

In deze notitie zal ik, namens de WRR, ingaan op het onderwerp ‘data en analytics’ met name in relatie tot de investeringsagenda van de Belastingdienst ten behoeve van het rondetafelgesprek op dinsdag 4 juli 2017.¹ Ik zal me daarbij in hoofdzaak baseren op het WRR-rapport ‘Big Data in een vrije en veilige samenleving’² en daarmee op het onderwerp van de grootschalige data-analyse.

Data en analytics: opportunity knocks

In een digitale wereld ligt het voor de hand dat de belastingdienst zijn pijlen (mede) richt op data analyse om zijn taken beter uit te voeren. De Belastingdienst heeft zeer ruime bevoegdheden om data te verwerken en heeft veel geïnvesteerd om zijn gegevensinfrastructuur te vereenvoudigen (de data laag) en om data te analyseren (broedkamer / data en analytics). De Belastingdienst gebruikt data-analyses om fraude te kunnen bestrijden, om onopzettelijk verkeerd ingevulde aangiften te kunnen traceren, en om zogeheten *green lanes* te faciliteren voor (rechts)personen waarvan in het verleden gebleken is dat zij hun aangiften juist invulden. Door data analyse centraal te stellen in de (toekomstige) aanpak verandert de werkwijze van de Belastingdienst – dat is ook de bedoeling – maar daarmee komen er ook andere vragen in beeld als het gaat om de relatie tussen overheid en burger. Die vragen gaan over (a) wat doet de nieuwe werkwijze van de Belastingdienst met de relatie tussen de Belastingdienst en de belastingplichtige (b) wat doet de nieuwe werkwijze met het toezicht op de belastingdienst bij de uitvoering van zijn taak.

A. Data en de werkwijze van de Belastingdienst

(Big) Data analytics en de belastingdienst

(Big) Data is meer dan alleen veel data. Het cruciale punt is dat de werkwijze anders is dan bij een gewone methode voor opsporing en controle. In een ideaaltypische Big Data analyse gaat men op zoek naar correlaties in een dataset zonder voorafgaande verdenkingen. Dat kan verrassende inzichten bieden maar geeft geen direct inzicht in causaliteit. Dat betekent ook dat de logica van Big Data altijd in de richting wijst van het verzamelen van méér data omdat dit de analyse ten goede komt. Er zit bij data-gestuurd werken geen natuurlijke stop op het verzamelen en analyseren van data, eerder het tegendeel. Een volledig ongestructureerde analyse zal in de praktijk van de belastingdienst niet vaak voorkomen. Wel zal men op basis van experimenten profielen opstellen en daarmee grote datasets doorkammen op zoek naar personen/organisaties/transacties die (in bepaalde mate) aan bepaalde profielen voldoen.

Meer is beter

De Belastingdienst heeft op basis van zijn eigen taken en registraties al de grootste dataset van Nederland: alle belastingaangiften van personen en ondernemers en massale data over betalingstransacties (1 miljoen transacties per dag). Veel van de data die geanalyseerd worden zijn dus traditioneel al intern aanwezig. Maar de belastingdienst staat ook te boek als een data hongerige organisatie. In het verleden is zowel door individuele burgers als

¹ Met dank aan dr. Erik Schrijvers (WRR) voor kritisch meelezen en waardevolle suggesties.

² Wetenschappelijke Raad voor het Regeringsbeleid (2016) *Big Data in een vrije en veilige samenleving*. Amsterdam: Amsterdam University Press. <https://www.wrr.nl/publicaties/rapporten/2016/04/28/big-data-in-een-vrije-en-veilige-samenleving>

bedrijven bezwaar gemaakt tegen het grootschalig opvragen van gegevens zonder dat er van verdenking sprake is. Dit gold bijvoorbeeld voor het opvragen van alle parkeergegevens van parkeerbeheerders, alle betaalgegevens van klanten van betaaldienstverleners (bijvoorbeeld European Merchant Services) en alle kentekenregistraties van ANPR-camera's, waardoor de Belastingdienst automobilisten bijna *realtime* kan volgen. De Belastingdienst werd vanwege dit gedrag door de rechter van een *fishng expedition* beticht, maar mocht deze handelwijze na hoger beroep alsnog voortzetten. Ook de digitale aangifte geeft aanleiding tot vragen aangezien het nu mogelijk is voor de Belastingdienst om alle data die te maken hebben met het invullen van het digitale aangifteformulier (waar, welk tijdstip, verschillende versies, twijfel bij invulling, velden wel en niet ingevuld etc.) allemaal gelogd worden.³ Of ze ook gebruikt worden bij het maken van risicoprofielen is niet helemaal helder, maar het is zeker mogelijk. Wel is zeker dat dankzij de technologische vooruitgang – ook bij andere partijen – de toegang van de Belastingdienst tot meer (nieuwe) data steeds groter wordt.

