BIG DATA & PRIVACY

ritassist®

INHOUDSOPGAVE

INLEIDING

Big Data is hot

De potentie van Big Data voor het veranderen van de manier waarop wij leven en de manier waarop bedrijven georganiseerd zijn is enorm. Volgens kenners is er een ware revolutie gaande. Het belonen van dus bevorderen van goed rijgedrag, het voorspellen van ziektegolven, het voeren van specifiekere advertenties, verzekeringen op maat maken door goed gedrag te belonen: de potententiele toepassingen en welvaartstoename zijn enorm.

Wat is Big Data

Big Data is de term die gebruikt wordt als datasets van gegevens te groot zijn om met reguliere databasemanagementsystemen onderhouden te worden[1]. Bij Big Data gaat het om persoonlijke gegevens (informatie), bij elkaar gebracht en (mogelijk) (deels) veranonimiseerd[2]. Voorbeelden zijn het verzamelen van internetsurfgedrag, het tracken van uw auto (bestemming en rijgedrag) en het verzamelen van klantkaartgegevens.

Doel van Big Data

De verzamelde (vaak) persoonlijke gegevens worden gecumuleerd tot Big Data. Met behulp van (zeer) ingewikkelde model-analyse kan er een “wetmatigheid” of ”voorspeller” uit worden gedestilleerd. Hierdoor kunnen er voorspellingen over bepaalde gebeurtenissen en gedragingen op zowel macro-schaal of micro-schaal (individu) gedaan worden. Ook kan het individuele gedrag worden bekeken. Dit resulteert weer in allerlei toepassingen waarbij markten anders worden ingericht, zoals de automotive markt (connected cars) en de verzekeringsmarkt (bijvoorbeeld door nauwkeurige premie-afstemming op individuele gedragingen van verzekerden).

---

[1] http://nl.wikipedia.org/wiki/Big_data

[2] Het verschil tussen ‘gewone data’ en ‘big data’ is aan te geven volgens de 4 v’s:
Volume (veel data), Veracity (onzekerheid van data), Velocity (snelheid van binnenkomst van de data), Variety (verschillende vormen van data)” – PM Bronvermelding

Privacy- en eigendomsvraagstuk

Aangezien het verzamelen en opslaan van persoonlijke gegevens in principe valt onder de wet op de privacy en daarom beschermd is en ook het eigendom (op basis van de auteurswet) in principe bij de creëerder ligt, moet heel kritisch worden gekeken naar het mogelijke spanningsveld tussen het uitnutten van de mogelijkheden van Big Data en waarborging/beveiligen van de privacy. Echter, zoals later in dit artikel zal blijken, kan deze schijnbare tegenstelling juist ook de smeerolie én de brandstof zijn voor de volle benutting van de potentie van Big Data.

Daarnaast is in het post Snowden tijdperk de bescherming van bedrijfsvertrouwelijke informatie een groot goed. Daarnaast, last but not least, hebben bedrijven ook goede prikkels om privacy te beschermen, om zo de individu überhaupt bereid te vinden om gegevens over te dragen.[3]

Aangezien we nu aan de vooravond staan van de revolutie, is zowel een goede discussie, een goede afweging als uiteindelijk een goed framework nodig om deze balans te vinden, teneinde de voordelen van Big Data te kunnen borgen.

De maatschappelijke discussie nu (medio 2014) is zeer noodzakelijk om bewustwording te creëren rond dit onderwerp en voor de juiste keuzes te zorgen op beleidsniveau en op bedrijfsniveau. Door het houden van de discussie nu, in ieders volle bewustzijn, en daarbij ook bewust wordend dat informatie over eigen gedragingen geld waard is, kan het aanbod van persoonsgegevens toenemen, waar men nu nog zeer huiverig is. Noodzaak is dan dus wel –ik zeg het nogmaals- om een aantal aspecten te borgen.

---

[3] PM [V: artikel 1] noemt nog een aantal andere factoren om persoonlijke gegevens te willen beveiligen.

