WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID

Big Data in een vrije en veilige samenleving

Amsterdam University Press

Big Data in een vrije en veilige samenleving

De Wetenschappelijke Raad voor het Regeringsbeleid werd in voorlopige vorm ingesteld in 1972. Bij wet van 30 juni 1976 (Stb. 413) is de positie van de raad definitief geregeld. De huidige zittingsperiode loopt tot 31 december 2017.
Ingevolge de wet heeft de raad tot taak ten behoeve van het regeringsbeleid wetenschappelijke informatie te verschaffen over ontwikkelingen die op langere termijn de samenleving kunnen beïnvloeden. De raad wordt geacht daarbij tijdig te wijzen op tegenstrijdigheden en te verwachten knelpunten en zich te richten op het formuleren van probleemstellingen ten aanzien van de grote beleidsvraagstukken, alsmede op het aangeven van beleidsalternatieven.
Volgens de wet stelt de WRR zijn eigen werkprogramma vast, na overleg met de minister-president die hiertoe de Raad van Ministers hoort.
De samenstelling van de raad is:
prof. dr. A.W.A. Boot
prof. dr. mr. M.A.P. Bovens
prof. dr. G.B.M. Engbersen
prof. mr. dr. E.M.H. Hirsch Ballin
prof. dr. J.A. Knottnerus (voorzitter)
prof. dr. M. de Visser
prof. dr. C.G. de Vries (adviserend raadslid)
prof. dr. ir. M.P.C. Weijnen
Secretaris: dr. F.W.A. Brom

Wetenschappelijke Raad voor het Regeringsbeleid
Buitenhof 34
Postbus 20004
2500 EA Den Haag
Telefoon 070-356 46 00
E-mail info@wrr.nl
Website www.wrr.nl

WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID

Big Data in een vrije en veilige samenleving

Amsterdam University Press, Amsterdam 2016

Vormgeving binnenwerk: Textcetera, Den Haag
Omslagafbeelding: cimon communicatie, Den Haag

ISBN	9789462983571
e-ISBN	9789048533923 (pdf)
e-ISBN	9789048534029 (ePub)
NUR	740

WRR

WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID

Aan de Minister-President
Voorzitter van de Ministerraad
De heer drs. M. Rutte
Postbus 20001
2500 EA Den Haag

ons kenmerk
2016010/AK/am

doorkiesnummer
070 3564691

onderwerp
WRR-rapportnr. 95
Big Data in een vrije en veilige samenleving

email
knottnerus@wrr.nl

datum
14 april 2016

Het doet mij genoegen u hierbij het rapport *Big Data in een vrije en veilige samenleving* aan te bieden. Met dit rapport geeft de raad antwoord op de adviesaanvraag van het kabinet over het thema 'Big Data, veiligheid en privacy'.

De raad constateert dat Big Data een belangrijke bijdrage kan leveren aan het bevorderen van de veiligheid. In dit rapport analyseert de WRR hoe de Nederlandse overheid Big Data kan gebruiken en richt zich daarbij specifiek op (Big) Data-analyse door politie en justitie, de inlichtingen- en veiligheidsdiensten en verschillende organisaties en samenwerkingsverbanden op het gebied van de fraudebestrijding. De geautomatiseerde analyse van grote, gecombineerde gegevensbestanden levert onder meer grote tijdwinst op en kan in nauwkeurigere uitkomsten resulteren, die bruikbaar zijn voor gerichte inspecties, de reconstructie van aanslagen en het in kaart brengen van criminele netwerken om opsporing van daders te vergemakkelijken. De overheid staat bovendien nog maar aan het begin van wat mogelijk is met Big Data-analyse. De grootschalige verzameling, het hergebruik van data en geautomatiseerde gegevensanalyses met profielen – de kern van Big Data-toepassingen - brengen echter ook risico's met zich mee, die te maken hebben met privacy, discriminatie en chilling effecten op de vrije meningsuiting.

De WRR meent dat de kansen die Big Data biedt voor opsporing en surveillance gepaard moeten gaan met sterkere waarborgen voor de vrijheidsrechten van burgers. Het reguleren van het *verzamelen* van data – het zwaartepunt in de huidige juridische kaders – moet daartoe aangevuld worden met de regulering van en het toezicht op de fases van de *analyse* en het *gebruik* van Big Data. De WRR formuleert hiertoe een samenhangend regulerend kader, en bepleit een versteviging van het onafhankelijke toezicht en meer mogelijkheden voor burgers en de organisaties waarin zij zich verenigen om besluitvorming op basis van Big Data-processen te bevragen. Dit regulerend kader is nodig voor het vertrouwen dat burgers moeten kunnen hebben dat de overheid niet sluipenderwijs in hun persoonlijke vrijheid penetreert. Ook houdt het de overheid scherp om legitieme Big Data-toepassingen in het veiligheidsdomein te blijven verkennen en implementeren.

