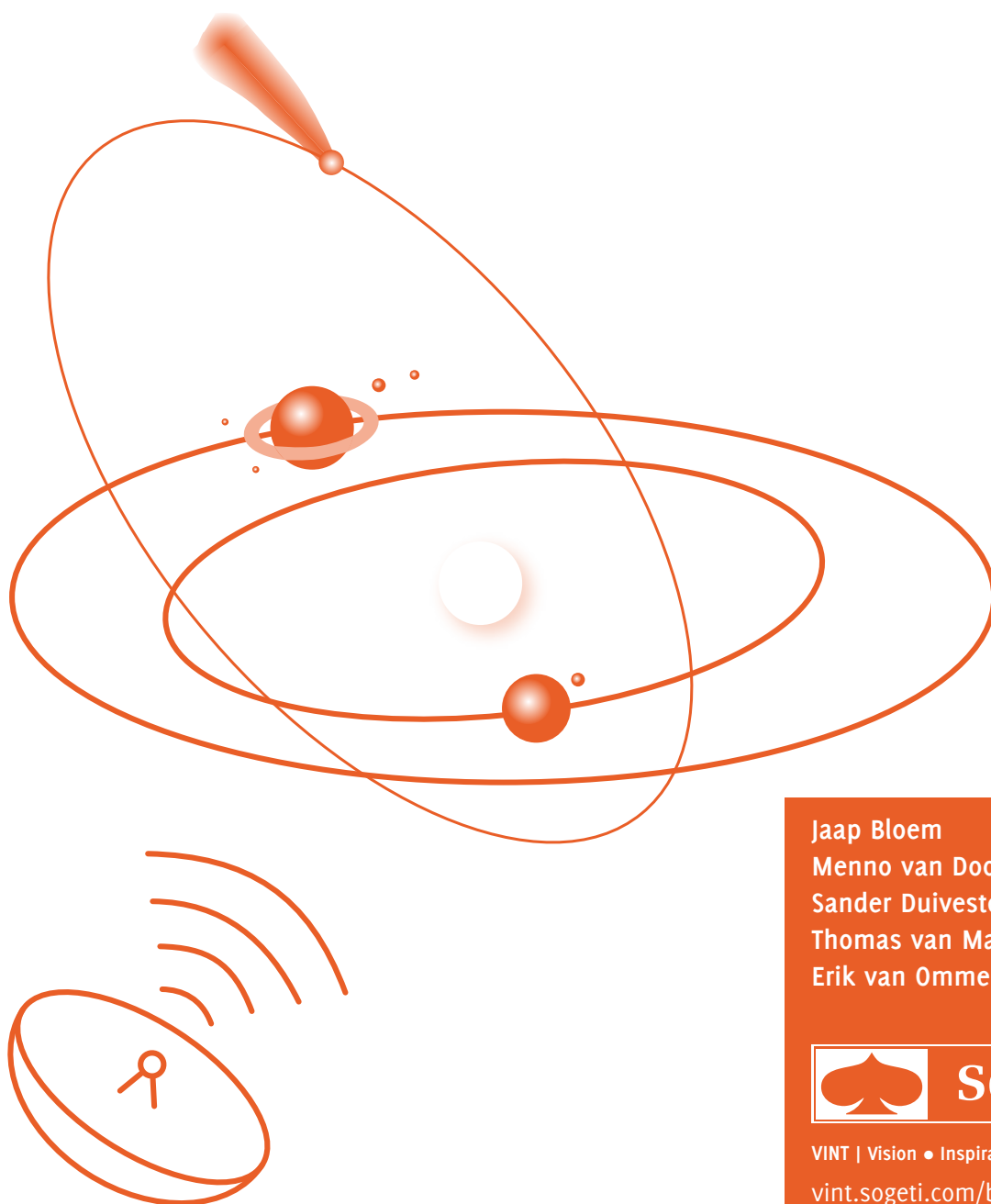


Big Social

Gedrag voorspellen met Big Data



Jaap Bloem
Menno van Doorn
Sander Duivestijn
Thomas van Manen
Erik van Ommeren



SOGETI

VINT | Vision • Inspiration • Navigation • Trends

vint.sogeti.com/bigdata

vint@sogeti.nl

Inhoud

De Big Data-onderzoeksnotities van VINT	3
1 What's next in Big Data?	
Negen waarnemingen met beide benen op de grond	4
2 Ritmes van menselijk handelen	8
3 Meer data voor betere antwoorden	10
4 Totaal datamanagement en de 'Big Five' van Social Data	13
5 Hoe Web en Social Analytics verweven raakten	16
6 Naar Next-Generation Analytics	18
7 Data en algoritmes in plaats van modellen	20
8 Sociale media als lensafwijking	24
9 De gereedschapskist puilt uit	25
10 Om te beginnen heel goed luisteren	28
11 De kracht van Big Social Data	34
12 Samenvatting en de organisatie van privacy	38
Acht centrale Big Social-definities	40
Literatuur en illustraties	43
Over Sogeti	47
Over VINT	47



Naamsvermelding-NietCommercieel-GelijkDelen 3.0
(CC BY-NC-SA 3.0)



De Big Data-onderzoeksnotities van VINT

Sinds 2005, toen het begrip Big Data pas werd gelanceerd – opmerkelijk genoeg vanuit O'Reilly Media, dat een jaar eerder met Web 2.0 was gekomen – is Big Data een steeds actueler onderwerp geworden. Qua technologieontwikkeling en businessadoptie is het Big Data-veld sterk in beweging, en dat is een understatement.

In *Helderheid creëren met Big Data*, onze eerste van in totaal vier onderzoeksnotities, gaven we antwoord op de vraag wat Big Data eigenlijk is, waarin het verschilt van bestaande dataduiding en hoe de transformatieve potentie van Big Data wordt ingeschat. De concrete adoptie en plannen in organisaties raken momenteel vooral het thema Big Social: de klantkant kortom, met name geïnspireerd door de sociale netwerkactiviteit van Web 2.0. De data-explosie vindt overal om ons heen plaats, maar een belangrijk deel van de discussie betreft de vraag hoezeer organisaties zich nu in Big Data moeten storten. Het antwoord luidt: met beleid. Dat geldt zeker ook ten aanzien van de privacy-issues, die uitgebreid in onze derde onderzoeksnotitie aan bod zullen komen.

Met vier Big Data-notities beoogt VINT helderheid te scheppen door ervaringen en visies in perspectief te presenteren: onafhankelijk en aangekleed met voorbeelden. Maar lang niet alle antwoorden zullen kunnen worden gegeven en er zullen zelfs meer vragen bij u opkomen. Bijvoorbeeld over de roadmap die u voor Big Data wilt hanteren. Voor dat onderwerp hebben we onze vierde notitie gereserveerd. Die gaat over management en governance. Vragen zijn er ook over hoe u uw organisatie misschien moet herinrichten. En over de privacy-issues die Big Data-analyse oproept, zoals ten aanzien van Social Analytics. Eveneens kijken we reikhalzend uit naar wat nieuwe algoritmes en systemen ons nog zullen brengen.

De nieuwe datafocus is een zoektocht met veel vragen aan het begin en zeker ook gedurende de reis. Daarom wisselen we graag met u van gedachten: online op www.sogeti.com/vint/bigdata/questions en natuurlijk in persoonlijke gesprekken. Door actief deel te nemen aan de discussie helpt u uzelf en ons om de gedachten ten aanzien van Big Data aan te scherpen. Om door voortschrijdend inzicht te komen tot heldere en verantwoorde beslissingen.

Ter inspiratie treft u in deze notitie ook weer zeven vragen aan waarover we graag uw mening vernemen. In de pdf van dit document kunt u op de betreffende buttons klikken. Vervolgens wordt u direct naar de discussie in kwestie geleid.



We zouden het zeer waarderen als u ons rapport wilt delen met een tweet naar uw volgers. Ze zien dan dat u het rapport leest of gelezen hebt en kunnen het dan eventueel zelf downloaden. Klik op bovenstaande knop en er verschijnt een tweet die u kunt versturen; de inhoud van de tweet kunt u eventueel nog aanpassen.



1 What's next in Big Data?

Negen waarnemingen met beide benen op de grond

Deze tweede onderzoeksnotitie over Big Data beginnen we heel bewust met beide benen op de grond. Als eerste vervolg op onze startnotitie *Helderheid creëren met Big Data* presenteren we hier negen waarnemingen. Voor het netelige onderwerp privacy en hoe we daarmee om kunnen gaan, onder meer met Big Brother-angst, hebben we het derde rapport gereserveerd. Deze tweede notitie, *Big Social* genaamd, biedt u een veelzijdige oriëntatie op de veelbelovende Big Data-ontwikkelingen inzake Social Analytics en Social Media. Sommige ontwikkelingen zijn daadwerkelijk veelbelovend, andere beloven eigenlijk alleen maar veel. Hoe steviger u in uw schoenen staat, des te beter bent u in staat om zelf te oordelen en hiernaar te handelen.

Dit rapport is toegespitst op de voorspelbaarheid van consumentengedrag. Het brede scala aan Big Data-gerelateerde onderwerpen vindt u op onze website <http://vint.sogeti.com/bigdata>. Daar vindt u ook de actuele discussies en de video's met expert-interviews. De hamvraag voor nu luidt simpelweg *What's next in Big Data?* Veel organisaties vinden namelijk dat ze al erg lang wachten op concrete handreikingen. Die moeten natuurlijk makkelijk te implementeren zijn en zichzelf ruimschoots terugverdienen.

Gretigheid en scepsis tegelijk kenmerken de huidige eerste fase van Big Data-business-praktijken, maar die zullen zich verder ontwikkelen en geïntegreerd raken. De data-explosie is een feit en de bijbehorende managementrevolutie, zoals Andrew McAfee en Erik Brynjolfsson het noemen, staat op stapel. Uiteindelijk zullen de *emerging next practices* van vandaag hun weg vinden naar de dagelijkse operatie van organisaties, want *'you can't manage what you don't measure'*. Deze voor de hand liggende wijsheid in de strijd tegen intuïtie en heroïsch management stamt al uit de tijd van W. Edwards Deming en Peter Drucker, die zich er beiden van hebben bediend. In het Big Data-tijdperk wordt het excuus dat we iets niet hadden kunnen weten met de dag minder geldig, maar tevens worstelen veel organisaties nog met de ROI van Big Data-inspanningen; niet in de laatste plaats omdat datamanagement de afgelopen decennia al flink is geëvolueerd.

Graag delen we in dit verband de volgende negen waarnemingen. Ze sluiten aan bij het ambivalente sentiment van vandaag, dat uiteenloopt van erkenning naar ontkenning van Big Data-waarde. Beschouw ze daarom vooral als centraal onderdeel van de dynamische Big Data-discussie die nu volop wordt gevoerd:

1. Gekwalificeerde best practices zijn nog 'under construction'.
2. De technologie breekt nu door.

3. Big Data zal de belofte van hypertargeting inlossen.
4. Big Data loopt het gevaar een mislukt IT-feestje te worden.
5. Bepaal steeds goed wat u daadwerkelijk nodig heeft.
6. Gedrag voorspellen kan vaak best zonder Big Data.
7. De Big Data-trend stoelt op de traditionele datafocus.
8. Big Data-ROI begint zich af te tekenen.
9. Social Analytics is momenteel het Big Data-troetelkind.

1 Gekwalificeerde best practices zijn nog ‘under construction’

Iedereen die zich wel eens heeft verdiept in Big Data kent natuurlijk het beloftevolle rapport *Big Data: The Next Frontier for Innovation, Competition, and Productivity* van het McKinsey Global Institute. Daar staat heel wat te gebeuren, denk je dan!

Des te opmerkelijker was de volgende ontboezeming die Michael Chui, een van de auteurs, deed op het MIT Sloan CIO Symposium in mei 2012. Precies een jaar na de publicatie van het *Next Frontier*-rapport zei Chui het volgende:

*‘There are no [Big Data] best practices.
I’d say there are emerging next practices.’*

Dit lijkt op gespannen voet te staan met de titel van het rapport, maar de overeenkomst die tevens de rek biedt, zit hem in het woordje ‘next’. Voor innovatie, concurrentie en productiviteit zal Big Data grote vruchten kunnen afwerpen, maar het overtuigende bewijs daarvan ligt nu nog even niet keihard op tafel. Inderdaad staat er heel wat te gebeuren, maar er moet nog flink worden gewerkt om de *emerging next practices* verder te ontwikkelen en te implementeren. Met alle investeringen die al zijn gedaan zitten veel organisaties daar niet op te wachten. Zeker niet gegeven het huidige economische tij.

Organisaties zijn aan het experimenteren, maar het is nog te vroeg voor echte schoolvoorbeelden, best practices, die anderen kunnen volgen. Het Big Data-domein is nog te zeer in ontwikkeling en indachtig de claim van het bevorderlijke effect van Big Data op innovatie en concurrentie, zou het zelfs heel logisch zijn dat dit zo zal blijven. Slagvaardig innoveren en concurreren impliceert namelijk doorlopende verandering en vernieuwing.

2 De technologie breekt nu door

Technologisch gezien is er in de context van Big Data van alles in beweging, bijvoorbeeld de opkomst van de data-analysetaal R. Een ander voorbeeld is AMPLab aan Berkeley University. Met Big Data als vertrekpunt richt AMPLab zich op de gecombineerde kracht van *Algorithms, Machines & People*. Een Big Data-mijlpaal uit de zomer van 2012 was de GraphChi-software, ontwikkeld aan Carnegie Mellon University. Daarmee zijn nu op een gewone pc analyses mogelijk, waar eerder grote computerclusters urenlang op stonden te stampen. Met als benchmark een Twitter-dataset uit maart 2010 bleek één enkele GraphChi-pc die in 59 minuten te kunnen doorzoeken. De vorige keer



Vraag 1
Zijn Social Data over drie jaar een drijvende factor in uw organisatie?

www.sogeti.com/vint/r2q1

dat dit was gedaan, waren 1000 grote computers er ruim 6,5 uur mee bezig geweest. De dataset in kwestie is gratis verkrijgbaar op de website Infochimps.com. Hij omvat 40 miljoen gebruikers, ruim anderhalf miljard tweets en 1,2 miljard connecties tussen gebruikers.

3 Big Data zal de belofte van hypertargeting inlossen

Hoe imposant ook, de grote vraag blijft wat de zin hiervan is. Niet iedereen is er even enthousiast over, maar velen, waaronder Jeff Dachis van de Dachis Group, zijn van mening dat Big Data op social media de glorierijke hypertargeting-toekomst van de marketing is. Ga maar na, zegt Dachis: honderden miljoenen mensen roeren zich op het sociale web om er onbekommerd hun complete leven met elkaar te delen. Dat is zomaar 500 miljard dollar aan brand-engagementwaarde.

Aan het eind van 2012 zullen er naar verwachting 250 miljoen actieve Twitteraars zijn, die samen per dag ruim 250 miljoen tweets versturen. Ter vergelijking: Facebook is inmiddels de magische grens van 1 miljard gebruikers overschreden en er zijn zo'n 200 miljoen mensen actief op LinkedIn. Deze 'consumentisering' van Big Data zal in de toekomst alleen nog maar grotere proporties aannemen.

Scepsis is er over het nut van advertenties op met name Facebook. Vlak voordat Facebook naar de beurs ging in mei 2012, zette General Motors nog rigoreus een streep door zijn advertentiebudget voor het sociale netwerk. Maar datzelfde GM geeft nog steeds driemaal dat bedrag uit aan engagement met mensen op Facebook. De grote uitdaging is om de ROI ervan te meten. Daar hebben we nu geavanceerde Social Analytics-tools voor en nieuwe algoritmes als GraphChi.