OVERWEGINGEN EN BELEIDSLIJN

De uitdaging is dus zoals genoemd, om een framework te vinden waarbij de individuele privacy wordt geborgd op een zodanige manier, dat de privacy-angsten worden wegenomen zodat de voorwaarden worden geschapen om de grote voordelen van het gebruik van big data te kunnen benutten.

Iets anders geformuleerd: hoe goed de mogelijkheden van big data kunnen worden benut, hangt af van de mate waarin bedrijven de bijbehorende privacy en security issues kunnen oplossen. Een recent voorbeeld is Volvo die zich daar zeer bewust van is. Daarvoor moet goed gekeken worden naar de kern van de privacy-overwegingen van (individuele) mensen, en hiervoor een voldoende comfortabele maar ook op maat gesneden oplossingsmogelijkheid voor te bedenken die op bedrijfsniveau kan worden geïmplementeerd.

De op maat gesneden oplossingsmogelijkheden zijn nodig omdat per geval de privacy-gevoeligheden kunnen verschillen. Denk bijvoorbeeld dat het recht op privacy, althans een onvoorwaardelijk beroep op dat recht, niet voor elk individu in alle gevallen aanwezig is. Het CPB constateerde al, dat mensen waarde hechten aan hun privacy, maar dat ze in de regel bereid om een deel van hun privacy te ‘verkopen’ of te ‘verhandelen’. Dit verschilt per individu en zal waarschijnlijk ook verschillen soort persoonlijke gegevens.

Met dit in het achterhoofd moet een framework worden ontworpen, dat de mogelijkheid biedt om de individuele optimale balans en keuzen in verschillende situaties te vinden.

Daartoe wordt eerst een korte inventarisatie van de algemene opties voor privacybescherming en security gegeven.

UITWERKING

Opties voor privacybescherming & security

De OECD en de APEC hebben diverse methoden beschreven hoe om te gaan met de security van privacy(gegevens). Bovendien is er een nieuwe EU privacy verordening in de maak. De methoden dienen dus tegen de nieuwe aanstaande EU privacy verordening te worden gehouden, zodat deze als het ware 'compatibel' worden gemaakt en elkaar niet tegenstaan, anders is de uiteindelijke uitkomst niet goed praktisch toepasbaar[4]. De uitkomst zal uiteindelijk resulteren in de zogeheten Optimal Individualized Privacy & Security Matrix OIPSM © in het volgende hoofdstuk.

De methoden van de OECD/APEC zijn: een beperking van het verzamelen van gegevens, doelspecificatie, beperking van het gebruik van deze gegevens, zorgen voor correctheid van de data, beveiliging van de opslagcapaciteit, openheid rond data-verzamelen en data-opslaan, individuele opties als opvragen van de persoonlijke data én het recht om verwijderd te worden, waarborgen en verantwoording, expliciete opt-in, beperking van de opslaglocaties van data, en bescherming tegen kwaadwillend personeel bij bedrijven.

De nieuwe EU-privacyverordening geeft onder andere aan, dat er duidelijk en expliciet toestemming moet worden gegeven voor het gebruik van de privacy gegevens. Verder stelt het, dat een aantal waarborgen worden ingebouwd om de privacygegevens te beveiligen.

Met inachtneming van de EU privacy-verordening, kunnen we de methoden van de OECD/APEC in een bepaalde volgorde neerzetten en vervolgens nader beschouwen. Ik noem dit de basisvereisten/voorwaarde voor beveiliging van privacy.

In essentie komt het op het volgende neer: zeg wat je doet, vraag daarvoor toestemming, doe (enkel) wat je zegt, laat zien dat je enkel doet wat je zegt, borg deze toezegging, en geef aan wat je er aan doet als het fout gaat. Het gaat erom dat er immers tevens om dat er vertrouwen is in een goed werkend framework.

---

[4] Nadeel hierbij is nog wel, dat de privacy verordening nog niet gereed is en nog niet definitief rond is.