Ingevolge de Instellingswet ziet de raad graag de bevindingen van de ministerraad tegemoet.

De voorzitter,

Prof. dr. J.A. Knottnerus

De secretaris,

Dr. F.W.A. Brom

Buitenhof 34, Postbus 20004, 2500 EA Den Haag
telefoon (070) 356 46 00, website: www.wrr.nl, e-mail: info@wrr.nl

INHOUDSOPGAVE

SAMENVATTING

De WRR analyseert in dit rapport hoe de Nederlandse overheid Big Data op een verantwoorde wijze kan gebruiken. Het rapport richt zich specifiek op (Big) Data-analyses door politie en justitie, de inlichtingen- en veiligheidsdiensten en verschillende organisaties en samenwerkingsverbanden op het gebied van fraudebestrijding. Big Data biedt zeker kansen voor opsporing en surveillance, maar vraagt tevens om sterkere waarborgen voor de vrijheidsrechten van burgers. Het zwaartepunt in de huidige juridische regelgeving ligt op de regulering van het *verzamelen* van data. 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.

BIG DATA EN VEILIGHEID

De hoeveelheid beschikbare data over personen en processen is de laatste jaren exponentieel toegenomen. Dat komt vooral omdat veel data tegenwoordig automatisch worden geproduceerd en het bijproduct zijn van dagelijkse handelingen zoals het gebruik van internet, sociale media, mobiele telefoons en verschillende applicaties. Hierdoor worden steeds meer handelingen van individuen digitaal geregistreerd. Bovendien verdubbelt de opslagcapaciteit ongeveer iedere drie jaar en nemen de kosten van dataopslag sterk af. De combinatie van steeds krachtigere computers, betere software, zelflerende algoritmen en *machine learning* biedt kansen voor Big Data-toepassingen. Door het koppelen van databases wordt het mogelijk om nieuwe, praktisch bruikbare kennis te construeren. Behalve commerciële partijen maken ook overheidsorganisaties steeds vaker gebruik van die nieuwe kennis- en datavergaring.

Het is niet gemakkelijk in kaart te brengen hoe en in welke mate Big Data zich in het (Nederlandse) veiligheidsdomein manifesteert. Dit heeft te maken met geheimhouding en het experimentele karakter van sommige toepassingen. Naast het domein van de inlichtingen- en veiligheidsdiensten, de Belastingdienst en fraudebestrijding zijn er vooralsnog weinig echte Big Data-toepassingen in gebruik. Dat komt mede doordat de omschakeling op nieuwe werkwijzen tijd en geld kost en verreikende aanpassingen in de werkorganisatie impliceert. Veel organisaties hebben nog niet de keuze gemaakt om grootschalige, data-gedreven analyses uit te voeren. Door de snelle technologische ontwikkelingen zal deze keuze voor een groeiend aantal organisaties desondanks in de nabije toekomst een reële optie zijn.

KANSEN EN RISICO'S

Big Data heeft voor specifieke vormen van criminaliteit (zoals fraude) al positieve resultaten laten zien, al ontbreekt in veel gevallen een betrouwbare onderbouwing en evaluatie van de effectiviteit van de gebruikte analysemethoden. De geautomatiseerde analyse van grote, gecombineerde gegevensbestanden levert grote tijdwinst op en kan – mits zorgvuldig uitgevoerd – ook in nauwkeurigere uitkomsten resulteren. Organisaties kunnen deze uitkomsten gebruiken om gerichte inspecties uit te voeren. Big Data is tevens nuttig bij de reconstructie van aanslagen en het in kaart brengen van criminele netwerken, met als doel de opsporing van daders te vergemakkelijken. Ook kan Big Data behulpzaam zijn bij het *real time* volgen van ontwikkelingen in crisissituaties of bij *crowd control* rond evenementen. Politiemensen en veiligheidsfunctionarissen kunnen dan snel een beeld krijgen van de situatie ter plaatse. Dit soort toepassingen zal in de nabije toekomst steeds belangrijker worden. We gaan toe naar steeds verdergaande koppelingen van databronnen. Ook zal het steeds aantrekkelijker worden om (ten minste een deel van) het analyseproces te automatiseren. De grens tussen data uit publieke en private bronnen zal vervagen.