4 Big Data loopt het gevaar een mislukt IT-feestje te worden

Een van de toepassingen van Social Analytics is om zoveel mogelijk informatie te verzamelen over het onlinegedrag van mensen – Big Social Data – met als doel te voorspellen wat ze gaan doen en wat ze willen kopen. Peter Fader, hoogleraar marketing aan Wharton Business School en mededirecteur van het Wharton Customer Analytics Initiative, plaatst hier een aantal pittige kanttekeningen bij. De gedroomde goudmijn van Big Social Data vergelijkt hij met Customer Relationship Management, dat begin jaren negentig doorbrak. In het begin was het de heilige graal, maar tegenwoordig luidt het harde ervaringsoordeel: één grote frustratie en veel te duur, kortom weer zo'n uit de hand gelopen IT-feestje. Fader is bang dat het met de huidige Big Data-hype dezelfde kant op zal gaan.



Vraag 2

Maakt het gedrag van uw klanten het noodzakelijk om u met Big Data bezig te houden?

www.sogeti.com/vint/r2q2

5 Bepaal steeds goed wat u daadwerkelijk nodig heeft

Zomaar de hele Twitter- of Facebook-'fire hose' door een of andere Social Analytics-raffinaderij halen is grote onzin, zegt Fader. Als je het over hypertargeting hebt, moet je op individueel niveau kijken naar tweets en die koppelen aan de transacties die een persoon doet. Maar online en mobiel vormen samen helemaal niet zo'n compleet nieuwe wereld als met name de Big Social Data-evangelisten ons willen doen geloven.

Natuurlijk kan meer informatie tot nieuwe inzichten leiden, maar de vraag is hoeveel data daarvoor nodig is. Hoe interessant is het eigenlijk om elke seconde te weten waar iemand die aan het winkelen is, zich bevindt en wat hij onderweg allemaal bekijkt? En welke gegevens moeten we daarvan bewaren?

6 Gedrag voorspellen kan vaak best zonder Big Data

De echte gouden jaren van 'predictive behavior' situeert Fader zo'n halve eeuw geleden. Toen was er bijna geen informatie. In de jaren zestig begon Lester Wunderman met wat hij 'direct marketing' noemde. Dat was met recht 'data science'. Alles wat er over een klant te weten viel, werd bijgehouden. Waar de direct-marketingpioniers uiteindelijk op uitkwamen was RFM: de relatie tussen *Recency*, *Frequency* en *Monetary value*. Het effect van F op M is duidelijk. R was de grote verrassing: mensen zijn makkelijk over te halen tot herhaald gedrag, zelfs als ze maar sporadisch iets kopen. Je moet er dan echter wel onmiddellijk bij zijn. In de marketing kent iedereen RFM, maar e-commerce mensen zegt het vaak weinig. Met een hoop Big Data kom je ongetwijfeld tot dezelfde conclusie, maar dat is wel zonde van alle tijd en moeite. Een mooie eye-opener in dit verband is het boek *How to Measure Anything: Finding the Value of "Intangibles" in Business* van Douglas Hubbard uit 2007. Dat staat vol voorbeelden en tips om op een praktische manier van alles aan de weet te komen.

7 De Big Data-trend stoelt op de traditionele datafocus

De meeste Big Data-exercities vinden grotendeels of helemaal in de bestaande data-omgeving plaats. Als dat een geavanceerd datawarehouse is met bijbehorende tooling, dan is daar flink in geïnvesteerd, wordt er al naar grote datasets gekeken en zal men niet zomaar geneigd zijn om vanwege een paar interessante experimenten groot geld te steken in een Big Data-infrastructuur en in oplossingen die nog volop in ontwikkeling zijn. Voeg daarbij het economische tij en het lijkt duidelijk dat de Big Data-euforie misschien niet onder zo'n heel gunstig gesternte staat. Dat de data-toevloed hand over hand toeneemt valt niet te ontkennen, maar die zien we al heel lang aankomen en is eerder een evolutionaire ontwikkeling.

8 Big Data-ROI begint zich af te tekenen

In het rapport *Big Data: The Next Frontier for Innovation, Competition, and Productivity* – nog steeds dé richtingwijzende publicatie – geeft het McKinsey Global Institute voor de Amerikaanse economie per sector aan hoe makkelijk of lastig het is om Big Data te verzamelen, wat Big Data-mining zou kunnen opleveren en hoe Big Data-maturity er stapsgewijs uitziet. Ondanks deels fundamentele slagen om de arm heeft dat bij elkaar de indruk gewekt van aantoonbare Big Data-ROI binnen afzienbare tijd. Anderhalf jaar na dato blijkt het nog steeds erg vroeg te zijn en gaan de ontwikkelingen niet zo hard als de verwachting die is gewekt. Er wordt zelfs gesteld dat Big Data helemaal niet zo'n handige term is, maar dat 'totaal datamanagement' of iets dergelijks de lading beter dekt – is dat dan waarover het gaat? Het antwoord is ja. Een totale-data-aanpak is essentieel en Big Data-ROI begint zich nu af te tekenen. In dit rapport geven we daar duidelijke en richtingwijzende voorbeelden van.



Vraag 3
Op welke manier staat de waarde van Social Data bij u op de balans?

www.sogeti.com/vint/r2q3

9 Social Analytics is het Big Data-troetelkind van deze tijd

Social Analytics, het station tussen Web Analytics en het zogeheten Next-Generation Analytics, is een krachtig middel tegen al te veel Big Data-scepsis. In Next-Generation Analytics, dat onder meer onderzoeksbureau Gartner ons in het vooruitzicht stelt, valt wederom dat verraderlijke woordje 'next' op en ook ten aanzien van Social Analytics wordt volop bediscussieerd of het glas momenteel halfvol is of eerder half leeg. Sullivan McIntyre van Radian6, onderdeel van Salesforce, is de eerste mening toegedaan en benadrukt het volgende: *'It becomes increasingly possible to make guesses about future behavior'*. Paul Barrett van Teradata laat zich een stuk relativerender uit wanneer hij zegt: *'We are still in the early, black and white TV stage of Social Analytics.'*

Heldere oriëntatie gewenst!

De vraag is nu hoe snel we van zwart-wit kunnen overstappen naar kleur, en daarna verder door naar HD en 3D. In deze onderzoeksnotitie, die we simpelweg *Big Social* hebben gedoopt, willen we u voorzien van voldoende ammunitie om zelf een oordeel te kunnen vellen en de ontwikkelingen in de gaten te houden. Dat doen we ook online op <http://vint.sogeti.com/bigdata>, waar we u graag bijpraten over de allernieuwste waarnemingen van VINT en een select aantal gewaardeerde kenners van het Big Data-veld.

In de volgende elf paragrafen van deze notitie kijken we naar de vele facetten van wat we kort en goed *Big Social* noemen. We staan om te beginnen stil bij de ritmes die in ons gedrag waarneembaar zijn; we belichten de waardebelofte van de digitale data-explosie die alleen maar groter wordt; we gaan in op de vijf grootste veroorzakers van Big Social Data; we volgen de lijn van Web Analytics via Social Analytics naar Next-Generation Analytics; we bespreken de lensafwijking die het inzoomen op sociale media kan zijn; we nemen u mee in de wereld van Social Analytics-tools, dashboards en listeningplatforms; en we belichten in concrete voorbeelden de kracht van Big Social Data. Tot besluit introduceren we kort het thema privacy, dat het onderwerp is van ons derde Big Data-rapport.

Natuurlijk kan er nog veel meer worden gezegd, zeker in deze tijden van snelle Big data-ontwikkeling, maar bij elkaar is dit voldoende basis om zelf uw oordeel te kunnen vormen, conclusies te kunnen trekken en daar vervolgens in uw eigen praktijk mee aan de slag te gaan.

2 Ritmes van menselijk handelen

In juni 2012 haalde een opzienbarende *emerging next Big Data practice* het nieuws. 24 uur van tevoren kan al duidelijk zijn waar iemand zich de volgende dag bevindt, op honderd en soms zelfs twintig meter nauwkeurig. Door de data van vrienden te

betrekken bij die van afzonderlijke individuen kon een *Synchronized Rhythm of the City* worden bepaald. Dat inzicht zat verborgen in een Big Data-set, afkomstig van de smartphones van 200 deelnemers. Zonder het gedrag van vrienden was de nauwkeurigheid van de locatiebepaling slechts 1 kilometer. Deze verbetering met een factor 10 tot 50 leverde drie Britse onderzoekers de eerste plaats op in de *Mobile Data Challenge* van het Nokia Research Center Lausanne.

Om ritmes van menselijk handelen in de vingers te krijgen, bijvoorbeeld iets relatief simpels als de bewegingspatronen van mensen in een stad, moeten grote hoeveelheden persoonlijke data worden verwerkt. Het iteratiemodel voor organisaties om daarmee aan de slag te gaan, is de volgende drietrapsraket, dat in deze variant weer betrekking heeft op het favoriete sociaal-mobiele thema locatie:



Door steeds meer data uit steeds meer bronnen bij analyses te betrekken, wordt menselijk handelen steeds beter voorspelbaar. De discipline die zich daarmee bezighoudt, wordt *Predictive Analytics* genoemd, en plaats en tijd zijn dan vaak geliefde parameters. Zoals in het voorbeeld, waar persoonlijke smartphone-data werden verrijkt met die van vrienden. Het resulterende algoritme leverde in dit geval ten minste een tienmaal betere voorspelling op. Hoewel het hier gaat om een researchprototype, is het tevens een overtuigende demonstratie van de kracht van meer data uit verschillende bronnen in een realtime setting. In deze context spelen dus vooral de V's van *Variety* en *Velocity* een rol.

Aan *Volume* is er overigens ook geen gebrek. Digitale sporen, die vaak veel verder gaan dan plaats en tijd, laten we overal achter. Bewust en meestal minder bewust. Als we winkelen, als we de ov-chipkaart gebruiken, op Google aan het zoeken zijn, een digitaal boek downloaden, digitale televisie kijken, muziek luisteren via Spotify, een belletje plegen, een mailtje versturen, een navigatiesysteem gebruiken enzovoort. Zulke data zijn volop beschikbaar om te worden bekeken en benut. Tel daarbij de

explosie van informatie op die we delen op sociale media als Facebook, Foursquare, Twitter, Zoover, Google+, Yelp, Pinterest, YouTube, Yammer en LinkedIn, en we zijn regelrecht in een marketing-walhalla beland, waar traditioneel focusgroeponderzoek, enquêtes en steekproeven tot voor kort de beperkte middelen waren in het kader van het genoemde trio *Understand, Predict* en *Act*. Dat is in het huidige Big Social-tijdperk wel even anders.

Wat we willen, is vanuit marketingperspectief in- en uitzoomen op menselijk gedrag met een analyse-instrument, dat microscoop en telescoop tegelijk is, en ook nog eens voor een groot deel alle data voor ons interpreteert, zodat we op tijd en midden in de roos een aanbod kunnen doen. Of liever nog: meteen afrekenen.

Als een nieuw product flopt in de markt, hoeven we geen maanden meer te wachten op de verkoopcijfers van de winkels. Signalen uit de sociale media kunnen vroegtijdig al voldoende indicatie geven over succes of falen. Dat heet *sentimentanalyse*, een begrip dat hard op weg is om buzzword van het jaar te worden. Op sociale media ventileren we immers doorlopend onze opinies over van alles en nog wat. Het is de grondstof voor de voorspelmechanismen van morgen. De belofte en het vooruitzicht is dat we bijvoorbeeld kunnen voorspellen wie er gaat frauderen, waar de volgende inbraak zal worden gepleegd en waar een klant volgende week zijn zinnen op zal zetten.

Veel dingen kunnen we al van tevoren weten, mits we maar uit andere data-vaten tapen. Social is daarbij dé grote opportunity, en met Big Data en nieuwe analysemethoden en -technieken kunnen we die kans inkoppen. Hypertargeting en personalisatie zijn nog nooit zo duidelijk binnen handbereik geweest als nu.

3 Meer data voor betere antwoorden

Stel, u heeft verschillende verklaringen voor de feiten waarover u beschikt. Maar als u nou moest kiezen, wat zou dan uw antwoord zijn? Tien tegen één precies wat een getergde Sherlock Holmes 120 jaar geleden ook al zei, namelijk:

'Data, data, data! I can't make bricks without clay!'

In onze Big Data-tijd is dit principiële antwoord uit *The Adventure of the Copper Beeches* (1892) relevanter dan ooit. We hebben meer data nodig en liefst misschien zelfs zo veel mogelijk. Kiezen bij gebrek aan gegevens is in elk geval een zwaktebod en volgens Big Data-evangelisten totaal onverantwoord. Want gokken leidt tot grote brokken. Bij waarheidsvinding is dat het geval, maar ook businesswise, namelijk als

onze concurrenten niet hoeven te gokken, omdat die de moeite hebben genomen om verder te kijken dan hun neus lang is.

Mysterie-oplossers – of ze nou Sherlock Holmes, Derrick of Baantjer heten – weten nooit waar ze hun data moeten zoeken. Detectives tasten in het duister en dat is heel frustrerend, zeker als de tijd dringt. Businesswise is er dan vaak geen optie: er moet snel een keuze worden gemaakt. Gelukkig staan de zaken er dankzij internet nu een heel stuk florissanter voor: Big Social Data kunnen steeds betere inzichten garanderen. Dat is hét grote verschil en daarmee, zo is de redenering, wordt het volgen van je intuïtie anno 2012 steeds meer synoniem met de weg van de minste weerstand kiezen. Immers: *'You can't manage what you don't measure'*, om Deming en Drucker nog maar eens aan te halen. Vermeende gaven, zoals ons onderbuikgevoel, kunnen ons makkelijk in de steek laten, maar digitale data niet en ze liggen voor het oprapen. Samen met de benodigde analysemethoden en -technieken. Die kunnen we onder de noemer brengen van Social Analytics.

Er zijn grofweg twee Social Analytics-scholen. De eerste (Roebuck, 2011) richt zich volledig op social media vanwege:

- de groei van social media;
- de groei van social media analytics-tools;
- de groei van social businesses op basis van deze twee.

De andere school, die wij hier zullen volgen, ziet Social Analytics in een breder perspectief en richt zich op de analyse en voorspelling van menselijk gedrag op basis van uiteenlopende bronnen; lang niet alleen social media. Tot die groep behoort onder andere Mary Wallace, Social Analytics Strategist bij IBM.

De volgende case is een sprekend voorbeeld van hoe in dit geval een warenhuisketen in feite vrij eenvoudig hele intieme feiten kon achterhalen op basis van koopgedrag.

De omzet van het Amerikaanse warenhuis Target steeg in tien jaar van 44 naar 67 miljard dollar, omdat men in staat was klantgroepen goed te segmenteren en gericht relevante aanbiedingen te doen. Voor een groot deel was dit te danken aan data-expert Andrew Pole. Zijn werk is het om grote hoeveelheden data net zo lang tegen het licht te houden tot er een harde, voorspellende waarde is bereikt. Zo keek Pole naar de verkoop van babyproducten in relatie tot veranderd koopgedrag in de maanden ervoor. Bijvoorbeeld lotion die minder sterk ruikt, omdat vrouwen tijdens de zwangerschap een sterkere reukzin ontwikkelen, voedingssupplementen als zink en foliumzuur en desinfectiesprays voor handen. Pole kwam uit op in totaal 25 producten, waarmee hij zelfs de datum kon bepalen waarop een zwangere vrouw was uitgerekend.



In Minneapolis kon de plaatselijke Target-manager, die geen flauwe notie had van hoe de aanbiedingen tot stand kwamen, tegenover een boze vader niet verklaren waarom zijn dochter kortingsbonnen voor babyspullen ontving. Bij nader inzien bleek de data-analyse van het warenhuis het bij het rechte eind te hebben. Dat is allemaal heel slim en waardevol voor in principe beide partijen, maar klanten moeten het wel willen en zich niet in hun privacy aangetast voelen. Begin 2012 publiceerde *The New York Times Magazine* het spraakmakende artikel 'How Companies Learn Your Secrets', onder meer op basis van deze Target-case. Op de cover werd groot uitpak met de tekst 'Hey! You're Having a Baby!', die helemaal was opgebouwd uit productverpakkingen. De pay-off: *How Your Shopping Habits Reveal Even the Most Personal Information*.