Basisvereiste 1: openheid van zaken

Een bedrijf dat data verzamelt dient openheid van zaken te geven en daarbij aan te geven voor welk doel de privacy gevoelige gegevens en de daaruit voortvloeiende big data gebruikt worden. Bedrijven moeten openheid van zaken geven, om zo het risico te verkleinen dat de data worden gebruikt met een onverenigbaar doel. Ook dient te worden aangegeven hoe de gegevens getransporteerd worden en hoe zij opgeslagen worden.

Overigens kunnen bij dataverzameling meerdere bedrijven betrokken zijn. Denk bijvoorbeeld bij surfen over internet dat zowel de internet service provider, de telecomaanbieder van de actieve laag én website (bijvoorbeeld Google) de data in principe zouden kunnen uitlezen.

Vragen die door de bedrijven (en ik bedoel daarmee de 'verzamelbedrijven') moeten beantwoord worden zijn:

- Welke data/informatie/gegevens worden opgeslagen?
- Hoe wordt deze data getransporteerd?
- Hoe wordt de data opgeslagen?
- Welke data/gegevens worden gebruikt voor analyses?
- Waarvoor worden de gegevens gebruikt, wat is het doel hiervan?
- Wordt het verwerkt tot bepaalde informatie? En zo ja, welke informatie?
- Voor wie is deze informatie, al dan niet veranonimiseerd[5], beschikbaar?
- Hoe wordt dat gebruikt voor feedback richting mij?
- En, meer algemeen: wat krijg ik er voor terug?

[5] Veranonimiseren: ontdaan van specifieke, individuele karakteristieken waardoor enkel direct de gegevens vallen terug te leiden op 1 individu.

De bedrijven dienen de privacy en security op de juiste manier zeker te stellen. Ze moeten dit aantoonbaar borgen.

Over wijzen van encrypted transport en opslag wordt in dit document niet (nog niet) in detail ingegaan. Tegen misbruik van gegevens, maar door inzet van de-identificatie methodes vergroten ze het risico dat de voertuigbezitters alsnog geïdentificeerd kunnen worden. Dat is een privacyrisico.

Basisvereiste 2: toestemming en controle

Alle bedrijven dienen de eigenaren van privacybronnen toestemming te vragen voor het gegevensgebruik. Dit moet volgens de nieuwe Europese privacyverordening expliciet worden gevraagd en er moet expliciet toestemming voor worden gegeven.

Er kan gedifferentieerd worden naar onderdelen van gegevens, dan wel naar wijze van opslag. Dit verschilt per soort gegevens. Zaak is wel om het niet te gecompliceerd te maken. Dat is niet praktisch, wekt geen vertrouwen, en is in strijd met privacyverordening die gewoon taalgebruik voor schrijft.

- Geeft u toestemming voor opslag en gegevensverwerking? (transport, opslag, verwerking/analyse, gebruik)
- Onder welke beveiligingsvoorwaarden?

  Denk hierbij aan de inzet van de-identificatie, intern bedrijfsbeleid bij de bedrijven, gebruik van beveiliging in verbindingen, etc.
- Wie mag deze gegevens inzien en gebruiken? Enkel de verzamelaar zelf, of ook anderen? En zo ja dit laatste, welke niet? En dit alles tegen welke voorwaarden?

Dit laatste kan lastig zijn, want tussen verzameling van aan de ene kant en verwerking en gebruik aan de andere kant kan een lange tijd zitten. Ook is van te voren niet duidelijk waar het gebruik precies voor is. Als niemand de koopgegevens van de consumenten in mocht zien, was de ‘griepvoorspeller’ nooit uitgekomen.

- Onder welke andere, eventuele voorwaarden? Wat wil de consument daarnaast NIET?

Basisvereiste 3: waarborgen

De bedrijven dienen de privacy en security op de juiste manier zeker te stellen. Ze moeten dit aantoonbaar en verifieerbaar borgen. De bedrijven informeren de consumenten hoe zij en hun werknemers verantwoordelijk zijn bij het gebruik van data.

Over wijzen van encrypted transport en opslag wordt in dit document niet (nog niet) in detail ingegaan. Tegen misbruik van gegevens, maar door inzet van de-identificatie methodes vergroten ze het risico dat de voertuigbezitters alsnog geïdentificeerd kunnen worden. Dat is een privacyrisico.