Aan Big Data-toepassingen kleven echter ook risico's. Een van de grootste zorgen is de grootschalige inmenging in de persoonlijke levenssfeer. De grootschalige verzameling, opslag en analyse van data door overheden, waaronder inlichtingen- en veiligheidsdiensten, kunnen ertoe leiden dat mensen het gevoel krijgen dat hun privacy en vrije meningsuiting in gevaar zijn, waardoor zij hun gedrag daarop aanpassen. Bovendien worden burgers steeds transparanter voor de overheid, terwijl de profielen, algoritmen en methoden die overheidsorganisaties gebruiken nauwelijks transparant of navolgbaar voor die burgers zijn. Nu met Big Data-toepassingen steeds grotere groepen burgers in beeld komen – naast verdachte ook niet-verdachte burgers – gaat dat gebrek aan transparantie steeds meer wringen. Daarnaast kunnen Big Data-toepassingen leiden tot een toename van sociale stratificatie, met een ongelijke verhouding tussen maatschappelijke groepen als gevolg. Dit gebeurt doordat Big Data onregelmatigheden en afwijkingen in datasets kunnen reproduceren, resulterend in uitkomsten die een onevenredige sociale impact hebben. Zonder correctie vertaalt zich dit op termijn in een cumulatief nadeel (discriminatie en oneerlijke behandeling) voor bepaalde groepen in de maatschappij. Ook zijn Big Data-toepassingen zeer gevoelig voor 'function creep', oftewel gebruik van gegevens anders dan voor het doel waarvoor de data zijn verzameld. De reden hiervan is dat het secundair gebruik van gegevens bij Big Data-toepassingen een grote meerwaarde oplevert.

MISMATCH TUSSEN BIG DATA EN DE HUIDIGE WET- EN REGELGEVING

De juridische kaders die van toepassing zijn op de gegevensverwerking binnen het veiligheidsdomein zijn vooral gericht op het verzamelen en delen van gegevens. Door het gebruik van Big Data ontstaat echter druk op belangrijke uitgangspunten

van deze kaders, zoals doelbinding en noodzakelijkheid. Want in zijn ideaalvorm is Big Data gebaseerd op het principe van ongerichte gegevensverzameling en secundair gebruik van reeds verzamelde gegevens voor andere doeleinden, hetgeen botst met de regelgevende kaders. De wettelijke normen voor het verzamelen van gegevens vereisen daarom aanvulling. Het verzwaren van het huidige kader – met de nadruk op het reguleren van verzamelen en de handhaving van doelbinding en noodzakelijkheid – zou echter een groot deel van de belofte van Big Data in de kiem smoren.

De WRR zet daarom in op een versterking van de regulering van de fases van de *analyse* en het *gebruik* van Big Data-processen. De bestaande en in ontwikkeling zijnde regulering voor het verzamelen van gegevens hebben desondanks ook in het Big Data-tijdperk onverminderd een belangrijke functie. De raad is echter van mening dat er meer winst te behalen valt in de latere fasen van Big Data-processen dan in een intensivering van de regulering van het verzamelen van data. Alleen door extra eisen te stellen aan het toepassen van Big Data-analyses kunnen burgers erop vertrouwen dat de overheid niet sluipenderwijs in hun persoonlijke vrijheid penetreert. Regulering van analyse en gebruik is volgens de raad een *conditio sine qua non* voor het niet zwaarder reguleren van het verzamelen van gegevens.

REGULERING VAN DATA-ANALYSE EN -GEBRUIK

Bij de regulering van de fase van de analyse van gegevens is sprake van een hiaat in de regelgeving. In Big Data-processen zijn de keuzes die in de analysefase worden gemaakt (algoritmen, categorisering, wegingsfactoren enz.) van eminent belang. Juist op dit vlak kunnen zich risico's voordoen zoals discriminatie en te veel pretenderende gegevensanalyses. Op de achtergrond hiervan dreigen negatieve sociale effecten op persoonlijke vrijheden zoals de vrije meningsuiting, die een vitale rol spelen in de democratische rechtsstaat.

Er is daarom aanvullende normering nodig in de vorm van een *wettelijk omschreven zorgplicht*, met algemene vereisten voor de kwaliteit van de data en van de deugdelijkheid van de gehanteerde analysemethoden.

De gegevensverwerkende partijen moeten desgevraagd altijd duidelijk kunnen maken hoe zij tot bepaalde uitkomsten komen. Dat vereist al tijdens de analysefase externe aandacht. Big Data-projecten en -toepassingen in het veiligheidsdomein moeten onderwerp zijn van een *externe review door de toezichthouder*. Bij gebleken gebreken in de rapportage kan de toezichthouder de audits aanscherpen, opvoeren en in uiterste gevallen overlaten aan een onafhankelijke derde. Een belangrijke vereiste voor deze verscherpte vorm van toezicht is dat toezichthouders zoals de Autoriteit Persoonsgegevens en de Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten (CTIVD) hun *technische en statistische capaciteit en expertise versterken*. Ook is het vanwege de potentiële impact van

data-gedreven toepassingen in het veiligheidsdomein belangrijk om bij Big Data-projecten vooraf een evaluatiemoment in te plannen. Grote dataverwerkingsprojecten binnen de overheid, vooral door de politie, inlichtingen- en veiligheidsdiensten, inspecties, de Belastingdienst en samenwerkingsorganen op het terrein van misdaad- en fraudebestrijding, moeten een *horizon van 3 tot 5 jaar* krijgen.