Hypertargeting als Big Social-trend

Zoals Target het aanpakt is een lichtend Big Data-voorbeeld, dat iedereen kan volgen. Big Social-hypertargeting kunnen we tegenwoordig zelfs gewoon inkopen, bij MyBuys bijvoorbeeld. Dat bedrijf specialiseert zich in cross-channel personalisatie voor onlinereetailers en consumentenmerken. Het doel is om engagement, conversie en omzet te stimuleren door consumentengedrag bij te houden en vervolgens met speciale algoritmes te voorspellen wat een klant graag zou willen kopen. MyBuys maakt haar personalisatieslagen op basis van een Big Data-repository met meer dan 200 miljoen consumentenprofielen en 100 terabyte aan data.

Dit type ontwikkeling heeft al geleid tot veranderingen in de structuur van verschillende organisaties. Net zoals de omgang met Big Data leidt tot de werving van een 'Data Scientist' en de aanstelling van een 'Chief Analytical Officers', maakt hypertargeting met Big Data de rol van 'Chief Customer Officer' wenselijk. De cco Council omschrijft die rol als: *'an executive that provides the comprehensive and authoritative view of the customer and creates corporate and customer strategy at the highest levels of the company to maximize customer acquisition, retention, and profitability.'*

4 Totaal datamanagement en de 'Big Five' van Social Data

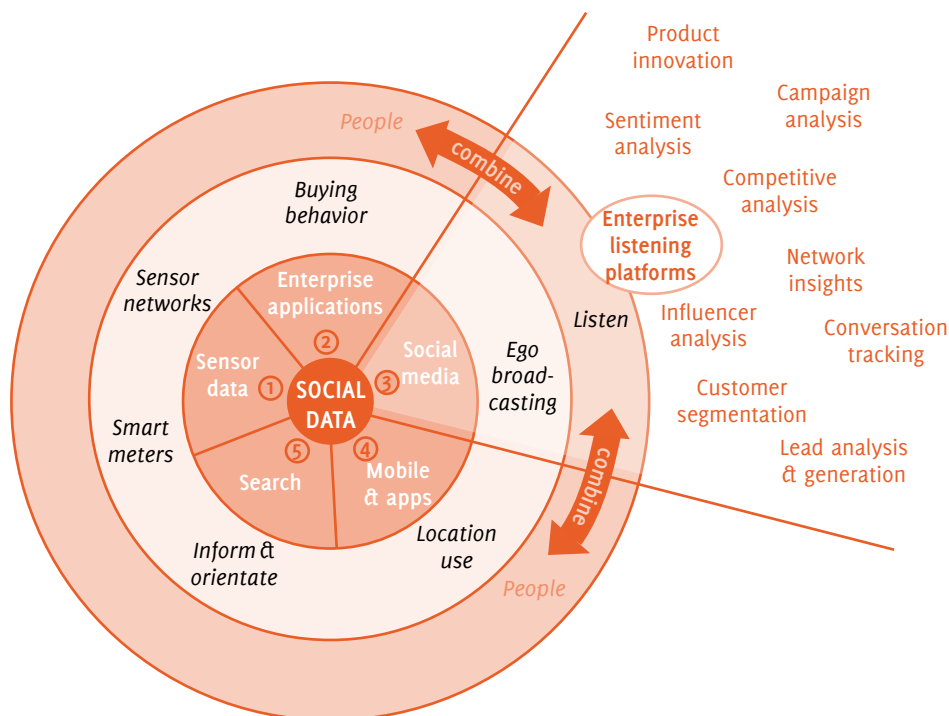
Aan gedragsanalyses kunnen tal van data ten grondslag liggen, bijvoorbeeld klantenkaarten, zoekwoorden op internet, aankopen en responses op kortingsbonnen. Behalve de traditionele enterprise-applicaties als databron zijn er tegenwoordig nog ten minste vier andere datacategorieën die de *emerging next Big Data practices* voeden van Social Analytics in de ruime zin van het woord 'social'. Het zijn: mobiele data en app-data, zoekmachinedata, sensordata en semantische data (zoals bijvoorbeeld bij *smart metering*), en natuurlijk Social Media Data.

Elk van deze 'Big Five'-databronnen heeft zijn eigen kenmerken. De ene zegt iets over hoe mensen zoeken op internet, een andere laat zien welke patronen er zitten in aankopen. De ene soort data komt uit eigen systemen, de andere komt van buiten. Social Media Data gaan over wat mensen beweegt. Daar luisteren we naar met *listening platforms* en we voeren er analyses op uit:



Vraag 4
Welke veranderingen brengen Big Social Data in uw organisatie teweeg?

www.sogeti.com/vint/r2q4



De 'Big Five' van Social Data met uitgelicht de huidige Social Analytics-praktijk voor sociale media.

Het is van belang om het geheel in beschouwing te nemen: een totaal datamanagement-strategie, rekening houdend met het specifieke karakter in sterktes en zwaktes van datasets

en de kracht van slimme combinaties in het kader van de bekende drietrapsraket *Understand, Predict* en *Act*. Maar laten we eerst eens nader naar elk van de 'Big Five' kijken.

1 *Sensordata*

Data van (netwerk)sensors, zoals bij slimme meters, die energieverbruik en energieproductie bijhouden per huishouden en wijk. Het netwerk bestaat uit apparaten die energie verbruiken, uit auto's voor de opslag van energie in accu's en uit mensen. Dit soort *neighbourhood analytics* is onderdeel van nieuwe productiesystemen: een soort *Social ERP*. Maar ook andere data, zoals het volgen met camera's van koopgedrag in winkels (*in-store analytics* en *anonymous analytics*) en tal van andere data uit mens-machine-interactie vallen hieronder.

2 *Enterprise Application Data*

Enterprise Application Data worden traditioneel gebruikt om sociale patronen te herkennen, zoals koopgedrag. Ze bevinden zich in gestructureerde databases van systemen voor Customer Relationship Management (CRM), Supply Chain Management (SCM), Enterprise Resource Management (ERP) of behoren tot de data van de eigen website ('owned media'). Denk aan on-site Web Analytics: de pagina's die mensen bezoeken, waar men op klikt, welke *landing page* het beste tot aankopen leidt enzovoort. Ook zijn er de zogeheten Cross Channel Attribution-tools, die grote hoeveelheden data uit diverse bronnen analyseren (in-store, online en offline sales). Enterprise Application Data zijn traditioneel de basis voor Business Intelligence, HRM-toepassingen en productie- en commerciële processen.

3 *Social Media Data*

Hier gaat het om data, veelal ongestructureerd, die afkomstig zijn van individuen die op sociale media aan 'ego-broadcasting' doen. Deze data zijn toegankelijk voor organisaties. Bedrijfseigen data uit interne microblogs of bedrijfseigen innovatieplatforms kunnen ook belangrijke bronnen zijn. Bij de analyse worden zogeheten *social listening tools* ingezet, die vervolgstappen in marketing en HRM mogelijk maken.

4 *Mobiele Data*

Data uit mobiele toepassingen, zoals de populaire apps-categorie. Spelers als Flurry verschaffen tools voor App Analytics en hanteren dezelfde soort metrieken als bij webgebruik. Mobiele data kunnen de basis zijn voor location based services die realtime worden geleverd. Sociale media op mobiele devices sturen zoals bekend vaak locatiegegevens mee.

5 *Search Data en externe internetdata (off-site)*

Search Data kunnen afkomstig zijn van zoekmachines, bijvoorbeeld Google Trends of Google Insights, van andere leveranciers die de zoekmachines 'scrapen', van 'phoning home'-software of van ISP's. De data worden gebruikt voor trendanalyses, Search Engine Optimization en Search Engine Marketing, bijvoorbeeld door gebruik te maken van keyword monitoring en diensten van Google Adwords. Off-site webdata,

die niet afkomstig zijn van de eigen website, geven bijvoorbeeld inzicht in de populariteit van websites, waar de meeste buzz is of waar comments gegeven worden. Dergelijke data kunnen uit logfiles van webserver komen of uit pagetagging met Javascripts. Dat gebeurt zeker ook on-site, op de eigen website.

Totaal datamanagement

Een totaal datamanagement-strategie op basis van deze 'Big Five' is niet alleen een vergezicht waar organisaties aan kunnen werken, het is ook een van de voornaamste trends die we zien bij *analysis vendors* als Alteryx, comScore, Datasift, IBM, Infochimps en Salesforce. Op Twitter profileert comScore zich kernachtig als volgt:

'comScore #measures the digital world. We manage #bigdata to bring you #mobile, #search, #video insights and more.'

Data-aggregator Alteryx lijkt wel heel bijzonder te zijn, namelijk:

'The only Business Intelligence company to offer packaged data from Tom Tom, Experian Marketing Services, Dun & Bradstreet and the 2010 US Census, and firmographics from the world's leading business data supplier in every license. This data, which spans spatial, demographics, household and firmographic market information, provides businesses a deeper understanding of where, why and with who events occur. By analyzing this market data in combination with their own data, businesses are able to perform analysis that drives highly targeted and localized decision making, and improve the overall ROI of every piece of data.'

Iets vergelijkbaars doet Salesforce met Social CRM door combinaties te maken van Social Media Data en CRM, of het bedrijf Datasift, dat een cloud-oplossing biedt om enterprisedata te verrijken met data afkomstig van sociale media.

In Israël werkt IBM aan een app om mobiele data te combineren met enterprisedata als historisch koopgedrag en persoonlijke klantvoorkeuren: een interessante Big Data-toepassing voor warenhuizen en supermarkten. De app plaatst alle data in een augmented reality-laag om klanten via speciale persoonlijke triggers en acties door de winkel te loodsen.

'Beyond what IBM's augmented reality app may offer retailers on a customer-by-customer basis, it has the potential to create a treasure trove of metadata regarding product sales trends, hot in-store selling spots, traffic patterns and inventory issues, all of which could be used to maximize revenue per square foot, a critical metric in retail.'

Google Now mixt ook allerlei sociale data, alleen niet voor bedrijven maar voor individuen, op basis van locatie, onze agenda, mail, search en video. Just-in-time kan

precies worden gepresenteerd wat we nodig hebben, rekening houdend met bijvoorbeeld vluchttijden en verkeersopstoppen.

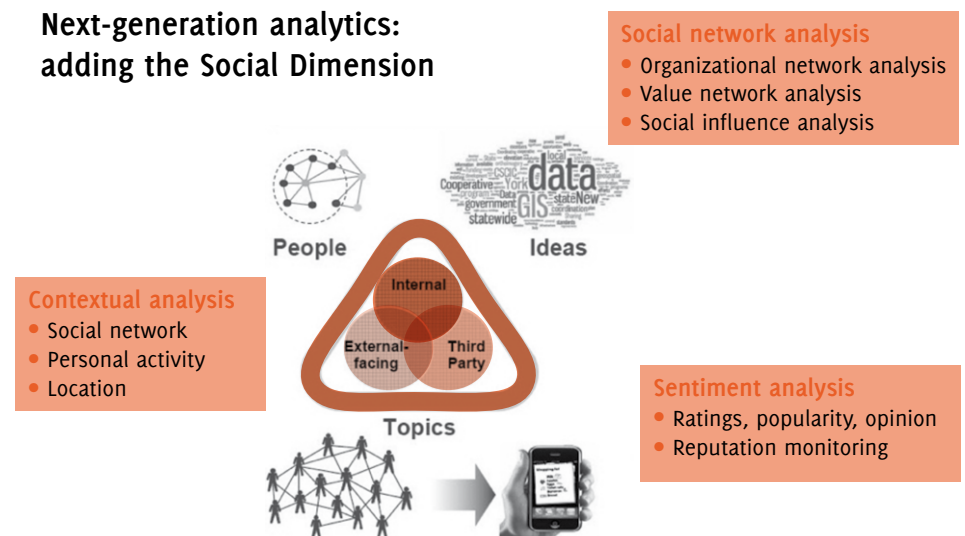
Op de Windows Azure Marketplace, de weblocatie van Microsoft voor waardevolle data en applicaties, is onder meer een dataset te verkrijgen met openbare kentekengegevens van onze RDW, de Rijksdienst voor het Wegverkeer.

Overall zien we momenteel deze convergentie van data en tools. Vanuit ons gedrag op internet ontstond die verwevenheid al heel vroeg. Voor een goed begrip van Social Analytics is het belangrijk om te weten hoe deze discipline zich uit Web Analytics ontwikkeld heeft. Vervolgens kijken we waar het straks naartoe gaat.

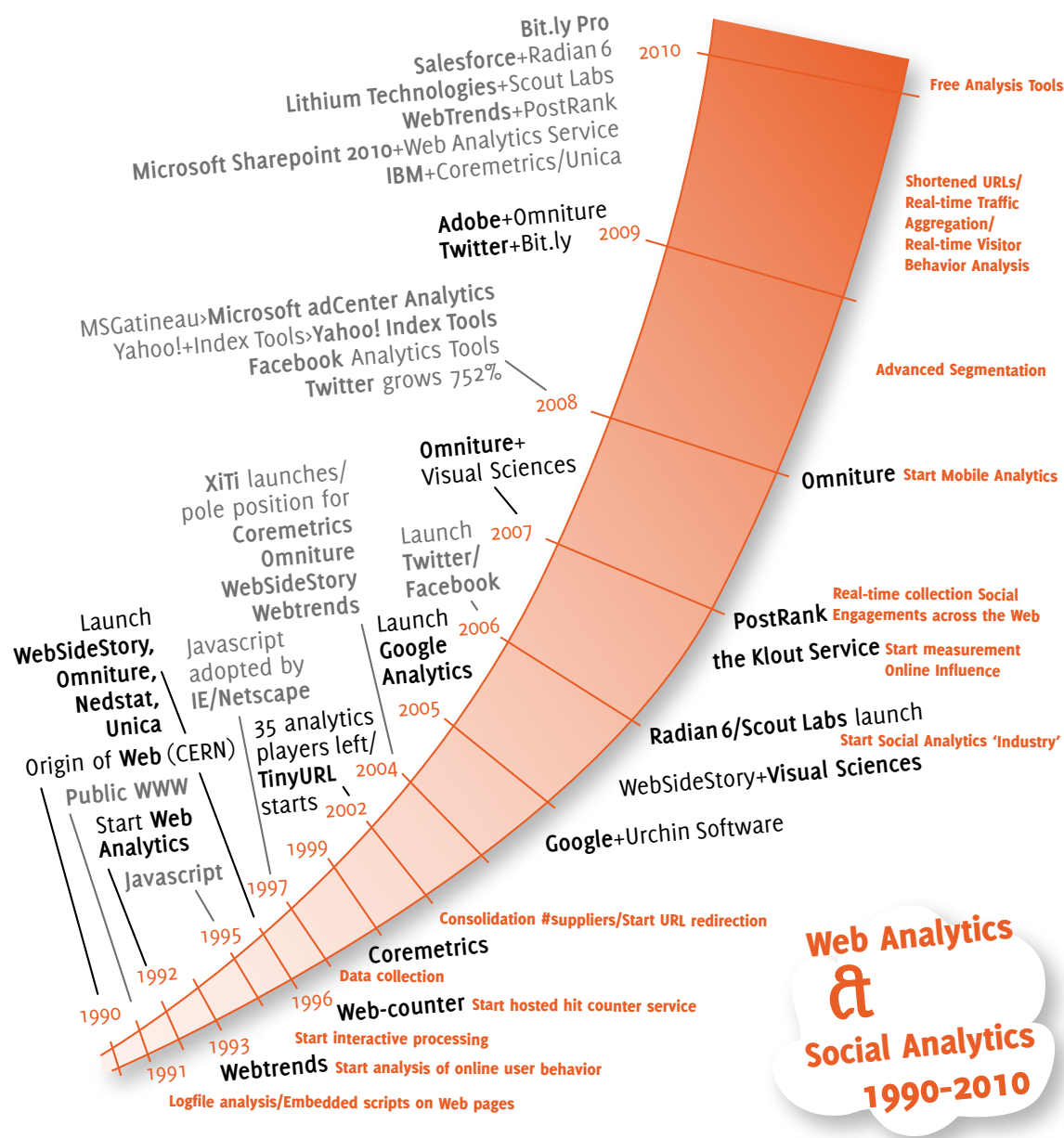
5 Hoe Web en Social Analytics verweven raakten

Met het oog op het gedrag van klanten, prospects en iedereen die zich bemoeit met ons merk en onze producten, gaat de aandacht tegenwoordig uit naar het (sociale) web. Daar is nu in veel gevallen vaak de meeste activiteit, representatieve activiteit, de best rederende activiteit, de activiteit die de meeste buzz genereert en de activiteit die het beste en makkelijkste matcht met andere soorten. Menselijk handelen op internet wordt al heel lang geanalyseerd. Vroege vormen van Web Analytics vonden twintig

Next-generation analytics: adding the Social Dimension



Bron: Gartner, <http://www.scribd.com/doc/81893261/February-9-Top-10-Strategic-Tech-Dclearley>



jaar geleden al plaats. Parallel daaraan is veel later, in 2006, Social Analytics ontstaan. In 2010 zijn Web Analytics en Social Analytics elkaar definitief in de armen gevlogen.