De grootschalige toegang tot data over burgers kan ook leiden tot *chilling effects*. De WRR wees in zijn rapport op de beschermende werking van afstand – van andere burgers, maar zeker ook de overheid – voor de persoonlijke vrijheid van burgers. Big Data doet afbreuk aan die afstand. Een samenleving waarin burgers zich voortdurend gevolgd en bekeken voelen kan ertoe leiden dat burgers zich anders gaan gedragen. Dit noemt men wel *chilling effects*. Met name als burgers gedrag gaan vermijden dat legaal is maar eventueel als verdacht uitgelegd zou kunnen worden, begeeft de overheid zich op glad ijs.

De belastingdienst als spil in samenwerkingsverbanden

De belastingdienst is partij in een groot aantal samenwerkingsverbanden waarin data wordt uitgewisseld. Het merendeel daarvan ligt vast in convenanten tussen de deelnemende partijen. Deze samenwerkingsverbanden zijn wat het kabinet betreft gewoon uitwerkingen van de wettelijke basis van de uitvoeringspraktijk. Volgens het kabinet bestaat er 'daarom geen aanleiding om de convenanten aan de kamer toe doen toekomen'⁴. Burgers en Tweede Kamerleden met een zeer onderzoekende instelling kunnen de convenanten desgewenst op een website van de overheid vinden en het wordt bij de Autoriteit Persoonsgegevens aangemeld maar die houdt geen actief toezicht. In een Big Data wereld is het de vraag of dit in termen van toezicht wel voldoende is.

Data en doelen: 'tax gaps' en 'data gaps'

In zijn brief over de investeringsagenda van 20 mei 2015⁵ stelt de staatssecretaris dat het doel van de Belastingdienst vanaf nu is om de 'tax gap' zo klein mogelijk te maken: er moet met de beschikbare capaciteit zoveel mogelijk belasting geïnd worden (p. 5 van 10). Eén van de voornaamste methoden om dat te bereiken is het gebruik van data analytics. Data analyse wijst de weg en bepaalt wat gecontroleerd wordt. De combinatie van data analyse (middel) en het dichten van de tax gap (doel) wijst in de richting van het laagst hangende fruit. Dat betekent ook dat alleen die vormen van fraude of vergissingen in beeld komen die door middel van data analyse ontdekt kunnen worden – of zelfs makkelijk ontdekt kunnen worden. Waar geen data over is, wordt op deze manier ook niet bestreden. De staatssecretaris stelt een 'nieuwe wereld in het vooruitzicht met meer opbrengsten door een dalende tax gap' maar geeft ook in een voetnoot aan dat de 'zwarte economie – het andere deel van de tax gap – niet met een informatie gestuurde aanpak aangepakt kan worden omdat deze niet in de gegevens van de belastingdienst voorkomt'. Met andere woorden: waar data ontbreekt, kan niet informatie-gestuurd gewerkt worden. Maar ook: als het doel onverkort het binnenhalen van het grootste

³ <https://accountantweek.nl/artikel/bond-voor-belastingbetalers-waarschuwt-voor-privacyrisico-online-belastingaangifte->

⁴ Peter Olsthoorn (2017) *Big Data voor fraudebestrijding*. WRR Working Paper nr. 21 :p. 77. <https://www.wrr.nl/publicaties/working-papers/2016/04/28/big-data-voor-fraudebestrijding>

⁵ <https://www.rijksoverheid.nl/documenten/brieven/2015/05/20/brief-over-de-uitwerking-brede-agenda-belastingdienst>

belastingopbrengst is met data analyse als middel heeft dat een selectieve werking waardoor sommige overtredingen wel in beeld en andere niet in beeld van de Belastingdienst komen – los van de inhoudelijke beoordeling van de overtreding.