Bij het gebruik van de gegevensanalyses in Big Data-processen is de problematiek rond het samenstellen van profielen van belang. De kracht van Big Data-analyses ligt voornamelijk in algemene conclusies en structurele patronen. Bij de toepassing daarvan op concrete situaties en specifieke individuen bestaat altijd een mismatch, omdat een profiel zowel over- als onder-inclusief is. De WRR beveelt aan om het *profileren strakker te reguleren* door nadere regels over toelaatbare foutmarges te stellen, het verbod op geautomatiseerde besluitvorming door computers strikter te handhaven en alert te zijn op semi-automatische besluitvorming. De Nederlandse overheid zou in Europa een voortrekkersrol moeten nemen en ervoor moeten zorgdragen dat geautomatiseerde besluitvorming achterwege blijft. Hiertoe behoren in feite ook de situaties waarin formeel een mens het besluit neemt, maar deze de facto niet afwijkt van het digitale advies.

In het verlengde hiervan beveelt de WRR aan om juridisch te verankeren dat dataanalyses en profielen niet kunnen leiden tot een feitelijke verlegging 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.

TOEZICHT, TRANSPARANTIE EN RECHTERLIJKE TOETSING

De toegenomen mogelijkheden om data te verzamelen en te analyseren, vragen om een *versteviging van het onafhankelijke toezicht*. Het toezicht op gegevensverwerking laat tot nu toe veel te wensen over, zeker in het licht van de huidige snelle ontwikkelingen op het gebied van Big Data. Zowel de Autoriteit Persoonsgegevens als de CTIVD is onvoldoende toegerust voor de uitdagingen van het Big Data-tijdperk in termen van bevoegdheden, expertise en financiële middelen. Vele partijen, waaronder de CTIVD zelf, zijn van mening dat de voorgenomen uitbreiding van bevoegdheden van de MIVD en AIVD vraagt om een significante uitbreiding van de capaciteit en expertise van de toezichthouder op alle niveaus. Voor de parlementaire Commissie voor de Inlichtingen- en Veiligheidsdiensten, die nauwelijks eigen ondersteuning heeft, geldt dat wellicht nog sterker. Hoewel de bevoegdheden en middelen van de Autoriteit Persoonsgegevens als gevolg van de nieuwe Europese verordening gegevensbescherming zullen worden verstevigd, is het voornamelijk aan de nationale wetgever om de bijbehorende financiële middelen, bevoegdheden en capaciteiten toe te kennen. Hier is dus actie van de Nederlandse regering nodig.

Ook op het punt van de transparantie van de dataverwerkingsprocessen van de overheid is nog een wereld te winnen. De gegevensverwerking is in veel gevallen een black box. Individuen kunnen vaak niet weten dat over hen gegevens zijn verzameld en zullen dus niet zo snel hun informatierecht inroepen. Hoewel binnen het veiligheidsdomein geen volledige transparantie kan bestaan vanwege geheimhouding, is desalniettemin op verschillende niveaus een *grotere mate van transparantie* mogelijk. Veel relevante informatie over gegevensverwerking binnen samenwerkingsverbanden op het terrein van fraudebestrijding staat bijvoorbeeld in convenanten en besluiten vermeld, die weliswaar openbaar maar niet erg toegankelijk zijn. Ook is het wenselijk dat organisaties die met Big Data-toepassingen aan de slag gaan, een beleidsplan opstellen waarin zij vermelden welke methoden zij gebruiken, en wat de kosten en de beoogde resultaten zijn. De inlichtingen- en veiligheidsdiensten zouden meer werk kunnen maken van het achteraf inzichtelijk maken van de frequentie waarmee zij bepaalde toepassingen hebben gebruikt en voor welke doeleinden. De beschikbaarheid van dergelijke informatie kan een bijdrage leveren aan een grotere maatschappelijke aanvaardbaarheid van de inzet van Big Data-toepassingen in het veiligheidsdomein.