Laten we in het kader van *emerging next practices* eens kijken hoe dat in zijn werk is gegaan aan de hand van een korte tijdlijn, die de laatste tien jaar van de vorige eeuw en de eerste tien van de huidige eeuw omspannt. De gegevens zijn afkomstig uit de infographic *History of Web and Social Analytics (1990-2010)*. Daarom eerst twee korte beschrijvingen van deze twee domeinen:

Web Analytics (1992–2010), Wikipedia:

‘The measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing web usage.’

Social (Marketing) Analytics (2006–2010), Gartner:

‘Social analytics describes the process of measuring, analyzing and interpreting the results of interactions and associations among people, topics and ideas. [...] Social network [or media] analysis involves collecting data from multiple sources, identifying relationships, and evaluating the impact, quality or effectiveness of a relationship.’

In 2010 kwam Social Analytics echt op de kaart te staan, toen Gartner het in de top-10 van strategische technologieën zette. Web en Social Analytics zijn nu soulmates. De hectische dynamiek die we zien in de illustratie *Web Analytics & Social Analytics 1990-2010* is er een van bedrijven die opkomen en samengaan (zwart), en *emerging next practices* (rood) die uiteindelijk allemaal mainstream zijn geworden. Dit gaat natuurlijk nog steeds door en het interessante is, dat in deze tijd van Big Data nu Web & Social Analytics snel evolueren naar een volgend stadium, dat Gartner inmiddels Next-Generation Analytics heeft gedoopt. In paragraaf 7 kijken we in het kader van *emerging next practices* nog een horizon verder.

6 Naar Next-Generation Analytics

Web Analytics en Social Analytics zullen nog lang door elkaar heen worden gebruikt, maar het zwaartepunt ligt op dit moment duidelijk bij Social. Als we hulde willen brengen aan wie die toekomst, dan is het Lars-Henrik Schmidt. Sinds begin jaren tachtig is deze Deen namelijk al bezig met de ontwikkeling van een beschrijvend filosofisch perspectief, dat hij Social Analytics noemt. Schmidt had een vakgebied voor ogen dat zich in zijn algemeenheid bezighoudt met rapporteren over ‘de trends van deze tijd’.

Die halen we nu eenvoudig uit de trending topics van Twitter. Per minuut zien we waar mensen het over hebben door te tellen hoe vaak een woord met een zogeheten *hashtag* (#) wordt getweet. Willen we de trends per land weten, dan gebruiken we een tool als Twirus. In een oogwenk rapporteren is met sociale media mogelijk geworden. Het enthousiasme over het gebruik van sociale mediatools is dus heel begrijpelijk. Analyses zijn razendsnel beschikbaar, er zijn gratis tools waarmee we meteen aan de slag kunnen en tegelijkertijd kunnen we uitzoomen voor de Big Picture en inzoomen op de (potentiële) kopers van producten en diensten.

Maar ook Social Analytics is een overgangsfenomeen, althans de overmatige nadruk daarop. In de Gartner-top 10 van veelbelovende trends voor 2011 stonden twee datamining-gerelateerde domeinen: Social Analytics en Next-Generation Analytics. Die volgende generatie is waar het natuurlijk naartoe gaat: de integrale Big Data-benadering van het zogeheten totaal datamanagement, dat we in notitie 1 *Helderheid creëren met Big Data* als de nieuwe focus presenteerden.

We maakten al eerder kennis met hoe Gartner Social Analytics definieert. Next-Generation Analytics wordt als volgt getypeerd:

'It is becoming possible to run simulations or models to predict the future outcome, rather than to simply provide backward looking data about past interactions, and to do these predictions in real-time to support each individual business action.'

Hoe het dan concreet verder gaat in die 'Next Generation' vertelt het rapport *A Framework for Social Analytics* van Susan Etlinger en Charlene Li, dat Altimeter Group in augustus 2011 publiceerde. Ook hier is de overmatige nadruk op *social* een 'passing phase', zoals dat heet, want *'Social is One of many Signals – Data is King'*, zo wordt geconcludeerd. Altimeter schat de toekomst van Social Media Measurement als volgt in:

"Social analytics" or "social intelligence" will become an integral – and eventually indistinguishable – element of the enterprise's ability to sense, interpret, and recommend actions based on signals from the market. [...] One of the greatest impacts of the transition to the adaptive business is the advent of "Big Data" – the algorithmic increase in unstructured data that will stem from continuous interaction with customers, communities, and markets. [...] Clearly this future state is several evolutionary steps away from social media monitoring and will require an entirely new processing and analytics approach that is able to make sense of both the unstructured nature and the sheer volume of data.'

Altimeter zegt daarmee precies hetzelfde als Gartner met zijn tweedeling in Social Analytics en Next-Generation Analytics, namelijk: we zijn op weg van A naar B, van Web Analytics en Social Analytics naar *Total Data Analytics*. Zo zouden wij het willen noemen, aanhakend bij totaal datamanagement.

Nota bene

Een beknopt overzicht van het definitielandschap vindt u in de paragraaf *Acht centrale Big Social-definities*, aan het eind van dit rapport.

7 Data en algoritmes in plaats van modellen

De juiste interpretatie en koppeling van data leidt nu al tot betere beslissingen, meer verkopen, risicobeperking en kostenreductie. Maar indachtig de *emerging next Big Data practices* van Michael Chui komen er nog legio verdere verbeteringen en mogelijkheden, immers:

‘We are still in the early, black and white TV stage of Social Analytics.’

Dat zegt Paul Barrett, Customer Management Director bij Teradata. Laten we eens kijken wat dat letterlijk genomen zou betekenen. Zwart-wit was de tv-periode waarin er nog maar weinig kanalen waren en we verschillende antennes nodig hadden om ze te ontvangen. Heel vaak stond er plotseling ‘STORING’ op de buis. We hadden last van ‘atmosferische storingen’ en zagen niets dan ‘sneeuw’ als harde wind de antenne een stukje had gedraaid. Het was een verre van ideale situatie met alleen maar grijs tinten om een bontgekleurde wereld weer te geven. Slechts op gezette tijden konden we steeds naar hetzelfde programma kijken.

Het is overdreven om Big Social zo weg te zetten, want met moderne Social Analytics is het zo ellendig zeker niet gesteld. Maar het kan nog veel beter: één antenne graag, scherper beeld, meer details, meer kanalen en bronnen, realtime, verschillende invalshoeken, meerdere aggregatieniveaus, patroonherkenning en vooral: gedrag kunnen voorspellen. Kortom, weten wat de mensen willen, ze op tijd kunnen bedienen, ze aan ons binden en met en via hen relaties op- en uitbouwen. Dat spel opzienbarend verbeteren is de commerciële Big Data-uitdaging voor organisaties.

Als een dataset groot en actueel genoeg is, werkt de empirische aanpak vaak beter dan een formule. We zouden een complex model kunnen opstellen om te bepalen hoeveel mensen ziek worden van de griep, maar zoekresultaten doorspitten levert sneller een duidelijker beeld op. Gunther Eysenbach, hoogleraar aan de universiteit van Toronto, was in 2006 de eerste die dat deed. Zijn conclusie:

‘The Internet has made measurable what was previously immeasurable: The distribution of health information in a population, tracking (in real time) health information trends over time, and identifying gaps between information supply and demand.’

Hetzelfde geldt voor veel andere dingen, zoals de juiste prijsstrategie voor tweedehands artikelen. Die vinden we natuurlijk onmiddellijk in data afkomstig van eBay, die ook beter inzicht kunnen geven in zaken als inflatie en consumentenvertrouwen. In grote datasets liggen kortom allerlei antwoorden besloten en die kunnen we te weten komen zonder ons met modellen te hoeven bezighouden.

Chris Anderson van het technologiemagazine *Wired* sprak in 2008 al provocatief over een Big Data-vergezicht, waarin zelfs theorievorming en de wetenschappelijke methode overbodig zouden worden, maar data spreken natuurlijk niet zo maar voor zichzelf. Empirie en theorie spelen hooguit haasje-over en dankzij de data-explosie ligt de nadruk momenteel meer op data en algoritmes dan op traditionele modellen. Die ontwikkeling is al enige jaren gaande; vergelijk bijvoorbeeld de statistische benadering met die van machine learning:

'Statisticians emphasize probabilistic models for learning, and techniques for quantifying variation in the estimated model that results from variation in the learning sample. For many machine learners, the algorithm is the model, and emphasis is placed on developing interpretable yet flexible methods of learning in challenging context (computer vision, natural language).'

http://www.crm.umontreal.ca/Machine06/index_e.html

Realistische Big Social-scenario's

De laatste ontwikkelingen op basis van Big Data en algoritmes gaan qua voorspellingskracht en ambitie veel verder dan traditionele Web Analytics, aangevuld met dashboards om Twitter- en Facebook-verkeer in de gaten te houden en snel te kunnen reageren. Voordat we daar in de volgende paragrafen op ingaan, kijken we hier eerst over de horizon heen naar scenario's die op de keper beschouwd niet eens al te fantasieerlijk zijn.

1 Conclusies trekken uit schijnbaar ongerelateerde zaken

We kennen dit uit online winkelen met aanbevelingen voor zaken waar we ook in geïnteresseerd zouden kunnen zijn. Soms zijn de artikelen logisch: bij de aanschaf van een digitale camera worden we gewezen op een extra batterij of een geheugenkaart. Maar wat nu als schijnbaar compleet ongerelateerde zaken voorspellend blijken te zijn? Bijvoorbeeld als de snelheid waarmee u langs de bakker rijdt, iets zegt over de kans dat u vanavond naar de film gaat? Of als het moment van de dag waarop u een spelletje Angry Birds speelt, een goede indicator zou zijn om u straks in de supermarkt te laten kiezen voor een duurdere fles wijn? Of misschien zelfs dat een 'like' op een foto op Facebook iets zegt over uw gezondheid? Zulke verbanden zitten verborgen in Big Data: patronen waar nog nooit iemand bij heeft stilgestaan. Het is niet ondenkbaar dat systemen zelf op zoek gaan naar correlaties en ze ons proactief voorschotelen.

2 Levenschte persona's creëren

In de marketing, en steeds vaker ook in de IT – want in de context van Business Technology zijn ze twee handen op een buik – werken we vaak met persona's: concrete persoonsbeschrijvingen die kenmerkend zijn voor typische klanten. Op het moment is het nog mensenwerk om op basis van onderzoek en inzichten persona's te ontwikkelen: wie ze zijn, hoe ze heten, wat voor soort baan ze hebben, welke voorkeuren



Vraag 5
Welke indicatoren (nieuwe klanten, behoud van klanten, klanttevredenheid of andere) gebruikt u om Big Social-investeringen te rechtvaardigen?

www.sogeti.com/vint/r2q5

enzovoort. Door voor persona's scenario's te bedenken kunnen waardevolle verbeteringen worden herkend die goed zijn voor de klant, het bedrijf, het proces, de keten enzovoort. Door de patronen achter alle klanten te combineren kunnen betere, rijkere en zinnellere persona's worden ontwikkeld.

3 Voorspellen op basis van 'live' gedrag

Nu is veel van de analyse nog gericht op feiten uit het verleden, gecombineerd met acties die op dit moment gebeuren. Maar wat als de realtime component nog verder wordt doorgevoerd? Als mijn telefoon op basis van mijn acties in de voorgaande twintig minuten, en de pagina's die ik online heb bekeken, de apps die ik heb gebruikt, de geluiden uit mijn omgeving, mijn agenda, de activiteiten van mijn Facebook-vrienden en actuele gegevens over de exacte locatie, precies doorgeeft wat ik denk, voel en wil? Dus een actueel inzicht in mijn 'state of mind' en waar ik vatbaar voor ben. Welke thema's zoomen er rond in mijn hoofd, wat is mijn gevoel, welke urgente behoeften heb ik, bewust of onbewust? Google Now is al een aardige eerste aanzet tot hoe dit zou kunnen werken en de invloed van een technologie als Google Glasses, die mikt op continue aanwezigheid, zou een flinke stap verder in deze richting kunnen betekenen.

4 Van voorspellen naar subtiel beïnvloeden

Als we zover zijn dat we de actuele geestgesteldheid van iemand met enige zekerheid kunnen inschatten, dient de volgende mogelijkheid zich onmiddellijk aan: welke kleine en onopvallende prikkels kunnen we iemand voeren om te zorgen dat hij in een bepaalde gemoedstoestand terechtkomt, eentje waarin hij tamelijk gelukkig is, openstaat voor experimenten en klaar is om zijn portemonnee te trekken? Misschien een muziekje uit de telefoon of een plaatje van Facebook naar boven halen om de juiste toon te zetten? Of de route iets aanpassen, zodat we door een straat met meer bomen rijden. Als maar genoeg mensen meedoen, kan door subtiel variatie de optimale invloed worden bepaald. Dat kan ten goede zijn: om mensen een simpel, waardevol en gelukkig leven te laten leiden, maar de mogelijke zwarte kanten van misbruik en verregaande manipulatie zijn niet minder realistisch.

5 Echt slimme organisaties

Alles in een organisatie is tegenwoordig digitaal: e-mail, telefoon, content, financiën, toegangsdeuren, het licht, klimaat, presentaties, trainingen, noem maar op. Kunnen we hier ook meer patronen uit gaan halen? In eerste instantie misschien alleen in grote bedrijven waar de hoeveelheid data groot genoeg is voor significante inzichten, maar op den duur kan het als service worden aangeboden aan elk bedrijf. Benchmarcken tegen de succesvolle bedrijven, processen optimaliseren op basis van hoe er op dit moment echt gewerkt wordt, problemen en kansen signaleren, sturen op risico's en misschien zelfs voorstellen doen voor de meer menselijke aspecten, zoals trainingen, vacatures of evaluaties.

6 Validatie door variatie

Big Data is nu nog vooral het aftappen van datastromen die al bestaan en daar proberen conclusies uit te halen. Maar wat gebeurt er als het systeem zelf op zoek kan naar nieuwe gegevens? Als het met intelligente machine learning zelf dingen kan proberen om te zien wat de effecten zijn? Bijvoorbeeld een nieuwsberichtje naar bepaalde mensen sturen om te kijken of ze er iets mee doen? Of een sms sturen met een feitje uit alle Big Data, om te kijken of een dreigend overschrijden van bepaalde meetwaarden kan worden afgewend? In de wereld van webadvertenties is dit al heel gewoon: het opdelen van alle bezoekers in groepen die elk net een iets andere variant van een advertentie gepresenteerd krijgen, of een iets andere plaatsing op de pagina. Door dan te meten welke configuratie het meeste effect heeft, wordt advertentieplaatsing door de tijd heen steeds efficiënter. Als Big Social-systemen autonoom acties kunnen ondernemen, zou variatie kunnen worden ingebakken om te zoeken naar maxima.

