Gids door het Big Data & Privacy Labyrinth

Auteur: Frank van Vonderen. Verdonck, Klooster & Associates

Big Data en Privacy hebben een bijzondere relatie met elkaar. Maar hoe zou je deze kunnen omschrijven? Opposites attract? Yin en Yang? Water en vuur? Vrijende egeltjes?

Feit is dat vanuit privacy perspectief vragen worden gesteld waar Big data mensen knorrig van worden zoals: welke persoonsgegevens heb je allemaal echt nodig, wat wil je er eigenlijk mee doen, wie hebben er eigenlijk toegang tot de gegevens. Het antwoord dat Big data mensen dan het liefst zouden geven is: “dat weten we nog niet exact, hangt ervan af”.

De big data mensen stellen op hun beurt vragen aan privacy mensen zoals: wat mag er nu wel en niet van de AVG. Het antwoord dat privacy mensen dan het liefst zouden geven is: “dat weten we nog niet exact, hangt ervan af”.

De meeste discussies die ik in de praktijk heb gezien verlopen chaotisch en zonder duidelijke uitkomsten. Maar ik merk ook dat wel een structuur kan worden gebracht in deze discussie: bij privacy mensen zie ik discussies aan de hand van 4 thema’s. Bij de big data toepassingen zie ik 3 fases. Ik licht hierna allereerst de thema’s en fases kort toe en daarna breng ik ze met elkaar in verband.

Wat zijn de hoofdfases van een Big Data initiatief?

Eerst de drie fases van Big Data:
Fase 1 – exploratief onderzoek. In deze fase heeft de organisatie nog geen duidelijk beeld welke data onderdeel is van big data en welke persoonsgegevens waarvoor worden gebruikt. Doel van deze fase is te kijken waar de waarde van de data zit en welke data wel of niet nuttig is.

Fase 2 – werken met hypotheses. In deze fase weet de organisatie al beter welke data wel en niet nodig is en wordt gerichter gekeken of de gewenste beleidsdoel-stellingen met de data worden behaald.

Fase 3 – exploitatie. In deze fase heeft de organisatie een helder beeld van wat zij met de data beoogt en nastreeft en levert de data gewenste en acceptabele resultaten op.

Wat zijn de hoofdvragen vanuit privacy perspectief?

Deze vallen uiteen in 4 hoofdthema’s:
Thema 1 – legitimiteit en transparantie. Is het wettelijk gezien wel passend om data voor analysedoeleinden te gebruiken (doelbinding en grondslag) en wordt hier helder over gecommuniceerd naar de betrokkenen(transparantie). De betrokkenen zijn de mensen wiens gegevens worden gebruikt.

Thema 2 – minimalisatie. Welke data is eigenlijk nodig? Wordt data die niet wordt gebruikt ook niet in het big data magazijn opgenomen. Wordt deze na gebruik verwijderd? Wordt gebruik gemaakt van anonimisering of pseudonimisering, technieken om persoonsgegevens te vervangen door gegevens die geen persoonsgegevens zijn.

Thema 3 – bescherming. Op welke manier wordt de toegang tot de (persoons)gegevens afgeschermd en op welke manier is er een controle op het gebruik van de gegevens (logging op gepast en afwijkend gebruik).

Thema 4 – gebruik en beheer. Op welke manier wordt verantwoording afgelegd over het gebruik van de gegevens en worden de effecten van wijzigingen (toevoegingen, andere zoeksleutels, et cetera) beoordeeld.

Waarom zijn de 3 fases en de 4 thema’s nu zo belangrijk: omdat in discussies over privacy en big data de fases en de thema’s door elkaar lopen. Door de 3 fases en thema’s kun je de wereld ‘in hokjes stoppen’ en zo een gestructureerde discussie voeren: zo wordt zichtbaar dat per fase in Big Data een ander risicoprofiel geldt ten aanzien van privacy-wetgeving. Als je de 3 big data fases en de privacy 4 thema’s tegen elkaar afzet, krijg je de volgende structuur in het labyrinth (of ‘praatplaat’, zo je wilt): *zie volgende pagina.*

"Dat weten we nog niet, hangt ervan af"