Veel grote dataverwerkingsprojecten overstijgen het individu in aard en omvang. Daarom is het belangrijk om, naast een inzet op toezicht en transparantie, ook de positie van NGO's en burgerrechtenorganisaties in juridische procedures te versterken. Het is weliswaar primair de verantwoordelijkheid van de wetgevende macht en van het parlement in zijn controlerende functie, om wetgeving en beleid omtrent Big Data-toepassingen te toetsen. Maar burgers kunnen ook direct of via belangenorganisaties stem geven aan hun belang bij vrijheid en veiligheid. In de huidige situatie is het klachtrecht sterk verbonden aan individuele schade en zijn er zeer beperkte mogelijkheden voor collectieve procedures bij de rechter. Dit geeft de burger – en organisaties waarin burgers zich verenigen – te weinig mogelijkheden om besluitvorming op basis van Big Data-processen te bevragen zolang zij geen gezamenlijke persoonlijke benadeling kunnen aanvoeren. Het is dus belangrijk dat de rechter selectief zaken toelaat die recht doen aan collectieve zorgen en bijdragen aan de *opbouw van jurisprudentie* op dit belangrijke en relatief onontgonnen terrein.

TEN GELEIDE

In dit rapport schetst de WRR zijn visie op het gebruik van Big Data door de overheid, in het bijzonder in het domein van het veiligheidsbeleid. Het rapport is opgesteld door een projectgroep onder leiding van prof. mr. dr. Ernst Hirsch Ballin, lid van de raad. De projectgroep bestond verder uit de stafleden prof. dr. Dennis Broeders (projectcoördinator), dr. Erik Schrijvers, mr. drs. Bart van der Sloot, Rosamunde van Brakel (MA) en mr. drs. Josta de Hoog. In een eerdere fase van het project heeft ook Sacha van Schendel (LLB) een bijdrage aan de werkzaamheden geleverd.

Dit rapport is tot stand gekomen op basis van een uitvoerige analyse van de rijke (internationale) wetenschappelijke literatuur, onderzoek dat in opdracht van de WRR is verricht, en bijeenkomsten en gesprekken met externe deskundigen uit diverse lagen van het bestuur, de politiek en de wetenschap. Deze deskundigen waren onder meer verbonden aan ministeries, toezichthouders (Autoriteit Persoonsgegevens, Commissie van Toezicht op de Inlichtingen en Veiligheidsdiensten), veiligheidsorganisaties (politie, inlichtingen- en veiligheidsdiensten, Openbaar Ministerie, inspecties en samenwerkingsverbanden op het terrein van fraude- en criminaliteitsbestrijding), uitvoeringsinstanties, kennisinstituten (Rathenau Instituut), Nederlandse en buitenlandse universiteiten, bedrijven en andere relevante instellingen. De raad is al deze personen zeer erkentelijk voor hun tijd, kennis en suggesties, die van onschatbare waarde zijn geweest.

In opdracht van de WRR is tevens ondersteunend onderzoek verricht, dat in twee vormen is gepubliceerd. Een deel is als hoofdstuk opgenomen in de achtergrondstudie *Exploring the Boundaries of Big Data* (Van der Sloot et al. 2016). Een ander deel is verschenen als Working Papers via de website van de WRR. Al deze publicaties kunnen via www.wrr.nl worden gedownload. De raad is alle auteurs en deskundigen erkentelijk voor hun bijdrage.

Tot slot geldt een bijzonder woord van dank aan enkele deskundigen die in de laatste fase actief commentaar leverden op de conceptrapportage, dan wel bij aanvang meedachten over de opzet van het project: prof. mr. dr. Mireille Hildebrandt, prof. dr. Nico van Eijk, prof. dr. Ronald Leenes, prof. dr. Bart Jacobs, prof. dr. Jeroen van den Hoven, prof. dr. ir. Arnold Smeulders en prof. mr. dr. Erwin Muller. De raad is mr. dr. Bart Schermer bijzondere dank verschuldigd voor zijn inzichten en actieve bijdrage in de slotfase van het project.

1 BIG DATA IN EEN VRIJE EN VEILIGE SAMENLEVING

1.1 INLEIDING: EEN AMBIGU FENOMEEN

Big Data[1] is een ambigu fenomeen. De technische mogelijkheden om zeer grote, deels ongestructureerde bestanden te analyseren roepen zowel groot enthousiasme als angstige gevoelens op. Deze gevoelens uiten zich in utopische en dystopische toekomstbeelden in de politiek, media en wetenschap. De beoogde voordelen van het gebruik van Big Data zijn onder meer een betere dienstverlening aan consumenten en burgers, en een grotere efficiëntie bij bedrijven en overheden. De oorzaak voor veel onrust is het feit dat Big Data het gebruik van enorme hoeveelheden data veronderstelt, waaronder geaggregeerde informatie over wie wij zijn en wat we doen. Hieraan verbonden zijn zorgen over privacy en vrijheid, zeker als het gaat om het gebruik van Big Data in het veiligheidsdomein.