Monitoring van opbrengsten en van de modellen zelf

In november 2016 publiceerde de Algemene Rekenkamer een rapport over het handhavingsbeleid van de Belastingdienst.⁶ Naast complimenten voor de aanpak, waren er ook vragen en verzoeken. Een belangrijke aanbeveling is om de keuzes in het handhavingsbeleid verder en met betere informatie te onderbouwen. Dan gaan het vooral om inzicht in de kosten en baten van handhavingsinstrumenten. Met andere woorden: het is nu moeilijk te zeggen hoeveel opbrengsten de investeringen opleveren. Ook op het punt van de modellen die worden gebruikt had de Rekenkamer twee belangrijke aanbevelingen. (1) Gerichtte controle op de aangiften van risicogeveallen is effectiever en efficiënter dan controle zonder risicoselectie. Toch blijven aselechte steekproeven nodig omdat deze methode onmisbaar is om te identificeren welke risicofactoren bestaan. (2) Voor het toezicht heeft de Belastingdienst ook informatie uit externe bronnen nodig, zogeheten contra-informatie. Het ontbreekt nu aan voldoende overzicht van deze bronnen van deze informatie. Beide aanbevelingen zijn erop gericht om de modellen die worden gebruikt te testen. Deze informatie is nodig om er zeker van te zijn dat de modellen daadwerkelijk doen wat ze moeten doen.

Risico's in grijstinten

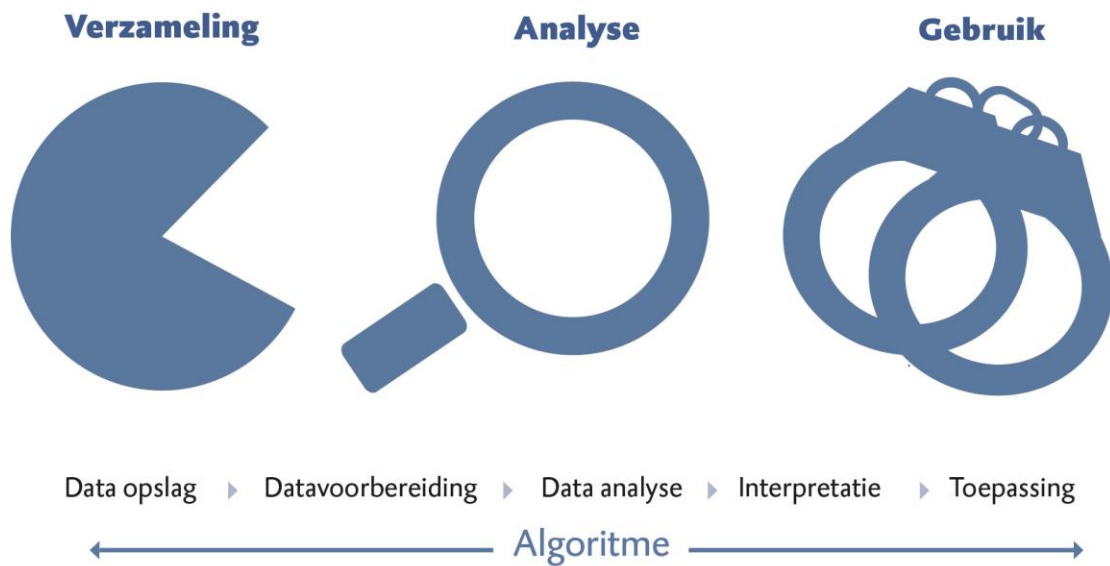
De patroonherkenning van Big Data brengt op basis van statistische waarschijnlijkheden veel meer belastingplichtigen onder de aandacht dan een steekproefsgewijs onderzoek of een onderzoek dat vanuit een concrete verdenking vertrekt. Enerzijds is dat winst: de onderzoeken worden beter toegesneden op risico's en de opbrengt van de onderzoeken gaat omhoog. Anderzijds heeft het risico's. Data-analyse, risicobepaling, patroonherkenning en uiteindelijk profilering van personen en bedrijven kunnen leiden tot omkering van bewijslast: personen en bedrijven zijn eerder en langer verdacht op grond van data en moeten het tegendeel - het profiel is niet van toepassing op mij - bewijzen.⁷

B. Versterking van het toezicht op de fase van analyse.

Het huidige juridische regime voor privacy en data bescherming legt de nadruk op het reguleren van het *verzamelen* van data. Belangrijke juridische ankerpunten als doelbinding en data minimalisatie zijn van toepassing op de verzameling van data. In het tijdperk van grootschalige data-analyse regeert de logica van het gebruik van veel data (N=all) en het secundair gebruik van data. Dat levert spanningen op met het huidige juridisch kader. De WRR pleit ervoor dat die bestaande wetgeving wordt aangevuld met de regulering van en het toezicht op de fases van de *analyse* en het *gebruik* van Big Data.