7 Het einde van onvoorspelbaarheid

Als concept is dit op zijn minst interessant te noemen: als de samenleving, het bedrijfsleven en overheden allemaal doordrongen zijn van het belang van data, van zoeken naar patronen en van betere voorspellingen over van alles en nog wat, kunnen we dan komen tot een situatie waarin we bijvoorbeeld met redelijke zekerheid een half jaar vooruit kunnen kijken? Kunnen we dan bewegingen op de markten, in koersen en innovaties zien aankomen? Veel van de huidige economie bestaat dankzij onvoorspelbaarheid: iemand die bereid is risico te lopen in ruil voor betaling. Als dit risico straks lager is, omdat we weten hoe mensen reageren, wat de risico's zijn en wat de kans is dat iets gaat gebeuren, wat wordt dan de kern van de economie? Zolang we het weer niet kunnen voorspellen, is het voorspellen van de samenleving misschien iets te hoog gegrepen. Hoewel, hoe onvoorspelbaar zijn mensen eigenlijk? Professor Dirk Helbing en consorten hebben voor hun *Living Earth Simulator* ofwel *FutureICT Knowledge Accelerator and Crisis Relief System* al 1 miljard euro gekregen van de Europese Unie.

8 De transparante mens

Er vindt veel interessant onderzoek plaats naar de relatie tussen het onderbewuste en het bewuste brein van mensen. De consensus lijkt te zijn dat het onderbewuste eigenlijk voor het grootste deel verantwoordelijk is voor ons gedrag en dat simpel gezegd identiteit en 'ik' eigenlijk meer toeschouwer zijn dan dat ze een sturende rol hebben. Het blijkt lastig te zijn om ons eigen gedrag in te schatten. We kunnen bijvoorbeeld denken dat we het belangrijk vinden om gezond te leven, maar alle daadwerkelijke handelingen duiden erop, dat gezondheid niet zo'n enorm hoge prioriteit heeft. De realiteit is natuurlijk veel complexer, met interessante feedback tussen gedrag, resultaat en 'ik', maar als een paal boven water staat, dat concreet gedrag meer zegt over wie we zijn dan wat we denken. Big Social kan een mooie rol spelen om mensen beter inzicht te geven in wie ze werkelijk zijn: wie bent u, in wat voor werk komen uw kwaliteiten het best tot zijn recht en welke relaties hebben de grootste kans van slagen? Harde data als koevoet om eindelijk eens de 'conditio humana' te kraken.

9 Een hackers-droom

Geen enkel systeem is ooit compleet veilig. De technologie rondom spam en cyber-attacks heeft laten zien dat mensen met minder idealistische bedoelingen ook IT inzetten voor hun eigen gewin. Ook Big Social-tools kunnen prachtig ten dienste staan van criminelen. Allerlei privégegevens kunnen worden gebruikt om wachtwoorden te raden. Of denk aan een automatisch systeem, dat zich perfect kan voordoen als goede vriend of vriendin om op die manier geld los te krijgen, of compromitterende gegevens? Gezien de hoeveelheid spam in de wereld deïnen mensen er niet voor terug enorme overlast te veroorzaken voor hun eigen geldelijke gewin, dus wat doet een crimineel met een perfecte sociale database?

8 Sociale media als lensafwijking

In de metafoor van in- en uitzoomen met verschillende soorten datalenzen zijn lensafwijkingen van groot belang. Voordat we de tools behandelen, die beogen uw zicht te verbeteren en uw blik te verruimen, brengen we eerst graag een zestal notoire afwijkingen onder de aandacht.

1 Niet alle Facebook-accounts zijn echt

Marketeers zijn dol op Facebook-‘likes’, al was het alleen om het succes van eigen acties te kunnen aantonen. Zulke likes moeten dan wel van klanten van vlees en bloed komen. Uit het onderzoek *Who likes my virtual bagels* van de BBC blijkt dat ook een onzinproduct succesvol kan zijn op Facebook dankzij likes van fake-accounts uit Egypte. Bijna 9 procent van de accounts is nep, geeft Facebook zelf toe. Dat zijn er ruim 83 miljoen. Bij 4,8 procent gaat het om dubbele accounts; 2,4 procent behoort niet toe aan echte personen – de doelstelling van Facebook – maar bijvoorbeeld aan een organisatie of een huisdier; en 1,5 procent van de Facebook-profielen wordt bijvoorbeeld gebruikt voor spam.

2 Ook Twitter is een lens met afwijkingen

Bijna de helft van alle Twitter-accounts zou fake zijn. Een bekend grapje luidt: als de wereld morgen vergaat maar computers nog gewoon zouden draaien, dan zijn er nog steeds volop trending topics op Twitter. Er zijn bedrijven die het kaf van het koren scheiden en databases bijhouden met alleen Twitteraars van vlees en bloed.

3 Begrijpt u wat er wordt gezegd?

Automatische tekstanalyse is niet perfect. Dat toont de kwaliteit van Google Translate wel aan. Tweets en comments kunnen ironisch zijn of staan misschien vol met ‘slang’. Hoe bepalen we dan waar het bericht precies over gaat en of het positief of negatief is? Maar ook hier schrijdt de technologie voort en er zijn partijen in de markt die beweren, dat hun computers en software wel degelijk snappen wat er wordt gezegd.

4 Wat meet u eigenlijk?

Was het maar zo eenvoudig dat het succes van reclame kon worden voorspeld aan de hand van wat mensen twitteren of op Facebook zetten. Er is tal van onderzoek naar het effect van reclame dat het tegendeel bewijst. Bijvoorbeeld het *Remember the ad, forget about the product*-effect: humoristische reclames scoren goed, maar de verkopen gaan niet altijd omhoog. En van wasmiddelreclame weten we dat die eigenlijk nooit goed scoort op likeability maar toch effectief kan zijn. Het aantal likes op Facebook is dus geen dwingende reden om door te gaan met een bepaalde campagne.

5 Hoe betrouwbaar is ego-broadcasting?

De meeste biografen waarschuwen voor autobiografieën. Waarschijnlijk terecht, want in de eigen geschreven geschiedenis brengen we het er zelf altijd beter vanaf. Over dat effect schreven we reeds in *Me the Media* (2008) – met name over al die ‘hyperego’s’ die veel overeenkomsten vertonen met Narcissus uit de Griekse mythologie. Wat zegt dat over sociale media? Hoe authentiek of ‘waar’ zijn die berichten? Misschien wordt al het onechte in de wereld straks als waar ervaren, een ‘hyperreality’ zoals onder anderen Umberto Eco het noemt, een Disneyland ‘*that can give us more reality than nature can*’. Maar ook zonder filosofische gedachten is het goed te beseffen dat data uit sociale media vaak met een korrel zout genomen moeten worden.

6 Hoort u ook wat er niet wordt gezegd?

Het bovenstaande is reden te meer om goed te luisteren naar wat er niet wordt gezegd. Sherlock Holmes loste ooit een moordzaak op doordat een hond niet blafte. Alles wat niet getwitterd, gefacebooked en gepinterest wordt, is ook belangrijke informatie. Al was het maar omdat we weten dat sommige dingen niet snel in autobiografieën zullen verschijnen – alle openheid en transparantie op internet ten spijt.

9 De gereedschapskist puilt uit

Ondanks, en natuurlijk eigenlijk vanwege, de kritische kanttekeningen over voorspellen op basis van online uitingen hierboven, is de Social Analytics-gereedschapskist voor bedrijven de laatste jaren explosief gegroeid. Hier is een kleine greep uit wat er zoal is bijgekomen:

- ♦ Insight van Adobe
- ♦ Visible Technologies
- ♦ Networked Insights
- ♦ Biz360 van Attensity
- ♦ Scoutlabs van Lithium
- ♦ Radian6 van Salesforce
- ♦ Cognos Consumer Insights van IBM
- ♦ De Azure/Hadoop-oplossing van Microsoft

- ♦ Een hele reeks tools voor Twitter en Facebook, zoals Osfoora, Tweetdeck, HootSuite, MentionMap
- ♦ De mobile-analytic tools van Google en Flurry

Een tamelijk actueel overzicht van diverse zogenaamde luisterplatforms treffen we aan op Wikispaces (zie: <http://socialmedia-listening.wikispaces.com/Tools>) of de Social Media Monitoring Wiki van Ken Burbary, medeauteur van het boek *Digital Marketing Analytics*, dat begin 2013 verschijnt bij uitgeverij Pearson Higher Education. Ideya Business Marketing & Consultancy telde in juli 2012 het aantal hulpmiddelen en kwam uit op een kleine 250 Social Analytics-tools, waarvan er zo'n 50 gratis zijn. Een deel van het rapport staat op <http://ideya.eu.com/publications.html>.

Sommige leveranciers brengen een groot aantal toepassingsgebieden van Social Analytics bij elkaar. Dat heet dan een *suite* of een *hub*. Om een idee te geven van belangrijke hoofddomeinen, hier de suite-onderdelen van Visible.



De Social Media Hub van Visible

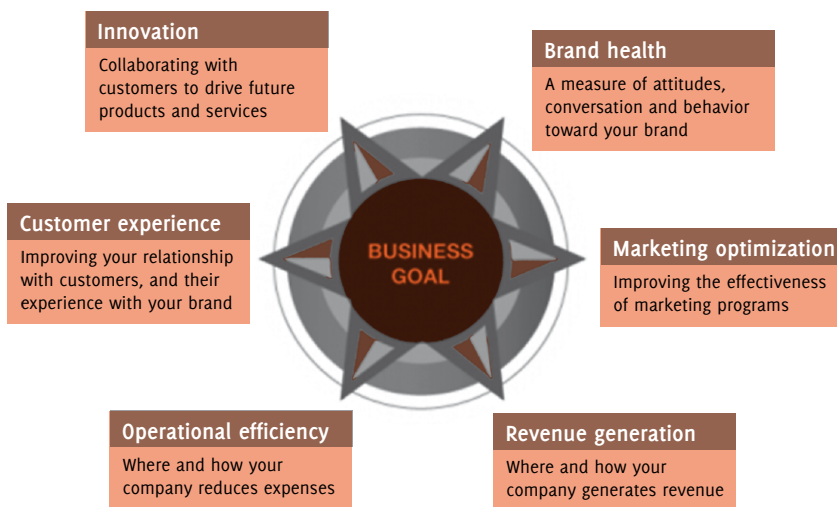
Deze tools beogen ons in staat te stellen om steeds verder en dieper te graven in de gedachten, intenties en gedragingen van mensen, teneinde de concurrentie te slim af te zijn.

Social Analytics kan worden toegepast van pre-sales, zoals *lead generation*, tot after-sales en *customer support*. We kunnen er het salesapparaat de maat mee nemen. Of belangrijke mensen in de organisatie ermee identificeren en belonen, bijvoorbeeld op basis van de *influencer scores* uit de chatter microblog van Salesforce. We kunnen er ook analyses mee doen op een hoger abstractieniveau en 'sentimenten' meten, bijvoorbeeld over bepaalde customer experiences. Allemaal ter inspiratie van het marketingapparaat of van productontwikkeling. We meten er 'rumour around the

brand', 'buzz', 'word of mouth' mee of maken vergelijkingen tussen ons merk en dat van de concurrent. We segmenteren markten, delen klanten in naar typen (persona's) of naar locaties die automatisch worden meegestuurd via mobiele apps. We kunnen meten of een reclamecampagne het goed heeft gedaan, al dan niet toegespitst op virale en sociale-mediacampagnes of traditionele tv-reclame en advertenties. We kunnen kijken wie er welke invloed heeft op sociale media en wat die mensen kunnen betekenen voor ons merk. Enzovoort.

Al dit soort gebieden hebben hun eigen terminologie en acroniemen. Zo kennen we *Brand Protection* en is *Influencer Marketing* de term voor het spotten en beïnvloeden van invloedrijke personen. Het komt allemaal bovenop de mogelijkheden uit het Google-tijdperk (SEO, linkbuilding, et cetera) en het enterprise application-tijdperk van CRM, ERP en SCM. Mogelijkheden te over, maar dat is ook niet vreemd, want *social* staat voor het leven zelf. De analyse van gedragingen is interessant voor tal van bedrijfstoepassingen en daarom schiet de tooling alle kanten op. Zulk gereedschap aanschaffen begint dus met de vraag wat we ermee willen bereiken.

Altimeter heeft een framework ontwikkeld voor Social Media Analytics, dat zes voor de hand liggende domeinen definieert om businessterreinen als marketing, innovatie en operatie aan doelen te koppelen. Inderdaad, het aloude *Management by Objectives*. In het rapport zien we dat de focus achtereenvolgens ligt op luisteren, inzichten krijgen, metrics bepalen en natuurlijk actie ondernemen. In de verschillende domeinen liggen voorspelmodellen ten grondslag aan de te ontplooien activiteiten.



Vraag 6
Welke stappen zet u op het grotendeels onbetreden pad om een Social Business te worden?

www.sogeti.com/vint/r2q6

10 Om te beginnen heel goed luisteren

Sullivan McIntire van Radian6 zegt dat sociale mediadata een belangrijke rol spelen in de stap van reactieve naar voorspellende analyse. Verweef je sociale mediadata met andere systemen en *'it becomes increasingly possible to make guesses about future behavior'*. Hij geeft drie criteria waaraan deze data dan moeten voldoen:

1 Zijn de data realtime?

Sociale data bewegen zich razendsnel online. De versheid van de data is cruciaal.

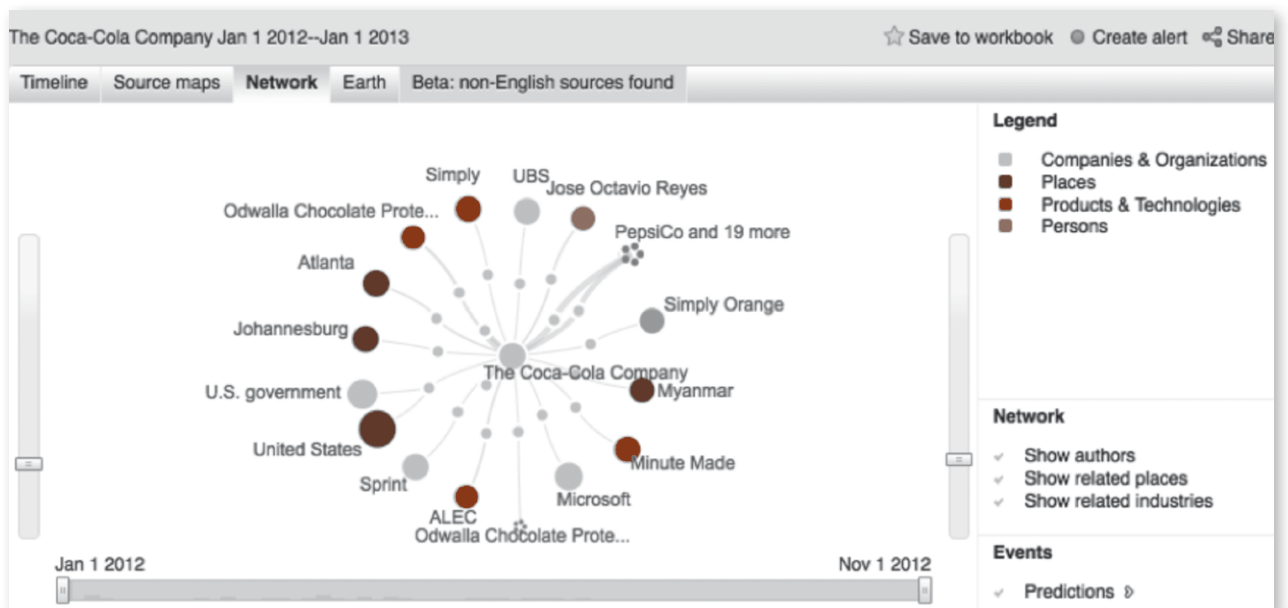
2 Zijn er interessante metadata?

Als we duizenden posts binnenkrijgen en alles met de hand moeten analyseren, dan gaat het moment voorbij. Met een rijke set van metadata kunnen we snel op trends inspelen.

3 Zijn de data geïntegreerd?

Data moeten zo veel mogelijk kunnen worden gekoppeld aan andere relevante bronnen om de juiste acties te kunnen ondernemen.

