dataverwerking eigenlijk zijn. Dat is: dingen uitproberen, voorbeelden verzamelen, publiek debat, onderwijs. Net zoals we sinds web 1.0 op basisscholen leren dat Google niet altijd gelijk heeft en dat Wikipedia mooi is maar dat je je oordeel op meer dan één bron moet baseren, zo zullen we na web 3.0 allemaal een beetje meer verstand moeten krijgen van de data om ons heen.

Noten

- 1 <http://www.frankwatching.com/archive/2009/01/05/web-webber-webstvan-10-naar-30-en-verder/>
- 2 <http://www.swell-project.net/>
- 3 <http://tweakers.net/nieuws/42385/gsm-masten-registreren-neerslag.html>
- 4 <http://www.nu.nl/tech/3844794/mit-reconstrueert-geluid-via-trillingen-van-zak-chips.html>
- 5 http://en.wikipedia.org/wiki/Right_to_be_forgotten
- 6 Evgeny Morozov gaat in zijn boek *To Save Everything, Click Here. The Folly of Technological Solutionism* (Perseus,

- 2013) overigens ongemeen fel tegen deze gedachte in en vindt dat we niet moeten proberen om elk probleem van een technologische oplossing te voorzien. Hij brengt waardevolle punten in, maar verzuimt aan te geven hoe we hier in de praktijk mee om zouden moeten gaan.
- 7 Kevin Kelly, *What Technology Wants*. Londen: Penguin, 2011.
- 8 Peter-Paul Verbeek, *Op de vleugels van Icarus*. Rotterdam: Lemniscaat, 2014.
- 9 A. Madho, *Big data. Gradueel of kwantitatief?* (masterthesis Public Information Management). Den Haag: HEC, 2014.
- 10 <http://www.bigdata-startups.com/transparency-in-big-data-is-not-enough/>