⁶ Algemene Rekenkamer (2016) Handhavingsbeleid Belastingdienst.
http://www.rekenkamer.nl/Publicaties/Onderzoeksrapporten/Introducties/2016/11/Handhavingsbeleid_Belastingdienst

⁷ Aldus Gerrit-Jan Zwenne (hoogleraar Recht in de Informatiemaatschappij), geciteerd in Olsthoorn 2017: 75



Zorgplicht

Gezien het belang van de analysefase vindt de WRR juist dat deze fase veel sterker onder toezicht zou moeten komen te staan. Dit zou moeten worden vastgelegd in een wettelijke zorgplicht waarin het borgen van de kwaliteit en de toegankelijkheid van de analyse voor het toezicht centraal staan. Centrale elementen daarvan zijn:

- Zorgen dat gegevens up to date zijn en dat gecorrigeerd wordt voor de bias in datasets. Deze plicht strekt zich ook uit tot gegevens die van derden worden verkregen.
- De algoritmes en methoden die bij data-analyses worden gebruikt moeten deugdelijk zijn, aan de wetenschappelijke criteria voor goed (statistisch) onderzoek voldoen en voor toezicht toegankelijk zijn.
- Onderzoeksresultaten, profielen en foutmarges moeten op hun merites worden beoordeeld: de Belastingdienst moet duidelijk kunnen maken hoe zij tot bepaalde uitkomsten komen.

De zorgplicht is de basis voor het versterken van het toezicht op de fase van de analyse. Dat vergt niet alleen inspanningen aan de kant van de Belastingdienst maar ook aan de kant van de toezichthouder zelf. De grootschaligheid van de data analyse, het groeiende aantal samenwerkingsverbanden waarin data wordt uitgewisseld en het feit dat de belastingdienst vol inzet op data gestuurde processen (en daarmee binnen de overheid voorop loopt) rechtvaardigt een grotere rol voor de Autoriteit Persoonsgegevens. Die moet daartoe wel in staat worden gesteld. Een belangrijke vereiste is dat de Autoriteit Persoonsgegevens zijn technische en statistische capaciteit en expertise kan versterken.

Transparantere werkwijze

Aangezien met een data gestuurde werkwijze veel meer mensen in de analyses van de Belastingdienst terecht komen, vereist dit meer transparantie. De afweging tussen noodzakelijke geheimhouding (voorkomen van ‘gaming the system’) en transparantie valt in Nederland te vaak – en met weinig onderbouwing – in het voordeel van de geheimhouding uit. Op sommige punten kan het eenvoudig beter, zoals meer openheid over data uitwisseling en samenwerkingsverbanden, op andere punten kan transparantie ook getraptd vorm worden gegeven. Door de toezichthouder meer inzicht te geven in de analysefase vergroot dat het inzicht in de werkwijze en in de risico's. De toezichthouder kan vervolgens aan uw kamer rapporteren. Omdat de verantwoording getraptd verloopt, is er ruimte is voor een grotere mate

van transparantie en detail in de rapportages aan de toezichthouder. In het openbare rapport van de toezichthouder aan de Kamer wordt 'het geheim van de smid' (de specifieke methoden) niet opgenomen. Maar het is wel op basis van kennis en inzicht daarvan opgesteld.

(Semi-) Automatische besluitvorming

In data gedreven methoden worden profielen steeds meer leidend in het maken van keuzes en beslissingen. De WRR beveelt daarom aan om het verbod op geautomatiseerde besluitvorming strikter te handhaven en alert te zijn op semi-automatische besluitvorming, waarbij formeel een mens het besluit neemt, maar deze de facto niet afwijkt van het digitale advies. Dit raakt ook aan het eerder gemaakte punt dat data-analyses en profielen niet mogen leiden tot een feitelijke omkering van de bewijslast. Dat speelt niet zozeer in het strafrecht – waar strikte regels voor bewijsvoering gelden – maar wel in verschillende vormen van surveillance, handhaving en fraudebestrijding. Het gevaar is dat bij geschillen niet van de overheid wordt verwacht dat deze aantoont dat een individu aan een profiel voldoet, maar dat een persoon moet bewijzen dat hij ten onrechte met een profiel wordt geassocieerd. Gezien het feit dat bij Big Data-processen in het veiligheidsdomein het machtsevenwicht verder verschuift ten 'gunste' van de overheid moet de burger meer grip krijgen op de beslissingen die hem aangaan.

Prof. Dr. Dennis Broeders
Wetenschappelijke Raad voor het Regeringsbeleid

Den Haag, 30 juni 2017