Om Social Media Analytics te kunnen toepassen, moeten we eerst goed luisteren naar wat er wordt gezegd. Dat doen we met *Enterprise Listening Platforms*. Onze drietrapsraket *Understand-Predict-Act* veronderstelt dat er wordt gezonden, wordt ontvangen en wordt geïnterpreteerd. Daarbij zijn de influencer scores van belang, profiling van de personen die gevolgd worden, een control center om alles bij te houden, netwerk-analyse om verbanden te leggen en sentimentanalyse.

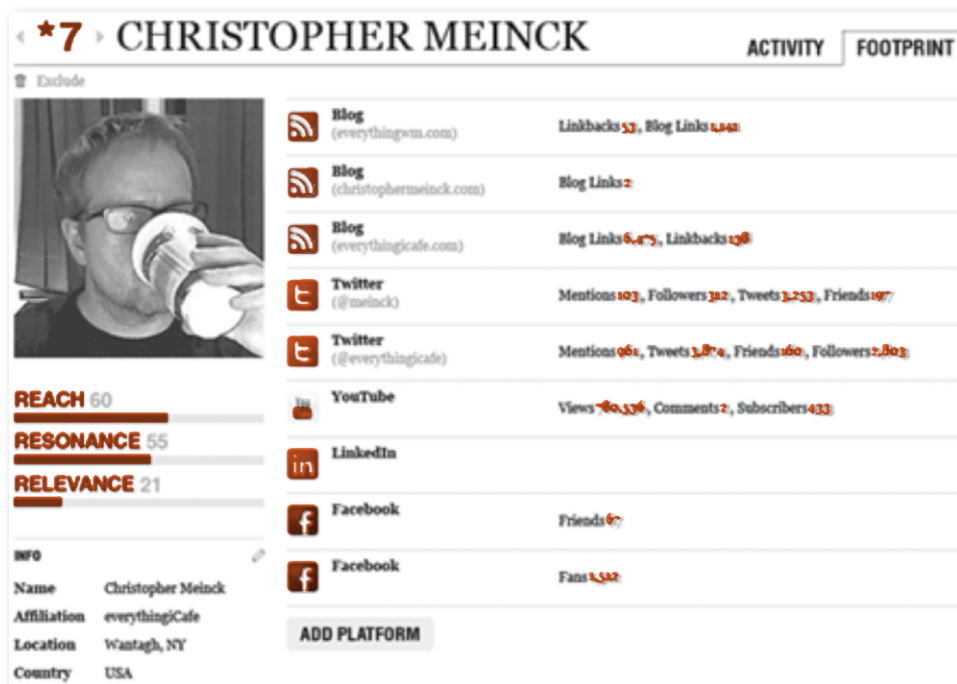


Luisterplatforms

Recorded Future is de naam van een luisterplatform met een hele specifieke rol. Het registreert gebeurtenissen die nog moeten plaatsvinden. Recorded Future wordt gefinancierd door Google en de FBI. De organisatie gebruikt blogs, websites en sociale media als input. Het zoekt op 'next week' en 'next month' en komt met antwoorden op vragen als 'Waar gaat Obama in augustus naartoe?' of 'Welke bierfestivals zijn er in oktober?' Ook de lancering van nieuwe producten en technologieën van concurrenten zijn op die manier te volgen, of gewoon de conversaties over een merk als Coca Cola.

Influencer scores

Er zijn ook platforms die de invloed bijhouden van mensen op sociale media, want influencers moeten we te vriend houden. Twenty Feet heeft zo'n tool en noemt het een 'ego tracking service'. Uiteindelijk krijgt iemand een score, zoals van het sociale netwerk Klout. Rapporten over influencers kunnen behoorlijk gedetailleerd zijn, getuige deze momentopname van Christopher Meinck die met Traackr is gemaakt.



Profiling

Maar er valt veel meer te weten dan of iemand een influencer is of niet. Vrouw of man, autorijder of niet, woonplaats, hobbies, getrouwd, gescheiden enzovoort. Al dit soort info staat op de servers van IBM, dat alle Twitter-accounts ter wereld verzameld heeft en daar de 'bots' uit heeft verwijderd. Vervolgens is er een profiel gemaakt op basis van de inhoud van de tweets en het eigen profiel. Op het moment dat bijvoorbeeld een mediabedrijf wil weten of een trailer voor een nieuwe film in de smaak valt bij een bepaald publiek, kan zo'n Twitter-profiel voorspellende waarde hebben. Een paar minuten na de uitzending van een trailer voor een nieuwe film wist een Ame-

rikaans mediabedrijf dat de beoogde doelgroep niet enthousiast was en ook wie de trailer wel kon waarderen.

Het Social Media Listening Center

Hier zien we een Salesforce Radian6-dashboard op verschillende beeldschermen in een Social Media Listening Center. Voor consumenten is het heel vervelend als een bepaalde service stukt. Via hun luisterplatforms vangen webcare-teams van organisaties dit soort berichten op en gaan ermee aan de slag.



Onder meer KLM werkt op deze manier. Een volgende keer dat iemand twittert, kan direct de vorige conversatie weer worden opgeroepen. Ook geven deze dashboards overzichten per regio. Via taal en land volgt men berichten op een wereldkaart. Sentimenten over de laatste televisiereclame of over de afhandeling aan de balie zijn te volgen op het dashboard in het control center.

Netwerkanalyse

MentionMap is een voorbeeld van een *network analysis tool* voor Twitter-conversaties. MentionMap toont tweets van personen en brengt onderlinge relaties in kaart. Wij voerden als voorbeeld de Twitter-ID in van Ben Lorica, Chief Data Scientist bij O'Reilly Media. Dat leverde de volgende MentionMap op:

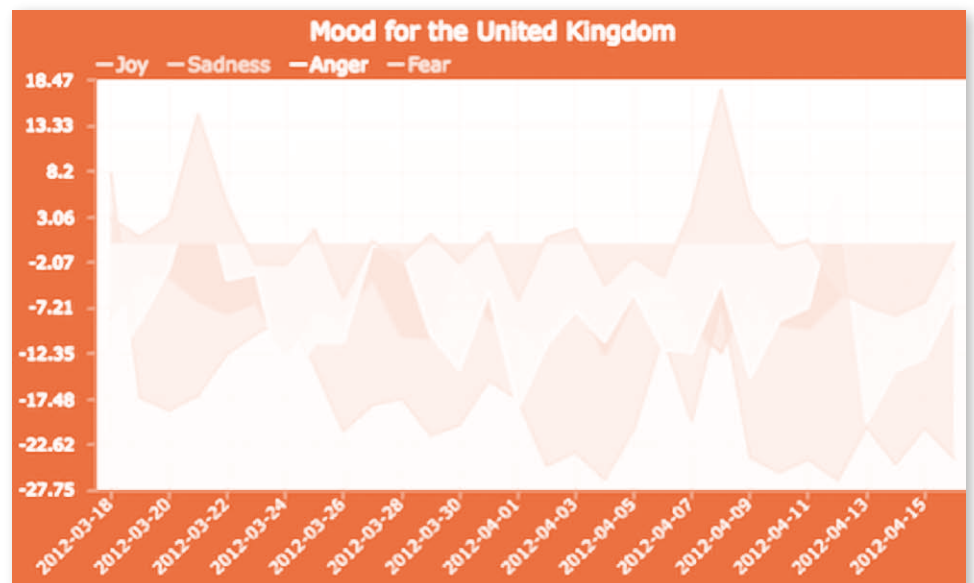


De conversaties die Lorica heeft gevoerd, leiden ons onder andere naar *The Economist*, Microsoft Research (MSFTResearch), Stanford University, Berkeley-student Neil Conway, Rafeboogs en David Wilson.

Sentimentanalyse

FoodMood.in combineert tweets met locaties om over de hele wereld het actuele sentiment te meten rond voedsel. Hier staan die van Nederland gevisualiseerd, met als populaire onderwerpen de pannenkoek, sushi, salade, maar ook kliekjes en worteltjes. Foodmood maakt gebruik van een sentimentclassificatie van Stanford University, een 'getrainde' classifier die miljoenen tweets aankan. Landen kunnen onderling worden vergeleken, er is een top-10 van de *happiest foods* per land en vergelijkingen tussen het GDP van landen en de scores van het voedsel zijn mogelijk.

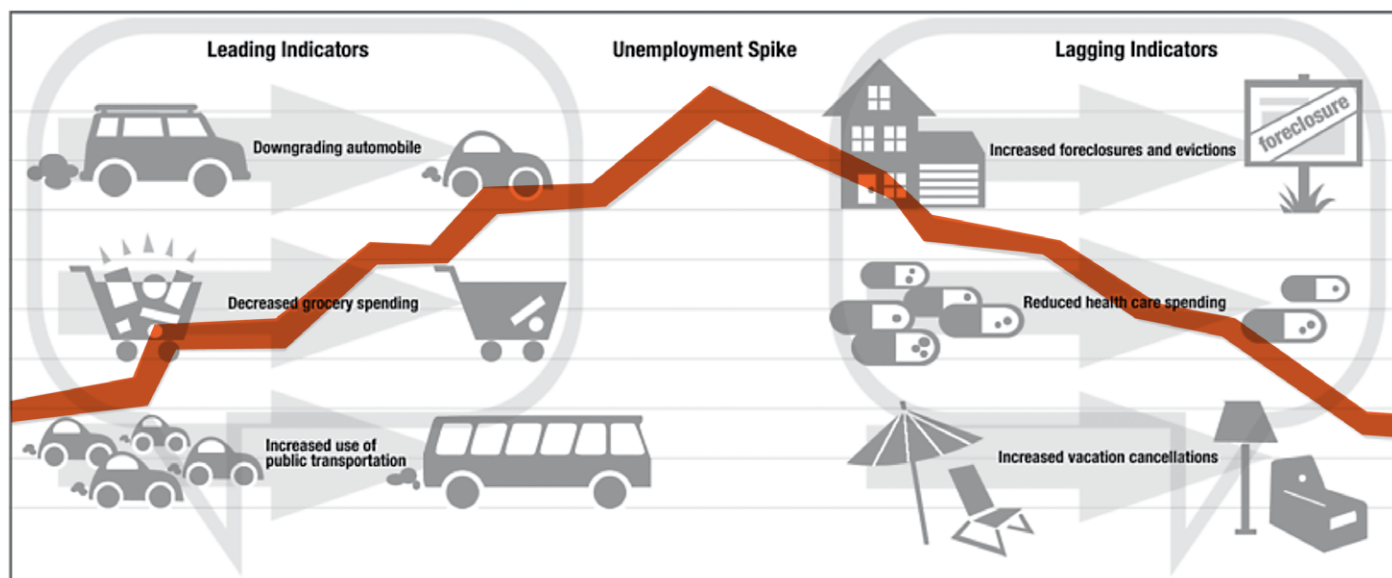
In Engeland meet de universiteit van Bristol emoties op Twitter en maakt daar overzichtjes van:



Angst en droefenis overheersten tijdens de zomeropstanden van 2011 en later ook toen de regering aankondigde fors in de uitgaven te gaan snijden. Aan de School of Informatics and Computing in Indiana is op basis van dit soort stemmingsanalyses een model gebouwd om koersfluctuaties te voorspellen. Onderzocht werd of het publieke sentiment correleerde met de waarde van de Dow Jones Industrial Average. Na een gedetailleerde tekstanalyse op basis van woorden als 'calm', 'alert', 'sure', 'vital', 'kind' en 'happy' was de conclusie dat in 86,7 procent van de gevallen de dagelijkse stijgingen en dalingen in de slotwaarde van de DJIA accuraat kon worden voorspeld.

De Amerikaanse Federal Reserve Bank maakte in oktober 2011 bekend op gelijksoortige manier het consumentenvertrouwen te gaan volgen via Facebook, Twitter, blogs, YouTube, fora, Associated Press, CNN en de *Wall Street Journal*. Het Global Pulse-project van de Verenigde Naties zet sentimentanalyse onder de bevolking onder andere in om de werkloosheid te voorspellen. SAS vond dienaangaande zes indicatoren in uitingen op sociale media, waaronder in een kleinere auto gaan rijden of de vakantie uitstellen.

'SAS compared mood scores and conversation volume with official unemployment statistics to see if upticks in those topics were indicators of spikes in unemployment. The analysis revealed that increased chatter about cutting back on groceries, increasing use of public transportation and downgrading one's automobile could, indeed, predict an unemployment spike. After a spike, surges in social media conversations about such topics as canceled vacations, reduced health care spending, and foreclosures or evictions shed light on lagging economic effects. Such information could be invaluable for policymakers trying to mitigate negative effects of increased unemployment.'

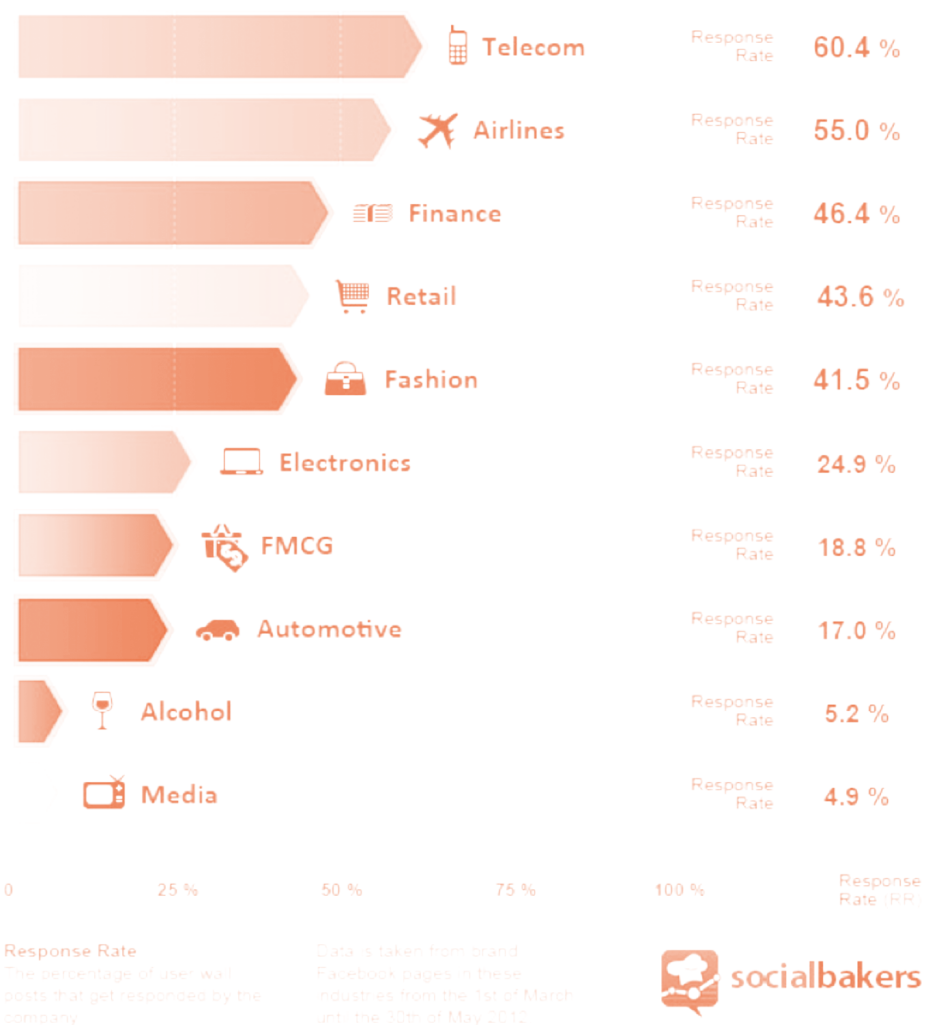


Analysis of social media using SAS shows increases in chatter about certain topics that are leading and lagging indicators of a spike in unemployment.