VERDONCK KLOOSTER & ASSOCIATES

Factor	Fase 1 Formeel: Exploratief Informeel: 'Grasduinen' Doel: Beleidsonderzoek	Fase 2 Formeel: Werken met hypotheses Informeel 'Gericht onderzoeken' Doel: Beleidsonderzoek	Fase 3 Formeel: Exploitatie / Business intelligence Doel: Operationele sturing
Legitimiteit - Doelbinding grondslag - Transparantie	**- Legitimiteit kan niet goed worden onderbouwd.** Er is nog geen goed idee wat met de data / analyse wordt nagestreefd	**? De legitimiteit wordt geleidelijk duidelijk** Langzaam ontstaat een scherper idee wat je met de data / analyse kunt gaan nastreven	**+ + Legitimiteit kan uitstekend worden onderbouwd** Er is een helder idee wat je met de data / analyse nastreeft
Minimalisatie - Welke gegevens zijn echt nodig - Kan pseud / anon worden toegepast	**- - Minimalisatie is niet gewenst** Zo veel mogelijk data is nodig om patronen te zoeken	**+ / - Geleidelijk is minimalisatie mogelijk** Er is nog steeds veel data nodig om hypotheses te toetsen. Stap voor stap kan onnodige data tijdens onderzoek worden verwijderd. Er kan gericht worden gepseudonimiseerd of geanonimiseerd	**+ + Minimalisatie is uitstekend mogelijk** Je gebruikt alleen de data die nodig zijn om analyse te doen / algoritme te voeden. Onnodige data is verwijderd. Anonimisering of pseudonimisering is optimaal toegepast
Bescherming - Toegang - Controle op gebruik	**+ + Uitstekend mogelijk** Alleen een selecte groep experts hebben gecontroleerde toegang en er controle op gebruik	**+** Er is sprake van zorgvuldige toegang en controle op gebruik. Net zoals je bij andere vertrouwelijke data doet	**+** Er is sprake van zorgvuldige toegang en controle op gebruik. Net zoals je bij andere vertrouwelijke data doet
Gebruik en beheer - Verantwoording - Uitlegbaarheid - Beheer, ontwikkeling, evaluatie en gebruik	**? nog niet bekend** Gebruik vindt plaats in een separate of geïsoleerde omgeving met eigen / virtuele machines of in een pilot achtige opzet	**+ / -** Het beeld over verantwoording en uitlegbaarheid wordt scherper. Er ontstaat zicht op passend gebruik.	**+ +** Geen enkele beperking voor verantwoording en uitleg-baarheid. Er zijn sturings-processen voor gebruik, beheer en ontwikkeling

Wat lees je nu uit deze tabel af?

Bij Fase 1 - exploratief onderzoek

Als je deze kolom leest, dan zie je per privacy thema een korte toelichting staan. Hieronder wordt de toelichting iets verder uitgewerkt.

Legitimiteit – De privacy wetgeving gaat er van uit dat je een goede reden moet hebben om persoonsgegevens te gebruiken. Dit wordt grondslag genoemd. Maar in deze fase ben je nog niet exact in staat om te duiden wat de analyse gaat opleveren. Dan is de grondslag ook niet goed te bepalen. Er is helaas geen grondslag: 'handig', of 'dat zien we later wel'. Kortom: naar de letter van de wet heb je geen grondslag. Betekent dat dat je niet aan business intelligence kan doen? Dat is een afweging die je als organisatie moet maken, maar maak deze niet licht: als je aan business intelligence gaat doen in deze fase, ben je dan bewust van de risico's en beheers deze.

Minimalisatie – De analist heeft nog geen idee welke data exact nodig is. De gangbare strategie in deze fase is om zoveel mogelijk te verzamelen om patronen te kunnen gaan zoeken. Later wordt wel gekeken waar de patronen toe leiden en of alle data nodig was. Bij dit type onderzoek is geen sprake van minimalisatie, maar juist het tegenovergestelde. Namelijk zo veel mogelijk verzamelen: data hoarding.

Bescherming – Er wordt heel veel gewerkt met persoonsgegevens, waarvan het nut nog niet duidelijk is en dus de risico's op onterecht gebruik groot zijn. Om dit risico te controleren wil je dat alleen maar enkele kundige mensen, experts, toegang hebben tot deze data. Bij wijze van spreken in een kluis. Ook wil je zeker weten dat je achteraf kunt bewijzen dat alleen de paar experts met deze data hebben gewerkt. Kortom: afscherming van fysieke en logische toegang en logging op passend en ongewenst gebruik. Deze laatste twee zijn ook specifieke speerpunten van de privacy toezichthouder.

Gebruik en beheer – In deze fase is eigenlijk nog niet duidelijk hoe de data exact gaat worden gebruikt en beheerd. Processen rondom beheer en transparantie zijn nog niet nodig. Gegevens worden geanalyseerd in vaak separate (virtuele) omgevingen, en dus los van de productie omgeving of data warehouses. Er is geen update verbinding met bronnen.

Bij Fase 2 – werken met hypotheses

Ook bij deze fase is in de kolom van boven naar beneden een korte toelichting opgenomen. Een iets uitgebreidere toelichting:

Legitimiteit – Langzaam krijg je een beeld wat de analyse gaat opleveren. Ook begint duidelijk te worden hoe de verwerking aansluit op je bestaande grondslag of dat een legitieme grondslag kan worden verkregen. Dit is ook het moment om na te denken hoe je tegemoet komt aan je transparantieverplichtingen en uit te leggen dat je deze gegevens gebruikt (en waarvoor).

Minimalisatie – De analist heeft intussen een aardig beeld van welke gegevens wel nuttig zijn en welke gegevens niet nuttig bleken. De gegevens die niet nuttig waren zijn verwijderd, of ontdaan van persoonsgegevens. Over de gegevens waarover nog twijfel is, wordt later een beslissing genomen. Daar waar mogelijk worden persoonsgegevens gepseudonimiseerd of geanonimiseerd.

