

Big data: hype of trend?

Citeersuggestie: A.P. Engelfriet, Big data: hype of trend? Tijdschrift voor Internetrecht 2014.

Big data, wat betekent het juridisch? Je hoeft maar een juridisch tijdschrift of cursusprogramma open te slaan om deze zin aan te treffen. Big data is happening, en we moeten er wat mee als juristen zijnde. Maar is big data nu werkelijk zo uniek?

Hoeveelheid, snelheid en diversiteit

De term *big data* wordt gebruikt om te verwijzen naar verzamelingen data die zo groot zijn dat ze met traditionele database-systemen niet goed meer te verwerken zijn¹. De snelle groei van verwerkings- en opslagcapaciteit van ICT-systemen maken het mogelijk dat hoeveelheden data worden verwerkt waar vroeger alleen van kon worden gedroomd. En bij zulke grote hoeveelheden data speelt ook nog eens het probleem dat de data zelden netjes gestructureerd of gelabeld is, zodat zoeken, combineren en analyseren van die data minstens zo'n uitdaging is als ze bij elkaar krijgen.

Volgens onderzoeks- en adviesbureau Gartner gaat het bij big data om drie factoren: de hoeveelheid data, de snelheid waarmee de data binnenkomt en opgevraagd wordt en de diversiteit van de data. Ook de complexiteit van de analyse is van belang. Het krantenarchief van de Koninklijke Bibliotheek is in die definitie geen big data, aangezien hier niet vaak complexe zoekopdrachten in moeten worden uitgevoerd. De verzameling met "Vind ik leuk's" die Facebook aanlegt, is wél big data, aangezien het bedrijf hieruit in real time moet achterhalen wat 'men' interessant vindt en er ook nog eens welke advertenties daar het beste bij passen.

Met big data kan worden gezocht naar combinaties of trends die in kleinere hoeveelheden data onzichtbaar zouden zijn gebleven. Met statistische analyses door krachtige computers kunnen dan verbanden worden gevonden die handmatig onderzoek nooit had kunnen vinden. Correleer tienduizend factoren over een periode van vijf jaar en je zult opmerkelijke dingen vinden: een patroon van frauduleuze transacties in de boekhouding, de kans dat een ontslagen patiënt op korte termijn terugkomt in het ziekenhuis of het tijdstip waarop een machine op de productievloer het beste vervangen kan worden.

Vind ik leuk

Een bekend voorbeeld van wat er zoal mogelijk is met big data is de Amerikaanse winkelketen Target die middels big data-analyse aan het koopgedrag van haar klanten kon achterhalen of deze vrouw waren – en zwanger. Zo blijkt het omschakelen naar geurloze lotions en de aanschaf van extra schoonmaakmiddelen opmerkelijk vaak door zwangere vrouwen te gebeuren. Toen een zestienjarige dame vervolgens een boekje kortingsbonnen voor zwangerschaps- en babyproducten ontving, verraste dat haar ouders nogal. Maar de computer van Target bleek het bij het rechte eind te hebben, en de boze vader bood dan ook zijn excuses aan voor zijn boze protest en de eerdere stap naar de media.²

Dit voorbeeld laat meteen zien waar het grootste juridische pijnpunt zit: veel big data die vandaag de dag wordt verzameld, bevat persoonsgegevens. En dat maakt het verwerken van big data lastig, want het verwerken van persoonsgegevens is aan strenge regels gebonden. De Wet bescherming persoonsgegevens eist in beginsel uitdrukkelijke toestemming van de betrokkene, alsook een informatieplicht voor de verantwoordelijke die de gegevens gaat verwerken. Die toestemming is doelgebonden – toestemming voor gebruik in een bestelling betekent nog niet toestemming voor uitgebreid analyseren in een big data-bestand. Weliswaar kent de Wbp een uitzondering voor historische, statistische of wetenschappelijke doeleinden (art. 9 lid 3) maar het doel van big data onderzoek is zelden onder dit kopje te rangschikken. Het gaat immers om het ontdekken van nieuwe commerciële mogelijkheden: welke productcombinaties kunnen we introduceren, wat werkt voor onze klanten en hoe kunnen we ze nog meer extraatjes verkopen.

De uitgangspunten van de Wbp zijn op zich overzichtelijk en helder. Alleen, ze zijn geschreven in een wereld waarin databanken kilobytes aan data bevatten – een klantenbestand, een jaaradministratie, een logboek of wat. Geen zettabyte – een triljard gigabyte – aan ongestructureerde data waar een gemiddeld databasepakket zich per direct verslikt.

¹ Manyika, James, et al. "Big data: The next frontier for innovation, competition, and productivity." McKinsey Insights, mei 2011.

² Hill, K. "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did", Forbes magazine 16 februari 2012.