11 De kracht van Big Social Data

In deze paragraaf zetten we vier concrete Big Social-cases op een rij in een dwarsdoorsnee van verschillende sectoren: retail, banken, verzekeraars, politie en productontwikkeling in de computerindustrie. Misschien is het nog te vroeg om algemene conclusies te trekken over de vraag in welke sector Big Social de meeste impact zal hebben. Als de activiteit van sectoren op Facebook indicatief is voor de voorspelling over de impact, dan zou het rijtje er zo uit kunnen zien:

Most Socially Devoted Industries on Facebook

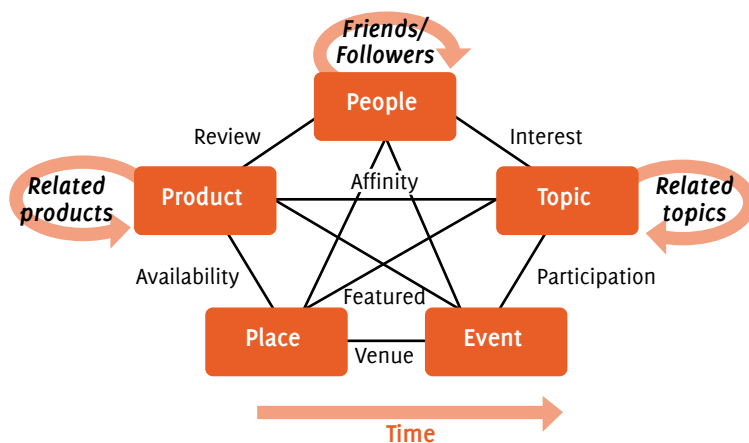


Onafhankelijk van een sectorindeling pleiten we voor een focus op processen en doelstellingen, zoals Altimeter ons al voorhield in het eerder aangehaalde rapport. De cases in deze paragraaf betreffen kostenbesparing, productinnovatie, marketing en

sales. Net als bij de Target-case in paragraaf 3 zien we dat doelbewust opschuiven in de richting van totaal datamanagement tegenwoordig gelijk vruchten afwerpt – laaghangend fruit – wanneer de insteek ‘social’ is. De case van Walmart Labs laat zien, dat voorkeuren op sociale media online en offline direct kunnen worden omgezet in aanbevelingen. De case van het Britse Wonga laat zien, dat Big Data meer bankproducten rendabel kan maken. Voor verzekeringen is het belangrijk om fraude te voorkomen en uitkeringen voor ziektekosten te beperken. Tot slot kijken we naar de rol van Big Social in misdaadbestrijding.

1 Walmart Labs & marketinginnovatie

Walmart Labs, de social-data R&D-afdeling van Walmart, is voortgekomen uit de acquisitie van Kosmix, een social-media-analysebedrijf dat vooral bekend is om de Twitter-filtertool TweetBeat. De belangrijkste daad van het lab is de ontwikkeling van een tool die semantische analyses doet op Twitter, Facebook en Foursquare. Zo wordt het zogeheten *Social Genome* in kaart gebracht, bestaande uit rijke profielen van klanten, onderwerpen, producten, locaties en evenementen.



Het gaat hier om de interpretatie van een netwerk aan de hand van relaties: een persoon is geïnteresseerd in een onderwerp, een persoon is bij een event, een event is gerelateerd aan een onderwerp, een organisatie wordt geassocieerd met een product enzovoort. Walmart maakt hierbij gebruik van publieke data op het web, eigen data en data afkomstig van sociale media.

De eerste resultaten zijn inmiddels zichtbaar. Met data van Facebook en Twitter kunnen bijvoorbeeld op de website en in de winkels betere aanbevelingen worden gedaan. Bijvoorbeeld met de nieuwe Walmart-app Shoppycat. Die doet kadosuggesties aan vrienden als iemands verjaardag nadert.

Join the
conversation

Vraag 7
Welke Social Analytics-
praktijken zou u in het bij-
zonder willen aanbevelen?

www.sogeti.com/vint/r2q7

2 Geld lenen met Big Data

Wonga, een Engels slangwoord voor geld, is een startup die kortlopende kleine leningen aanbiedt zonder menselijke tussenkomst. Deze voor traditionele banken onrendabele markt cultiveerde Wonga op basis van Big Data tot een winstgevende niche. Mensen die geld nodig hebben, spiegelen je van alles voor, maar de harde data liegen niet, aldus Wonga-oprichter Errol Damelin.

Het begon met het experiment SameDayCash en de beta-periode verliep zoals verwacht. Voor elke succesvolle lening, faalde er een ander. Niks bijzonders, maar er werden belangrijke feiten verzameld: over mensen en hun gedrag. SameDayCash voedde het huidige Wonga-algoritme. In het eerste jaar werden 100.000 leningen verstrekt met een gezamenlijke waarde van 20 miljoen pond. Het algoritme bekijkt allerlei informatiebronnen, waaronder sociale media. Er wordt gewerkt met een basis-set van 30 informatiepunten, dat wordt verrijkt met duizenden andere datapunten. Het algoritme legt anomalieën in de relaties bloot. Omdat de leningen kort lopen, is er een constante toevoer van nieuwe data. Financieel is deze aanpak haalbaar, omdat het om kleine bedragen gaat.

3 Fraudedetectie als Big Social killer-app voor insurance

Fraude kostte de Amerikaanse verzekeraar Property & Casualty zo'n 30 miljard dollar per jaar: naar schatting 10 procent van alle schadeclaims. Door historische data over claims te koppelen aan demografische profielen kan de fraudekans al worden ingeschat. Het netwerk van een persoon zegt nog veel meer. Als drie Facebook-vrienden al eens zijn betrapt, dan kan een extra controle geen kwaad. Predictive en Social Analytics helpen zo met rijkere profielen de belangrijkste kostenposten van verzekeraars – fraude en risicoanalyse – te verkleinen. Ook tekstanalyse speelt een grote rol bij het doornemen van schadeformulieren.

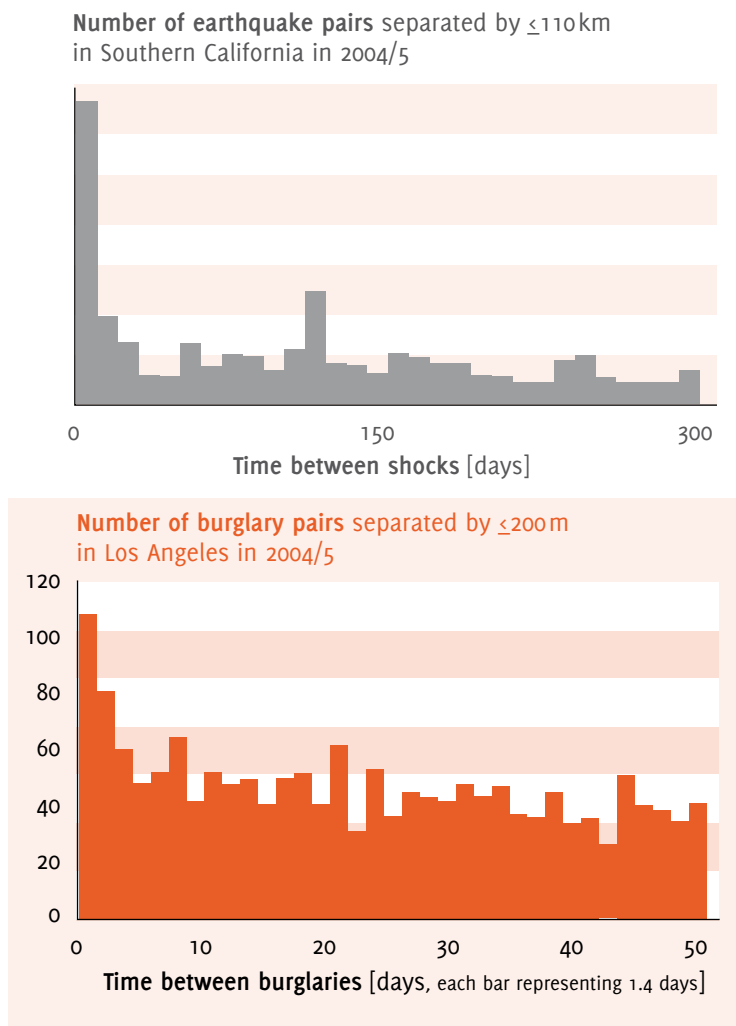
Traditionele risicoanalyses vonden plaats op basis van bijvoorbeeld sociaal-demografische gegevens, rijgedrag en kredietgegevens. Nieuwe datapunten dragen bij aan betere risicoanalyses. Levensverzekeraars passen hun prijzen al aan op basis van medische gegevens die een gezonde levensstijl suggereren. Veel Facebook-likes voor extreme sporten op een profiel zouden straks kunnen leiden tot hogere premies.

4 Misdaadbestrijding dankzij Predictive Police

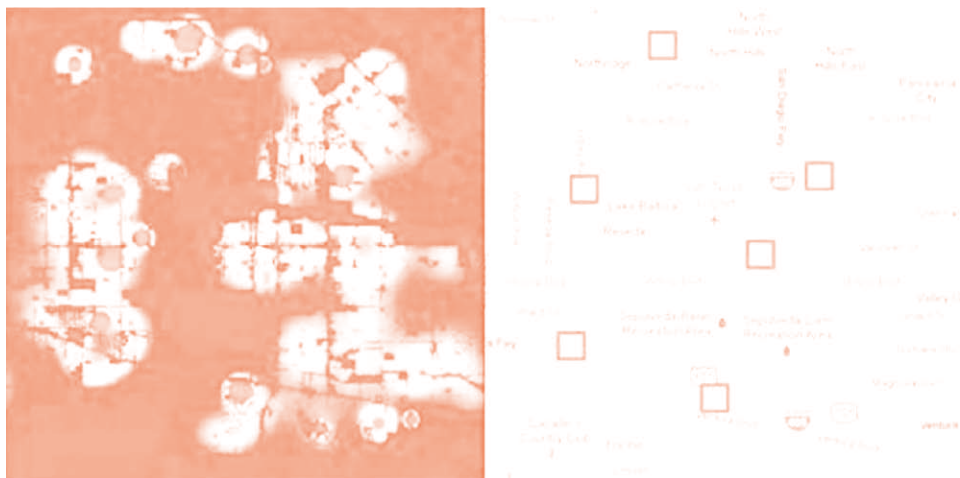
In Nederland gaan politieagenten op pad met een smartphone om in de buurt de signalen uit sociale media op te pikken. Zo kunnen ze al even hun gezicht op een schoolplein laten zien, voordat er iets gebeurt dat niet in de haak is. Dit soort initiatieven wordt gesteund en gedeeld door *politie 2.0*, een community die de Big Social-verandering nastreeft.

In Santa Cruz, Californië, maakt de politie gebruik van een algoritme dat in staat is om misdaden te voorkomen. De software is ontwikkeld door twee wiskundigen, een antropoloog en een criminoloog van het bedrijf PredPol, en gebaseerd op een model dat de naschokken van een aardbeving voorspelt. De praktijk leert namelijk dat een

crimineel vaak terugkeert naar de plek waar hij eerder actief is geweest. *Aftercrimes* volgen hetzelfde patroon als de *aftershocks* van een aardbeving.



Al langer maakten diverse politieafdelingen gebruik van software van CompStat om misdaden te voorspellen. Deze computerprogramma's beschikten slechts over historische data van minimaal een week oud. De PredPol-algoritmes nemen ook realtime data mee. Op basis van locatie, tijd en type misdaad is de software in staat om *prediction boxes* met een nauwkeurigheid van 500 vierkante meter te definiëren.



In het eerste half jaar van de pilot daalde het aantal misdaden met een kwart. Het werkt heel simpel: *'The suspect is showing up in the area where he likes to go. They see black-and-white talking to citizens – and that's enough to disrupt the activity.'*

12 Samenvatting en de organisatie van privacy

De concrete adoptie en plannen in organisaties voor Big Data raken momenteel vooral het thema Big Social: de klantkant kortom, met name geïnspireerd door de social networkactiviteit van Web 2.0. Maar naarmate we het begrip 'social' ruimer opvatten, komt er steeds meer Big Data-potentie vrij. Dat is kort en goed de mars-route die we sinds begin jaren negentig hebben gevolgd: eerst met Web Analytics, vervolgens met Social Analytics en nu met Next-Generation Analytics. In deze tijd van Big Data ligt het voor de hand om door te groeien naar Total Data Analytics en totaal datamanagement.

Een belangrijk deel van de discussie betreft de vraag hoezeer organisaties zich nu in Big Social Data moeten storten. Het antwoord luidt: met beleid. Slim ondernemen in de groeiende datastroom is het devies om zo nieuwe krenten uit de pap mee te kunnen pikken. De vraag of een organisatie dan bezig is met echte Big Data(sets) of niet is eigenlijk irrelevant. Opschalen gaat vanzelf en daar zitten onder meer nog behoorlijk wat privacyvraagstukken aan vast. Die behandelen we uitgebreid in onze derde onderzoeksnotitie. Nummer vier zal gewijd zijn aan een Big Data-roadmap vanuit verschillende invalshoeken.

Moderne Social Analytics-toepassingen stellen organisaties in staat om ritmes van menselijk handelen te begrijpen, daar voorspellingen aan te verbinden en dienstovereenkomstige acties te plannen en uit te voeren: *Understand*, *Predict* en *Act*. De mogelijkheden van personalisatie en hypertargeting nemen steeds meer toe en de gereedheidskist puilt uit. Maar willen klanten dat wel? Het geeft velen op zijn minst een onbehaaglijk gevoel om te beseffen wat commerciële organisaties allemaal over individuen en groepen weten. De organisatie van privacy en de waarborging van onze persoonlijke integriteit is daarom misschien wel het domein bij uitstek om aandacht aan te besteden. Big Data, Big Social en Big Brother liggen dicht bij elkaar – zeker in de menselijke perceptie.

Het begin van deze tweede onderzoeksnotitie staat in het teken van de vraag *What's next in Big Data?* Veel organisaties kijken graag nog even de kat uit de boom, want er zijn nog geen best practices en het financiële risico is daarmee groot. Maar klanten hebben een heel andere mindset. Die zijn voornamelijk gealarmeerd door bijvoorbeeld berichten over premies die stijgen, omdat ze op internet direct of indirect te enthousiast hebben geëtaleerd dat ze aan waaghalzerij doen in hun vrije tijd, of dol zijn op sigaretten en bier. Hoe organisaties ook over Big Data en Big Social mogen denken, de perceptie van de klant zal ons dwingen om er serieus mee om te gaan, standpunten in te nemen en die uit te dragen.

De technologie schrijdt voort, we kunnen steeds beter voorspellen en stappen moeiteloos van Web en Social Analytics door naar Next-Generation Analytics. De nadruk komt steeds meer te liggen op data en algoritmes dan op modellen, kortom: de commerciële kracht van Big Social Data is onmiskenbaar en groeiende. Op zijn minst brengt dit toenemende verantwoordelijkheden en waarborgen met zich mee, waar het gaat om thema's als privacy, persoonlijke integriteit en vooral ook de perceptie en het sentiment daaromheen. Dat is misschien wel de allereerste waarneming die we met beide benen op de grond moeten doen.

Big Data is momenteel nog echt een discussiethema. Op het gebied van technologie, van organisatie, van ROI enzovoort. Daarom blijven we graag met u in contact naar aanleiding van alle *next practices* die momenteel worden ontwikkeld en op waarde getoetst: online op <http://vint.sogeti.com/bigdata> en natuurlijk in persoonlijke gesprekken.

Acht centrale Big Social-definities

Zoals we in dit rapport hebben gezien omvat het domein van wat we gemakshalve Big Social noemen, verschillende vormen van analyse. Daarom is deze paragraaf gewijd aan de vijf belangrijkste onderwerpen op dit gebied. We geven in een logische volgorde acht definities: een voor Predictive Analytics, een voor Sentiment Analysis, een voor Web Analytics, vier voor Social Analytics en een voor Next-Generation Analytics.

Vandaag de dag gaat de meeste aandacht uit naar Social Analytics. We beginnen met de oorspronkelijke en breedste invalshoek van Lars-Henrik Schmidt, gevolgd door de opvatting van Kevin Roebuck over de dominantie van social media en de standaard-definitie van Gartner, en sluiten af met de convergerende visie op sociale informatie en systemen van Mary Wallace. Deze convergerende manier van kijken naar Big Social ligt ook ten grondslag aan dit rapport.