Ook noodzakelijk zijn voorwaarden voor werknemers/gebruikers (wie kan er bij, wie heeft inzage, geen data mee naar buiten nemen). Hier speelt certificering een grote rol.

Basisvereiste 4: verantwoordelijkheid/accountability

De bedrijven dienen volledige verantwoordelijkheid te nemen voor de juiste gang van zaken. Daarnaast dienen zij op een vooraf afgesproken manier te worden afgerekend op eventuele schendingen, bijvoorbeeld een bepaald bedrag ter compensatie, persoonlijke vervolging (strafrecht).

Al deze zaken moeten in een beslissingschema worden neergezet. Dit gebeurt in het volgende onderdeel.

De onderstaande matrix biedt een overzicht van de keuzes die de individuele consument moet maken. Hij moet dit voorafgaand het afgeven van persoonlijke informatie doen, zo legt de nieuwe (en huidige) Europese privacyverordening ons uit.

Kernvraag: geeft u toestemming om uw gegevens te verzamelen (transport en opslag) en vervolgens verwerken (veranonimiseerd) en te gebruiken (individuele aanbiedingen)?

Transport: NEE /

Ja mits: beveiligd / versleuteld/

KEUZES VOOR DE INDIVIDUELE CONSUMENT

Optimal Individualized Privacy & Security Matrix OIPSM ©

Wel/niet door derden

Opslag: NEE /

Ja mits: beveiligd / versleuteld/ veranonimiseerd/

Wel/niet door derden

Verwerking (algemene analyse en verbanden): NEE /

Ja mits: beveiligd / versleuteld/ veranonimiseerd/ En niet voor de … sector.

Wel/niet door derden

Gebruik (individuele aanbiedingen): Nee /

Ja mits beveiligd/Versleuteld (graad van versleuteling/ En niet voor de … sector.

Algemene uitzonderingen:

Geen retargeting

Geen verslechtering van mijn positie t.o.v. een anonieme gebruiker.

Met het boven beschreven framework én de wijze van aanbieden aan de consument, wordt de mogelijkheid geboden om de individuele optimale balans en keuzen in verschillende situaties te vinden. Daarmee kunnen de voordelen van Big Data maximaal worden benut. Wat hier nog niet op ingegaan is, is hoe precies de beveiliging etc. er uit ziet. Deze behoefte zal mogelijk wel bestaan bij de mensen die de gegevens leveren.

Bedrijven hebben zoals gezegd naast de regelgeving ook goede prikkels om privacy te beschermen, om zo de individu bereid te vinden om gegevens over te dragen. Dit systeem werkt bovendien beter dan regelgeving. Regelgeving is traag en loopt achter de feiten aan, zeker in een dynamische omgeving. Goede prikkels werken beter dan stijve regels.

VOORLOPIGE CONCLUSIE

Aangezien het gebruik van big data enorme voordelen oplevert en we ze pas net zijn begonnen te ontdekken, zullen bedrijven bereid om voor de security van gegevens te betalen en om deze waarborgen te verzorgen –zeker bij het groter worden van de big data markt.

Er zal daarnaast een nieuwe markt ontstaan voor advisering over en implementering van beveiliging van persoonsgegevens.

In een aantal eerder aangehaalde uitspraken van diverse mensen in de sociale media, blijkt dat er nogal wisselend wordt gedacht over het benutten van de mogelijkheden van Big Data. Het komt er op neer, dat zij de voordelen van Big Data –for the greater good- niet laten opwegen tegen hun eventuele individuele gevolgen. Soms om principiële redenen. Een ondertoon die daarbij gevoerd wordt, is dat het de solidariteitsgedachte van het verzekeringsysteem zou ondergraven. Dit is echter nogal eenzijdig en kortzichtig bekeken.

EEN LAATSTE OPMERKING

Over solidariteit

Hieronder wordt geanalyseerd welke gedragseffecten optreden en wat dat betekent voor “solidariteit”.