Deze zorgen worden versterkt door onthullingen over massasurveillance door de Amerikaanse National Security Agency (NSA) en de grote hoeveelheden persoonlijke data die voor commerciële of veiligheidsdoeleinden ongemerkt de grens passeren. Ook bestaat er bij burgers onzekerheid over de betrouwbaarheid van al die data, waarvan de opslag bovendien regelmatig verre van veilig blijkt te zijn (Goodman 2015). Hiertegenover staat dat Big Data de maatschappelijke veiligheid lijkt te kunnen bevorderen. Zo kan Big Data leiden tot een meer doeltreffende inzet van de brandweer, een betere organisatie van humanitaire hulp in de nasleep van een grote ramp, en een effectievere bestrijding van misdaad. David Lyon (2015: 11) vat deze dubbelzinnigheid mooi samen als hij stelt: "Surveillance and big data are not inherently 'good' or 'bad' – but they are never 'neutral' either. They have to be probed and assessed further."

1.2 ADVIESAANVRAAG VAN DE REGERING

Dit rapport verkent de inzet van Big Data door de overheid op het terrein van het veiligheidsbeleid. De nadruk ligt daarbij op het gebruik van (Big) Data-analyse door politie en justitie, de inlichtingen- en veiligheidsdiensten en verschillende organisaties en samenwerkingsverbanden op het gebied van de fraudebestrijding. De WRR doet dat op verzoek van de Nederlandse regering, die advies heeft gevraagd over het brede thema 'Big Data, privacy en veiligheid'. Het kabinet zoekt naar mogelijkheden om Big Data – liefst op een transparante wijze – te gebruiken om de effectiviteit van het veiligheidsbeleid te vergroten én de bescherming van

privacy en persoonsgegevens beter te waarborgen. De adviesaanvraag van de regering – die als bijlage in dit rapport is bijgevoegd – bevat vier hoofdvragen, die als volgt zijn samen te vatten:

1. Kan en, zo ja, moet er een scherper onderscheid worden gemaakt tussen *toegang* tot, de *verzameling* van en het *gebruik* van gegevens in privacyrecht en databescherming?
2. Hoe kunnen processen als *profiling*, *datamining* en andere analysetechnieken voldoende transparant zijn zonder dat zij de effectiviteit van het veiligheidsbeleid doorkruisen?
3. Wat betekent de komst van kwantumcomputers voor het proces van gegevensverwerking en bescherming (encryptie)?
4. Welke gevolgen heeft Big Data voor de gegevenshuishouding van de overheid en hoe kan de burger daarop invloed hebben?

Deze vier hoofdvragen zijn in de adviesaanvraag uitgewerkt in een groot aantal subvragen. Om de regering en de lezer een samenhangend en leesbaar rapport te kunnen presenteren, heeft de WRR de vrijheid genomen deze vragen en de onderliggende subvragen niet letterlijk als uitgangspunt voor de studie te nemen. Er is een aantal voorvragen toegevoegd – zoals de vraag wat we onder Big Data kunnen verstaan – en omwille van de helderheid is er afgebakend naar het veiligheidsbeleid, de rol die Big Data daarin kan spelen en welke kansen en risico's daaraan verbonden zijn. Dit rapport beantwoordt de volgende vragen:

1. Wat is Big Data? (hoofdstuk 2)
2. Hoe en in welke mate manifesteert Big Data zich in het (Nederlandse) veiligheidsdomein? (hoofdstuk 3)
3. Wat levert een sterkte-zwakteanalyse van Big Data en veiligheid op? Wat zijn de belangrijkste beloftes, beperkingen, risico's en randvoorwaarden? (hoofdstuk 4)
4. Waar liggen de spanningen tussen de huidige (juridische) kaders voor gegevensverwerking in het veiligheidsdomein en de ontwikkeling van Big Data? (hoofdstuk 5)
5. Wat is nodig om het beleid en het (juridische) kader zo te organiseren dat een verantwoorde inzet van Big Data mogelijk wordt, fundamentele rechten van burgers beschermd zijn en een goede verhouding ontstaat tussen transparantie en effectiviteit van het veiligheidsbeleid? (hoofdstuk 6)

Als overkoepelend antwoord op de adviesaanvraag presenteert de WRR in hoofdstuk 6 dus een kader voor het gebruik van Big Data in het veiligheidsbeleid, waarbij de bescherming van ieders vrijheid en veiligheid, ongeacht ras, geloof, gender of

herkomst het uitgangspunt is. De vier hoofdvragen van de adviesaanvraag van de regering komen op verschillende punten en in iets aangepaste vorm in de tekst terug en zullen in de verschillende hoofdstukken kort aangestipt worden.