Niet alle big data bevat overigens persoonsgegevens: de deeltjesversneller *Large Hadron Collider* in Genève verzamelt bijvoorbeeld 40 miljoen sensormetingen per seconde in de zoektocht naar nieuwe elementaire deeltjes. Veel *bigger* dan dat wordt het niet, en er moet ook snel kunnen worden gezocht naar statistische afwijkingen die kunnen wijzen op natuurkundige noviteiten. Big data dus, maar het zal niet meevallen hier juridische vraagstukken bij te bedenken.³

Een aandachtspunt bij big data zal wel altijd de *service level agreement* zijn: hoe snel worden de data verwerkt, met welke snelheid wordt er gereageerd op problemen, en hoe vaak wordt er gebackupt. Dat laatste is per definitie niet triviaal: hoeveel harde schijven heeft u die een zettabyte aan data aankunnen?

22 miljard Vind-ik-leuks

Veel big data bevat dus persoonsgegevens. En dat maakt werken met big data lastig: het verwerken van persoonsgegevens is aan strenge regels onderworpen. Het begint al bij de eenvoudige vraag hoe men toestemming krijgt van alle betrokkenen. Waar dat bij een aanmelding op een website nog wel lukt, is het bij 4.5 miljard Vind-ik-leuks per dag⁴ al iets moeilijker om per geval op een zinvolle manier toestemming te vragen. En hoe aan de informatieplicht moet worden voldaan, is ook een vraag waar een aflevering van dit tijdschrift kan worden gevuld.

Hier wreekt zich het verschil tussen Europese en Amerikaanse privacy. Het Amerikaanse recht kent het concept 'persoonsgegeven' zoals wij dat kennen niet.⁵ Hoewel ook in de VS privacy een grondrecht is, wordt dit uitsluitend gerelateerd aan de relationele privacy – *the right to be let alone*. Informatieprivacy – zeggenschap over wat men wéét over jezelf – en dan met name de invulling die de zeggenschap over persoonsgegevens bij de betrokkene zelf legt, is een Europeesrechtelijk concept.

Ook in de VS wordt wel om toestemming gevraagd voor gebruik van persoonsgegevens, maar dat is eigenlijk altijd vanuit het perspectief van mensen met

rust laten. Een nieuwsbrief dringt hinderlijk binnen in je inbox, dat raakt aan de relationele privacy. Daarbij komt dat in de VS de beleidskeuze medio jaren negentig was dat zelfregulering door het bedrijfsleven de beste resultaten zou geven.⁶

Vogelvrije persoonsgegevens

Persoonsgegevens die niet gekoppeld zijn aan een naam, adres of ander contactgegeven worden in de Amerikaanse visie op privacy al snel als vogelvrij gezien. Daar heb je als privépersoon immers geen last van, niemand kan je contacteren dus je wordt keurig met rust gelaten.⁷ En dat je het alsnog vervelend vindt dat men weet wat jouw online alias Henkiepenkie123 allemaal uitspookt, tsja dan had je die gegevens maar niet moeten achterlaten op sites als Google of Facebook. Dat was jouw keuze en jij bent akkoord gegaan met de *terms of use* en de *privacy policy* van het bedrijf.

Big data is vrijwel nooit gekoppeld aan contactgegevens. Het gáát bij big data ook helemaal niet om individuele personen. Dat ik op Facebook zit, is niet interessant. Dat er 43.548 Nederlanders zijn in mijn inkomstencategorie die de advocatenserie *Suits* kijken, dat is wel interessant want daar kunnen dan specifieke advertenties bij neergezet worden.

Natuurlijk is er altijd wel iets nodig om die 43.548 mensen te herkennen en te onderscheiden. Hiervoor zijn technieken als cookies of IP-adressen zeer geschikt. Een cookie bevat een uniek getal dat aan een big data database-record te koppelen is, en hetzelfde kan worden gedaan met een IP-adres. Dit verklaart de heftige weerstand vanuit Amerikaanse bedrijven wanneer dergelijke getallen door Europese privacytoezichthouders ineens tot persoonsgegevens worden verklaard:⁸ daar gaat hun mooie advertentieverkoopbedrijfsmodel. En zonder gerichte advertenties geen inkomsten. Uit vele onderzoeken blijkt dat 'gewone' advertenties veel minder succesvol

³ Een onrechtmatige-daadsactie wegens het mogelijk veroorzaken van het einde van de wereld bleek in 2008 geen haalbare kaart. Hoewel niet uit te sluiten viel dat de LHC een zwart gat zou kunnen veroorzaken waarin de gehele aardbol verzwolgen zou worden, weigerde de Amerikaanse rechter de zaak met een beroep op gebrek aan jurisdictie. Johnson, A. "The Black Hole Case: The Injunction Against the End of the World". *Tenn. L. Rev.* 76 (819): 5480. arXiv:0912.5480

⁴ Facebook annual report 2013.