Geen voordelen Big Data als er geen gegevens worden opgeslagen en geanalyseerd

Er zal altijd een categorie (blijven) bestaan die optie X kiest in het OIPSM. Dit betekent, dat zij niet willen dat hun gegevens op enigerlei manier wordt gebruikt. Er wordt niks verwerkt, niets opgeslagen, niets geanalyseerd, etcetera. Het moge duidelijk zijn dat als iedereen dit doet (of heel veel mensen), er dan geen Big Data is/kan zijn, dus dat de voordelen ervan ook niet benut kunnen worden. Er zal dus geen welvaartstoename zijn.

Diverse effecten als niet iedereen gegevens ter beschikking stelt voor Big Data

Daarnaast zullen er mensen zijn die om die reden wel veranonimiseerde gegevens willen laten gebruiken, maar dat zij dan niet willen dat zij er zelf slechter van worden dan iemand die niet zijn of haar gegevens laat opslaan etc.

We beschouwen nu de situatie dat een deel van de mensen de gegevens geheel niet wil laten gebruiken dan wel dat zij niet slechter willen worden van mensen die de gegevens in het geheel niet laten gebruiken.

Eerste effect: premiestijging voor de ‘minder goede’ risico’s

We gaan in eerste instantie uit van een gemiddelde verzekeringspremie voor iedereen. We kunnen er van uit gaan dat degenen die een ‘goed’ risico hebben (op basis van gedrag!) en daarom relatief goedkoper uit zouden zijn als zij individueel beoordeeld zouden worden, eerder geneigd zijn om hun gegevens te delen en de voordelen daarvan te laten terugkomen in individuele aanbiedingen. Dat houdt ook in dat verzekeraars weten wie de ‘minder goede’

risico's zijn. Als de goede risico's een betere premie krijgen, moeten verzekeraars om hun verzekeringspremie voor de 'minder goede' risico's aanpassen om hun winst op peil te kunnen houden. Deze populatie is nu bekend: namelijk degenen die hebben aangegeven niet hun gegevens te willen laten gebruiken, dan wel de personen die niet slechter willen worden dan mensen die de gegevens in het geheel niet laten gebruiken. De verzekeraar kent tenslotte de totale populatie. Hierop kan dan de nieuwe, gemiddelde premie worden aangepast.

Dit effect zal zich een aantal keren herhalen, omdat bij een hogere premie de groep relatief 'goede' risico's ten opzichte van de verzekeringspremie iets groter wordt. Uiteindelijk blijft een kleine groep slechte risico's over op basis waarvan de standaard verzekeringspremie wordt bepaald. De rest heeft dus te maken met een geïndividualiseerde premie op basis van hun gegevens. Het zogeheten adverse selection probleem wordt hiermee deels gecounterd.

Tweede effect: gedragseffecten bij de goede risico's

De groep 'goede risico's' stelt dus zijn gegevens beschikbaar. Omdat zij nu vrij kort gevolgd worden, zullen zij zich (nog meer dan ze al deden) inspannen om hun verzekeringspremie naar beneden te krijgen. Er ontstaat dus een gedragseffect bij de goede risico's: zij vertonen beter gedrag zoals voorzichtiger met auto rijden, op hun voeding letten, sporten, etc. Hierdoor kan de premie nog verder voor hun omlaag gaan. Ook het moral hazard probleem wordt dus aangepakt.

Derde effect: gedragseffecten bij de minder goede/slechte risico's

De groep 'minder goede' risico's en de 'slechte' risico's echter kunnen nu ook worden geïdentificeerd. Zij worden door het naar buiten komen van de overige risico-groepen als het ware ook ontmaskerd. Dit betekent dat zij een andere premiebenadering krijgen. Echter, dit hoeft niet negatief te zijn. Nu ook zij op een indirecte wijze gevolgd worden, zullen zij zich ook inspannen om hun verzekeringspremie naar beneden te krijgen. Dit betekent, dat na verloop van tijd ook deze categorie 'beter' gedrag zal gedragen. Voor hen zal het echter de grootste inspanning vergen, aangezien zij niet eerder uit zichzelf/intrinsiek gemotiveerd waren.