Bescherming – Om de hypotheses te toetsen zijn meer mensen nodig. Deze mensen gaan zorgvuldig om met de (persoons)gegevens die ze hierbij gebruiken. Er is sprake van passende toegangscontrole (need-to-have rechten, logging).

Gebruik en beheer – Langzaam begint zich het beeld te ontwikkelen hoe je de data straks kunt gaan gebruiken, wat de mogelijkheden zijn en wat de beperkingen. De tijd van de speeltuin is voorbij, het is tijd voor de eerste protocollen, die later verder moeten worden aangevuld.

Fase 3 – exploitatie

Tenslotte is ook voor deze fase een korte toelichting in de tabel opgenomen. Wat wordt met de korte toelichting bedoeld:

Legitimiteit – De grondslag is onkreukbaar. Als je niet zeker wist dat je de data zou mogen gebruiken was je niet deze fase beland. Je kan volledig transparant zijn (intern en extern) over wat je met deze data doet en je hebt een antwoord voor iedere criticaster.

Minimalisatie – Je weet precies welke data je nodig hebt en welke niet. De data die je niet nodig hebt, heb je ook met precisie uit de bestaande omgeving en uit je archief weggehaald. Daar waar mogelijk worden persoonsgegevens gepseudonimiseerd of geanonimiseerd.

Bescherming – Meer mensen hebben toegang tot de gegevens. Deze mensen gaan zorgvuldig om met de (persoons)gegevens die ze hierbij gebruiken. Er is sprake van passende toegangscontrole (need-to-have rechten, logging).

Gebruik en beheer – Er worden beslissingen genomen op basis van data. Belangrijke beslissingen, waarbij de datakwaliteit onkreukbaar moet zijn. Mensen moeten blind kunnen vertrouwen op de data waar ze gebruik van maken. Voor het zo maar koppelen van andere gegevensbronnen of het ‘even uitproberen’ is geen plaats meer, dat wordt vooraf zorgvuldig getoetst.

Samen de weg vinden in het labyrinth

In mijn ervaring beseffen Big Data specialisten en Privacy specialisten vaak heel goed dat ze elkaar nodig hebben. Maar toch hebben ze regelmatig moeite om elkaar uit te leggen wat de mogelijkheden en de beperkingen zijn.

Waarom kijkt de Big Data specialist met argusogen naar de Privacy specialist – Een Big Data specialist is een vrijgevochten professional. Hij of zij wordt geacht inzicht te verschaffen uit een grote berg data. Dat vraagt creativiteit, ruimte en innovativiteit. Denken buiten de gebaande paden en denken buiten de bestaande kaders. De privacy wetgeving zo’n kader. Op lastige vragen als: “maar welke data heb je nu echt nodig” of “wat wil je er mee doen” zit een Big Data specialist niet te wachten.

Waarom kijkt de Privacy specialist met argusogen naar de Big Data specialist – Voor de Privacy specialist gaat bezinnen vóór beginnen en geldt dat minimalisatie (zo min mogelijk gebruiken van persoonsgegevens) altijd de te prefereren oplossing is. De hele dag krijgt de specialist de vraag of iets wel mag of niet. En bij een verkeerd antwoord sta je in de krant, is je baas boos of krijg je een boze brief van een toezichthouder.

Waar vinden ze elkaar – Niet verbazend: in het midden! Door ruimte te geven om te exploreren en grenzen aangeven om te exploiteren. Vanuit privacy perspectief betekent dit gepaste ruimte geven aan exploratie, ook al is de grondslag (nog) niet evident. Vanuit Big Data perspectief is het jezelf tijdig opleggen van kaders, zodat verantwoord én privacy vriendelijk kan worden gemaakt van het inzicht dat is ontstaan uit de data.

Probleem opgelost? Ja en nee. Ja, want met de matrix in dit artikel kunnen beide specialisten elkaar in gesprek gaan en de weg uit het Labyrinth vinden. En dan komen ze ongetwijfeld tot goede afspraken. Nee, omdat ontwikkelingen niet stil staan en door de technologische vooruitgang ('Deep Learning' of 'Neurale netwerken') er weer nieuwe redenen zullen zijn om elkaar eens kritisch te bevragen.

Of omdat er nieuwe bronnen zijn die aanvullend data kunnen ontsluiten, of dat je gaat delen met andere (keten) partijen of … De Privacy specialist en de Big Data specialist zullen dus met enige regelmaat samen het Big Data-Privacy Labyrinth moeten doorlopen om blijvend respect te kunnen houden voor elkaars vakmanschap.

Over de auteur – Frank van Vonderen is partner bij VKA. Doordat hij bij veel opdrachtgevers komt, ziet hij veel praktische problemen en oplossingen. Hij spreekt en schrijft hier vaak over. Frank blogt veel op www.vka.nl en is onder andere auteur van 'Help! Een audit', 'Privacy by Design' en 'Hoera! Een DPIA'.

Een engelstalige vertaling van dit artikel is verschenen in:
*DATA, CYBERSECURITY & PRIV*ACY APRIL 2019 (www.dcsp.nl)