⁵ Kamaal Zaidi, *Harmonizing U.S.-EU Online Privacy Law: Toward a U.S. Comprehensive Regime For the Protection of Personal Data*, 12 *Mich.St. J. Int'l L.* 169 (2003).

⁶ William J. Clinton & Albert Gore, Jr., *A Framework for Global Electronic Commerce*, July 1, 1997.

⁷ Vgl. Newell, B. 'Rethinking Reasonable Expectations of Privacy in Online Social Networks', XVII *RICH. J.L. & TECH.* 12 (2011).

⁸ *Opinion 4/2007 on the concept of personal data*, Article 29 werkgroep WP 136, 2007.

zijn dan gerichte, op bezoekersprofielen afgestemde advertenties.⁹

Nog lastiger wordt het als deze verwerking leidt tot beslissingen over de betrokken personen. Via big data analyse zijn potentiële wanbetalers snel te herkennen en te weren in de webwinkel. Mag zoiets van de privacywet?¹⁰ Is het wel wenselijk, dat een grote computer op basis van een zetabyte aan data en een vaag algoritme dat werkt met waarschijnlijkheden concludeert dat ik iets niet mag, of juist dat ik iets juist wel zou willen?

De oplossing is een database

Big data is geen hype maar een trend. Binnen de ICT geldt immers: ongeacht het probleem, de oplossing is een database. Hoe meer data, hoe beter. En vanuit zakelijk perspectief is dat best te begrijpen. Maar vanuit privacyrechtelijk perspectief is big data een big threat: er worden zéér uitgebreide informatieprofielen opgebouwd over mensen, en hiermee worden volautomatisch conclusies getrokken en acties ondernomen die die mensen raken.

Big data aan banden leggen zal echter niet eenvoudig zijn. Het gevecht om de Privacyverordening is een eerste voorbeeld; het 'heftigste lobbycircus ooit' ging in feite om de vraag of internetbedrijven met Big Data Europese computergebruikers mogen registreren, volgen en verkopen aan adverteerders. Dit tij keren zal niet eenvoudig zijn, met name niet omdat er niet echt een reëel alternatief is. Generieke advertenties leveren niet genoeg op om gratis diensten te kunnen ondersteunen. En de consument laten betalen voor online diensten lukt maar mondjesmaat.

Dit creëert een buitengewoon lastig dilemma. Big data verbieden zal internetondernemers zwaar raken in hun commerciële mogelijkheden, en gezien de belangstelling voor hun diensten bij het publiek is een verbod moeilijk uit te leggen.¹¹ Big data ongebreideld toestaan leidt echter weer tot onwenselijke uitsluitingen en beperkingen: "sorry, de computer zegt nee". De enige echte compromisoplossing voor dit soort dilemma's is *privacy by design*: zorgen dat technologie inherent rekening houdt met privacy. Dat er niet meer gegevens verwerkt of vrijgegeven

worden dan nodig. Maar juist dat is weer praktisch onuitvoerbaar, want privacy is een rekbaar begrip. Welke privacy design je in? Hoe ver ga je daarin?

Het liefst zou ik zien dat de wetgever een kader introduceert van wat wel en niet wenselijk is. Los van de techniek: wat voor verwerkingen vinden we acceptabel, welke zijn dubieus en welke kunnen echt niet. Ter vergelijking wijs ik op de zwarte en grijze lijsten van algemene voorwaarden: welke bedingen zijn acceptabel in algemene voorwaarden, welke zijn dubieus en welke willen we gewoon niet hebben. Maar voordat we een dergelijke set lijsten hebben opgesteld, is big data wellicht alweer achterhaald door iets nóg groters en nieuwers.

Mr.ir. A.P. Engelfriet is ICT-jurist en partner bij ICTRecht te Amsterdam.

⁹ Farahat, Ayman, and Michael C. Bailey. "How effective is targeted advertising?." Proceedings of the 21st international conference on World Wide Web. ACM, 2012. Zie ook Ur, Blase, et al. "Smart, useful, scary, creepy: perceptions of online behavioral advertising." Proceedings of the Eighth Symposium on Usable Privacy and Security. ACM, 2012.

¹⁰ Artikel 42 Wet bescherming persoonsgegevens: "Niemand kan worden onderworpen aan een besluit waaraan voor hem

rechtsgevolgen zijn verbonden of dat hem in aanmerkelijke mate treft, indien dat besluit alleen wordt genomen op grond van een geautomatiseerde verwerking van persoonsgegevens bestemd om een beeld te krijgen van bepaalde aspecten van zijn persoonlijkheid."

¹¹ Zeker sinds de ondernemersvrijheid expliciet ook als grondrecht is erkend, Hof van Justitie EU 24 november 2011, C-70/10 (Scarlet Extended NV/SABAM).