Desalniettemin zullen zij geprikkeld worden en hun gedrag aanpassen, uiteindelijk voor hun eigen voordeel. Ook voor deze groep wordt het moral hazard probleem dus aangepakt.

Dit laatste effect is verreweg het grootste, belangrijkste effect, en zal leiden tot een enorme kostenbesparing op macro-economisch niveau en waarschijnlijk ook welzijn voor mensen. Waar allerlei campagnes voor gezond leven zich op stuk bijten, wordt nu via big data een nieuw mechanisme gecreëerd wat waarschijnlijk effectiever is.

De gedragseffecten leveren een enorme welvaartswinst op, aangezien adverse selection and moral hazard sterk wordt tegengegaan. Deze effecten zullen binnen tientallen jaren honderden miljarden vermijdbare kosten uit de economie kunnen halen door een eerlijkere, nauwkeurige bepaling van risico's en kosten en een gedragseffect bij de actoren.

Wij kunnen, met alle problemen die wij op de aarde continue op te lossen hebben, waaronder aanpassingen en tegengaan van klimaatproblemen, deze potentiële winst simpelweg niet negeren.

Wat is nu: solidariteit?

Is het solidair om te verwachten dat iemand die gezond eet, beweegt etc., beperkt mag worden in zijn beloning om ook een betere verzekeringspremie af te sluiten, ten koste van diegene die dat niet doet en daarom een slechtere verzekeringspremie heeft?

Of is solidariteit zoiets als: we betalen liever een algemene premie, waar iedereen onder valt, dus ook de minder goede risico's, zodat de individuen met goed gedrag betalen voor slecht gedrag.

Het lijkt me dat het laatste niet de voorkeur moet hebben, ook niet omdat de gedragseffecten in het vorige onderdeel genoemd niet kunnen plaatsvinden.

In het bovenstaande verhaal wordt er vanuit gegaan dat alle risico's en dergelijke beïnvloedbaar zijn. Dit is natuurlijk –helaas- niet zo. Gedrag is één, maar vervelende aanleg en andere 'domme pech' is iets heel anders. Hoe werkt dit door in de solidariteitsvraag?

Het antwoord lijkt redelijk simpel. In de analyses van big data zal uiteraard een toevalsfactor zitten. Dit betekent, dat de voorspelkracht vrij groot kan zijn, maar nooit helemaal volledig. Iemand die heel gezond eet, goed sport etc. maar toch een aandoening krijgt, zal wat dat betreft ook niet in de vooraf betaalde premie ‘afgerekend’ worden. Kortom: goed gedrag wordt beloond, maar –ondanks dat- blijft eventuele domme pech daarbij verzekerd.

CONCLUSIE

Hoewel privacy en Big data op gespannen voet lijken te staan, is juist het borgen van de privacy op een voor elke consument voor hem of haar passende wijze essentieel. Het zorgt ervoor dat gegevens mogen worden gebruikt, mits onder bepaalde voorwaarden dus.

Deze voorwaarden moeten vooraf bekend worden gemaakt en moeten duidelijk worden geborgd. Denk daarbij bijvoorbeeld aan het veranonimseerd opslaan en het encrypted transport. Per bedrijf dat gegevens transporteert etc. moet hier een duidelijk protocol voor komen, zowel vanwege regelgeving als vanwege eigen voordelen. Als Big Data tot volle wasdom kan komen, dan zijn de mogelijkheden extreem groot. Bovendien zijn de gedragseffecten extreem groot, met miljarden aan onnodige kosten die uit de economie kunnen worden gehaald door een eerlijkere, nauwkeurige bepaling van risico's en kosten en een gedragseffect bij de actoren.

Deze potentiële winst hoeft niet gelaten te worden vanwege het vraagstuk van solidariteit. Solidair zijn betekent niet dat je een ander niet zou mogen prikkelen om zich ook in te spannen om verantwoordelijker gedrag te vertonen. Sterker nog: als je echt solidair bent, dan zou je dat juíst moeten doen.

--