1.2.1 KERNBEGRIPPEN

Voor een goed begrip van de thematiek van dit rapport bespreken we hierna de belangrijkste begrippen.

Big Data

Dit rapport bevat geen scherp omlijnde definitie van Big Data. In plaats daarvan richten we ons op een drietal hoofdkenmerken van Big Data, die als leidraad voor de analyse dienen:

1. *Data*: het gaat om grote hoeveelheden gestructureerde en ongestructureerde data uit verschillende bronnen.
2. *Analyse*: de analyse is 'data-driven' en zoekt geautomatiseerd naar correlaties. De grootste potentie wordt verwacht van realtime en voorspellende analyses.
3. *Gebruik*: de analyses moeten leiden tot 'actionable knowledge' (ingrepen in de realiteit op basis van bestandsanalyses).

Big Data is bovendien omgeven met allerlei technische en statistische begrippen. De meeste daarvan zijn niet nieuw noch specifiek voorbehouden aan Big Data. Enkele veelgebruikte technische begrippen zijn de volgende:

– *Profiling*: het gebruik van patroonherkenning bij data-analyse. Een profiel kan vooraf opgesteld zijn of zelf het resultaat zijn van een data-analyse.
– *Algoritme*: een geautomatiseerde reeks stappen die inputdata in outputdata omzet. Een algoritme 'weegt' de verschillende data ten opzichte van elkaar.
– *Machine learning*: het verschijnsel dat computers het vermogen verwerven om iets te doen waarvoor ze niet expliciet zijn geprogrammeerd. Dit werkt op basis van 'zelflerende' algoritmen.
– *Datamining*: een methode voor het vinden van anomalieën, patronen en correlaties in grote datasets.

Veiligheid

Het begrip veiligheid is in de adviesaanvraag aan de WRR niet uitgewerkt of gedefinieerd, maar de achterliggende notitie *Vrijheid en veiligheid in de digitale samenleving. Een agenda voor de toekomst*, waarnaar in de aanvraag is verwezen, gaat uit van een breed veiligheidsbegrip. Ook in dit rapport wordt veiligheid breed opgevat zonder daarbij volledigheid na te (kunnen) streven. Omwille van de breedte van de empirische voorbeelden kijkt de WRR naar (Big) Data-processen bij de politie, justitie en smart-citytoepassingen (openbare orde en veiligheid), migratiebeleid en grensbewaking, inlichtingen- en veiligheidsdiensten (nationale veiligheid) en

fraudebestrijding. Politie, justitie en de veiligheidsdiensten zitten in het hart van de overheidsverantwoordelijkheid voor veiligheid, de overige bevinden zich meer aan de randen daarvan maar kennen een grote dynamiek op het gebied van de dataverzameling, -uitwisseling en -analyse. Deze brede aanpak betekent dat er aandacht is voor een aantal verschillende juridische kaders waarin veiligheid, gegevensverwerking en privacy op een specifieke manier worden afgewogen. Met name politie, justitie en de inlichtingen- en veiligheidsdiensten vallen onder eigen wetgeving met specifieke bevoegdheden en waarborgen.

Privacy en andere fundamentele rechten

De adviesaanvraag richt daarnaast terecht de aandacht op de bescherming van de persoonlijke levenssfeer en op beginselen zoals doelbinding en dataminimalisatie om het verzamelen van persoonsgegevens aan banden leggen. Deze aandacht komt tevens tot uitdrukking in de titel van de adviesaanvraag, die behalve de begrippen 'Big Data' en 'veiligheid' ook het begrip 'privacy' omvat. Dit rapport plaatst de thematiek van de adviesaanvraag echter uitdrukkelijk in een breder kader van fundamentele rechten, waartoe ook de vrije meningsuiting, het verbod op discriminatie en het recht op een eerlijk proces behoren. Het gebruik van Big Data raakt behalve aan privacy ook aan deze en andere rechten, zeker wanneer dat binnen het veiligheidsdomein gebeurt. Zo betogen diverse auteurs dat de toepassing van algoritmes en profiling – noodzakelijk om grote databestanden te analyseren – een verhoogd risico op discriminatie en structurele benadeling met zich meebrengt (Zarsky 2014; Pasquale en Citron 2014; White House 2014). Dit betekent dat op diverse plekken in het rapport behalve privacy ook andere fundamentele rechten aan de orde zullen komen. Deze rechten vormen naast het fundamentele recht van veiligheid het kader voor de aanbevelingen in het slothoofdstuk.