De laatste definitie heeft betrekking op Next-Generation Analytics. Deze definitie is breed genoeg om alle toekomstige Big Data-ontwikkelingen die al expliciet in de definitie van Predictive Analytics genoemd worden, eronder te kunnen scharen.

Predictive Analytics (Wikipedia)

‘Predictive Analytics omvat diverse statistische technieken variërend van modelleren, machinaal leren (machine learning), datamining en speltheorie die huidige en historische feiten analyseren om voorspellingen te doen over toekomstige gebeurtenissen. [...] Invloeden van technologie en Big Data op Predictive Analytics: [...] Het volume, de verscheidenheid en de snelheid van Big Data creëren over de hele linie uitdagingen op het gebied van het vastleggen, opslaan, zoeken, delen, analyseren en visualiseren van data. Voorbeelden van Big Data-bronnen zijn weblogs, RFID- en sensordata, sociale netwerken, zoekindexen voor internet, gespreksgegevensregisters, militaire bewakingsgegevens en complexe data uit de astronomie, biogeochemie, genomica en atmosfeerwetenschappen. [...] Het is nu mogelijk om enorme hoeveelheden gestructureerde en ongestructureerde data te verzamelen, analyseren en exploiteren om tot nieuwe inzichten te komen. Het onderzoeken van Big Data en het gebruiken van Predictive Analytics ligt nu binnen het bereik van meer organisaties dan ooit tevoren.’

Sentiment Analysis (Mejova, 2009)

‘In reactie op de toenemende beschikbaarheid van informele, opiniërende teksten zoals blogs en productrecensies op internet is de afgelopen tien jaar het vakgebied van de Sentiment Analysis ontstaan om antwoord te geven op de vraag wat mensen van een bepaald onderwerp vinden. Sentiment Analysis brengt onderzoekers uit de computerwetenschappen, computerlinguïstiek, datamining, psychologie en zelfs sociologie bijeen en breidt zo de traditionele, op feiten gebaseerde tekstanalyse uit om op opinies georiënteerde informatiesystemen mogelijk te maken. Sentiment Analysis is

nauw verwant aan (of kan beschouwd worden als een onderdeel van) computerlinguïstiek, natuurlijke taalverwerking en text mining. Dit vakgebied, dat voortborduurde op het onderzoek naar de affectieve staat (psychologie) en beoordelingen (appraisaltheorie), probeert met behulp van nieuwe tools uit datamining en computerlinguïstiek antwoord te vinden op vragen die al lang door andere disciplines worden onderzocht. Sentiment Analysis heeft vele namen. Het wordt vaak subjectivity analysis, opinion mining of appraisal extraction genoemd, met enige relatie tot affective computing (herkenning en uitdrukking van emoties door computers). [...] Dit zijn doorgaans losse woorden, zinsdelen of zinnen. [...] Een mening kan op twee manieren in tekst worden uitgedrukt: expliciet, wanneer de subjectieve zin een directe mening uitdrukt (“Het is een prachtige dag vandaag”), en impliciet, wanneer in de tekst een mening doorschemert (“De koptelefoon ging na twee dagen al kapot”). Het meeste werk tot nu toe op dit gebied gedaan, richt zich op de eerste vorm, omdat die gemakkelijker te analyseren is.’

Web Analytics (Wikipedia)

‘Het meten, verzamelen, analyseren en rapporteren van internetdata met als doel het gebruik van het web beter te begrijpen en te optimaliseren.’

Social Analytics (Wikipedia)

‘Social Analytics is een filosofisch perspectief dat vanaf de vroege jaren tachtig van de vorige eeuw is ontwikkeld door de Deense ideeënhistoricus en filosoof Lars-Henrik Schmidt. Het theoretische onderwerp van dit perspectief is socius, een soort “gemeenschappelijkheid” die noch een universeel verhaal noch een gemeenschappelijkheid gedeeld door alle leden van een lichaam is. [...] Je zou kunnen zeggen dat dit perspectief probeert de geschilpunten tussen filosofie en sociologie te articuleren. Social Analytics doet verslag van de tendensen in een bepaalde periode.’

Social Analytics (Roebuck, 2011)

‘Social Analytics heeft betrekking op het bijhouden van diverse vormen van content, zoals blogs, wiki’s, microblogs, sociale netwerksites, websites voor het delen van video’s en foto’s, forums, messageboards en door gebruikers gegenereerde content in het algemeen, als middel voor marketeers om te bepalen hoeveel van dergelijke content er over een merk of onderwerp is en wat de opinies daarover zijn in de social media.’

Social Analytics (Gartner, 2010)

‘Social Analytics beschrijft het proces van het meten, analyseren en interpreteren van de resultaten van interacties en connecties tussen mensen, onderwerpen en ideeën. Deze interacties kunnen plaatsvinden met behulp van sociale softwareapplicaties die worden gebruikt op de werkplek, in naar binnen of naar buiten gerichte community’s of op het sociale web. Social Analytics is een overkoepelende term die een aantal gespecialiseerde analysetechnieken omvat, zoals social filtering, social network ana-

lysis, sentiment analysis en social media analytics. Tools voor social network analysis zijn nuttig voor het onderzoeken van sociale structuur en onderlinge afhankelijkheid en ook de werkpatronen van individuen, groepen en organisaties. Bij social network analysis worden data uit meerdere bronnen verzameld, relaties in kaart gebracht en de impact, kwaliteit en effectiviteit van een relatie geëvalueerd.'

Social Analytics (Wallace, 2011)

'Als we kijken naar de academische definitie van Social Analytics, "het proces van het meten, analyseren en interpreteren van de resultaten van interacties en connecties tussen mensen, concepten en feiten", en deze breder toepassen op de business, dan gebeuren er een paar dingen. Ten eerste wordt het dan mogelijk om sociale inzichten te vergaren uit bestaande enterpriseapplicaties en daarop actie te ondernemen, ten tweede creëren we zo een brug die ons helpt legacy business solutions te koppelen aan het social business-platform van de nieuwe generatie, en ten derde verhogen we zo aanzienlijk de ROI die we kunnen realiseren met onze sociale investering.'

Next-Generation Analytics (Gartner, 2010)

'Het wordt mogelijk om simulaties of modellen uit te voeren om de toekomstige resultaten te voorspellen, in plaats van simpelweg terugkijkende data over interacties uit het verleden te leveren, en om deze voorspellingen in real time te doen ter ondersteuning van elke individuele zakelijke activiteit.'

Literatuur en illustraties

- AMP Lab: UC Berkeley Algorithms, Machines and People Lab, <http://amplab.cs.berkeley.edu/>
- Anderson, C. (2008): 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Barrett, P. (2012): '10 questions CMOS need to ask about social media', <http://www.asterdata.com/blog/2012/04/26/10-questions-cmos-need-to-ask-about-social-media>
- Cellan-Jones. R. (2012): <http://www.bbc.co.uk/news/technology-18812126>
- Center for Economics and Business Research (2012): *Data equity: Unlocking the value of big data*, <http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf>
- Chief Customer Officer Council (2012): 'The Role of the CCO', <http://www.ccocouncil.org/site/the-role-of-the-cco.aspx>
- Conan-Doyle, A. (1892): 'The Adventure of the Copper Beeches', http://sherlockholmes.wikia.com/wiki/Story_Text:_The_Adventure_of_the_Copper_Beeches
- Dachis, J. (2012): 'Big Data Is The Future Of Marketing' <http://www.businessinsider.com/big-data-is-the-future-of-marketing-2012-7>
- Duhigg, C. (2012): 'How Companies Learn Your Secrets', <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
- Economist, The*: 'Special Report: International Banking (2012), Crunching the numbers. Banks know a lot about their customers. That information may be valuable in more ways than one', <http://www.economist.com/node/21554743>
- Economist, The*: 'Insurance data (2012), Very personal finance. Marketing information offers insurers another way to analyse risk', <http://www.economist.com/node/21556263>
- Etlinger, S. (2011): 'Research Report: A Framework for Social Analytics', <http://susanetlinger.wordpress.com/2011/08/10/research-report-a-framework-for-social-analytics>, <http://www.altimetergroup.com/research/reports/a-framework-for-social-analytics>
- Eysenbach, G. (2006): 'Gunther Eysenbach coins the term "Infodemiology" and wins AMIA award', <http://www.ehealthinnovation.org/?q=node/89>
- Fader, P. (2012): *Customer Centricity: Focus on the Right Customers for Competitive Advantage*, <http://executiveeducation.wharton.upenn.edu/resources/upload/Wharton-ExecEd-Customer-Centricity-excerpt.pdf>
- FuturICT: <http://www.futurict.eu>
- Gartner (2010): 'Gartner Identifies the Top 10 Strategic Technologies for 2011', <http://www.gartner.com/it/page.jsp?id=1454221>
- Gartner-plaat pag. 16, <http://www.scribd.com/doc/81893261/February-9-Top-10-Strategic-Tech-Dcearley>
- Global Pulse: <http://www.unglobalpulse.org/>
- Gomes, L. (2012): 'Is There Big Money in Big Data?', <http://www.technologyreview.com/news/427786/is-there-big-money-in-big-data/>
- GraphLab: GraphChi, <http://graphlab.org/graphchi/>

- Hardy, Q. (2012): 'Big Data for the Poor', <http://bits.blogs.nytimes.com/2012/07/05/big-data-for-the-poor>
- Hausmann, V. et al. (2012): 'Developing a Framework for Web Analytics', http://uni-koblenz-landau.academia.edu/PetraSchubert/Papers/1862203/Developing_a_Framework_for_Web_Analytics
- Hickins, M. (2012): 'Taking Small Steps to Big Data', <http://blogs.wsj.com/cio/2012/05/24/taking-small-steps-to-big-data/>
- History of Social Media from 550 BC to 2010: http://www.webanalyticsworld.net/wp-content/uploads/blogger/10-socialMediaTL_05.png
- Hubbard, D. (2007): 'How to Measure Anything: Finding the Value of "Intangibles" in Business', <http://www.hubbardresearch.com/wp-content/uploads/2011/08/TAC-How-To-Measure-Anything.pdf>, <http://howtomeasureanything.com/>
- Ideya (2012): *Social Media Monitoring Tools and Services Report 2012* <http://ideya.eu.com/images/SMMTools%20ReportExcerpts%2009072012Final.pdf>
- McAfee, A., Brynjolfsson, E. (2012): 'Big Data, the Management Revolution', <http://hbr.org/2012/10/big-data-the-management-revolution/>
- McIntyre, S. (2012): 'From Reactive to Predictive Analytics', <http://www.radian6.com/blog/2011/08/from-reactive-to-predictive-analytics/>
- McKinsey Global Institute (2011): 'Big Data: The Next Frontier for Innovation, Competition, and Productivity', http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation
- Mejova, Y. (2009): 'Sentiment Analysis: An Overview', <http://homepage.cs.uiowa.edu/~ymejova/publications/CompsYelenaMejova.pdf>
- MyBuys (2012): 'MyBuys Named the Leader in Personalization for Third Year in a Row', <http://finance.yahoo.com/news/mybuys-named-leader-personalization-third-110000414.html>
- Nerny, C. (2012): 'Point Smartphone, Get Data: IBM to Unveil Augmented Reality App for Retail', <http://data-informed.com/point-smartphone-get-data-ibm-to-unveil-augmented-reality-app-for-retail>
- Nokia Mobile Data Challenge: <http://research.nokia.com/page/12000>
- Police 2.0. The Possibilities of the Digital Revolution for the Dutch Police Force (in Dutch): <http://criminaliteitswijzer.ning.com/>
- Recorded Future - Unlock The Predictive Power Of The Web: <https://www.recordedfuture.com>
- Rezab, J. (2012): '70% of Fans Are Being Ignored By Companies – Now what?', <http://www.socialbakers.com/blog/655-70-of-fans-are-being-ignored-by-companies-now-what>
- Roebuck, K. (2011): *Social Analytics: High-impact Emerging Technology - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*
- SAS (2012): 'Claims Fraud: Prevent fraud before claims are paid', <http://www.sas.com/industry/ins/fraud.html>
- SAS, UN (2012): 'Can a country's online "mood" predict unemployment spikes?', <http://www.sas.com/news/preleases/un-sma.html>

Shaw, W. (2012): 'Cash machine: Could Wonga transform personal finance?', <http://www.wired.co.uk/magazine/archive/2011/06/features/wonga>

Social Media Listening, Monitoring, Measuring, and Management Tools, <http://socialmedia-listening.wikispaces.com/Tools>

Talbot, D. (2012): 'A Phone That Knows Where You're Going', <http://www.technologyreview.com/news/428441/a-phone-that-knows-where-youre-going/>

Ungerleider, N. (2011): 'The Federal Reserve Plans To Monitor Facebook, Twitter, Google News', <http://www.fastcompany.com/1786730/federal-reserve-plans-monitor-facebook-twitter-google-news>

VINT (2007): *Me the Media: Rise of the Conversation Society*, <http://www.MeThemedata.com>

VINT (2012): 'Creating clarity with Big Data', <http://blog.vint.sogeti.com/?p=4559>

Wallace, M. (2011): 'Social Analytics is more than just Social Media...', <http://allthingsanalytics.com/2011/11/04/social-analytics-is-more-than-just-social>

Wallace, M. (2011): 'What's in a name?', <http://allthingsanalytics.com/2011/10/13/whats-in-a-name>

Walmart Labs: 'Social Genome', <http://www.walmartlabs.com/social/social-genome>

Wiki: A Wiki of Social Media Monitoring Solutions, <http://wiki.kenburbary.com/social-media-monitoring-wiki>

Webtrends/DK New Media (2011): 'History of Web and Social Analytics', <http://www.marketingtechblog.com/infographic-history-web-social-analytics/>

Over Sogeti

Sogeti is een toonaangevende speler op het gebied van professionele ICT-dienstverlening, gespecialiseerd in applicatiemanagement, infrastructuurmanagement, high-tech engineering en testen. Sogeti werkt nauw samen met haar opdrachtgevers en maakt technologische innovatie mogelijk dankzij maximale resultaten. Bij Sogeti werken meer dan 20.000 professionals, verspreid over ruim 15 landen en meer dan 100 locaties in Europa, de VS en India.

Over VINT

Alle ontwikkelingen volgen op IT-gebied is voor veel organisaties een zware opgave. Vaak staan nieuwe IT-mogelijkheden immers ver af van het primaire bedrijfsproces. Bronnen die deze ontwikkelingen inzichtelijk en pragmatisch benaderen, door ook het mogelijke gebruik te belichten, zijn dun gezaaid. VINT, het Verkenningsinstituut Nieuwe Technologie van Sogeti, geeft invulling aan die koppeling tussen bedrijfsprocessen en nieuwe IT.

In elke rapportage over een verkenning die het instituut heeft uitgevoerd, zoekt VINT het juiste midden tussen feitelijke beschrijving en beoogde toepassing. Op die manier inspireert VINT organisaties om nieuwe technologie in beschouwing te nemen of zelfs te gaan gebruiken.

Big Social

Gedrag voorspellen met Big Data

www.sogeti.com/vint/r2q1

Vraag 1

Zijn Social Data over drie jaar een drijvende factor in uw organisatie?

www.sogeti.com/vint/r2q2

Vraag 2

Maakt het gedrag van uw klanten het noodzakelijk om u met Big Data bezig te houden?

www.sogeti.com/vint/r2q3

Vraag 3

Op welke manier staat de waarde van Social Data bij u op de balans?

www.sogeti.com/vint/r2q4

Vraag 4

Welke veranderingen brengen Big Social Data in uw organisatie teweeg?

www.sogeti.com/vint/r2q5

Vraag 5

Welke indicatoren (nieuwe klanten, behoud van klanten, klanttevredenheid of andere) gebruikt u om Big Social-investeringen te rechtvaardigen?

www.sogeti.com/vint/r2q6

Vraag 6

Welke stappen zet u op het grotendeels onbetreden pad om een Social Business te worden?

www.sogeti.com/vint/r2q7

Vraag 7

Welke Social Analytics-praktijken zou u in het bijzonder willen aanbevelen?

Doe mee aan onze Big Data-discussie op
www.sogeti.com/vint/bigdata/questions