1.2.2 WAT BIG DATA MET VEILIGHEID EN VRIJHEID DOET

De opkomst van Big Data heeft invloed op de vrijheid en veiligheid van burgers en de samenleving waarin zij verkeren. De thematiek van de adviesaanvraag raakt daarmee aan de grondslagen van de rechtsstaat.

De noodzaak van distantie

Vrijheid veronderstelt distantie, dat wil zeggen een zekere mate van afstand tussen de een en de ander. Distantie creëert een ruimte waarin je als mens jezelf kunt zijn, zonder voortdurend te worden gadegeslagen en zonder dat je over het eigen doen en laten verantwoording hoeft af te leggen aan anderen, bijvoorbeeld aan de overheid. In de geschiedenis van de moderne staat heeft distantie ten opzichte van degenen die ons gedrag willen observeren en dirigeren (de overheid) geresulteerd in een vergroting van de persoonlijke vrijheid. In een vrije samenleving worden burgers niet beoordeeld aan de hand van wie zij zijn. Voor het overheidstoezicht is het eerder van belang hoe burgers zich gedragen in een bepaalde relatie of situatie die de legitieme belangen van anderen raakt. In een sociale rechtsstaat betekent dit

dat gekeken moet worden naar wat de wezenlijke levensbehoeften zijn voor een menswaardig bestaan. De overige gedragingen van burgers, hun intenties en gevoelsleven zijn in dit opzicht irrelevant voor de wet. Op die manier verzekert een rechtsstaat zijn burgers een zo groot mogelijke persoonlijke vrijheid. Ook die vrijheid is een dimensie van hun veiligheid.

Tegelijkertijd heeft de overheid de dure plicht om haar burgers te beschermen en de veiligheid te vergroten, juist om ervoor te zorgen dat zij in vrijheid kunnen leven. Blootstaan aan geweld of dwang ontneemt mensen mogelijkheden om zelf, zonder pijn en angsten, inhoud te geven aan hun leven. De overheid zal daarom zorg moeten dragen voor de maatschappelijke en individuele veiligheid, onder meer door informatie in te winnen, waakzaam te zijn en bronnen van onveiligheid te bestrijden. Hierbij moet de overheid voldoende afstand houden van het persoonlijke leven van die burgers. Deze afstand onderscheidt een rechtsstaat van een totalitaire samenleving en bepaalt de mate van persoonlijke en maatschappelijke veiligheid. Het veiligheidsbeleid van de overheid moet zo ingekaderd zijn dat het zowel de persoonlijke als de maatschappelijke vrijheid beschermt. Als dat niet gebeurt, ondermijnt dat beleid immers precies datgene wat beschermd moet worden.

Veiligheidsbegrip in beweging

Veiligheid is geen statisch begrip, maar verandert mee met technologische en nationale en internationale ontwikkelingen. Zo is het veiligheidsbegip na de val van de Berlijnse Muur sterk verbreed, terwijl tegelijkertijd het vijandbeeld versplinterde (De Graaf 2011). En na de aanslagen in New York op 9 september 2001 is veiligheid nationaal en internationaal een politieke topprioriteit geworden. Daarbij is technologie in het veiligheidsbeleid een steeds prominentere rol gaan spelen. Dat geldt bijvoorbeeld voor de investeringen in grensbewaking en het managen van internationale mobiliteit (Adey 2004; Broeders 2007; Broeders en Hampshire 2013); het beveiligen van de openbare ruimte met behulp van cameratoezicht; innovaties in de jeugdzorg (Keymolen en Broeders 2013; Van Brakel 2016a); en beleid voor nationale veiligheid en antiterrorisme (Deibert 2013; DeNardis 2014; Schneier 2015; WRR 2015).

De grote en groeiende politieke aandacht voor veiligheid is eveneens een stuwende kracht achter de inzet van Big Data in het domein van veiligheid. De Verenigde Staten lopen daarin duidelijk voorop, met gigantische investeringen in dataopslag en analysecapaciteit voor veiligheidsdiensten als de NSA, CIA en FBI. Maar ook in een bredere opvatting van veiligheid worden (Big) Data-oplossingen ingezet om de veiligheid te vergroten. Denk hierbij aan de eerdergenoemde inzet van de brandweer (Mayer-Schönberger en Cukier 2013); de regulering van verkeersstromen en beveiliging van openbare ruimten in slimme steden (Kitchin 2014a);

het monitoren en indammen van epidemieën (Brownstein et al. 2009; Talbot 2014); en internationale reacties op humanitaire tragedies zoals oorlogen en natuurgeweld (Zook et al. 2010; Kaldor 2014).

Big Data kruipt burgers dicht op de huid

Big Data doet echter wel een aanslag op de beschermende werking van distantie. Tot de recente technologische doorbraken kon nog worden volgehouden dat mensen zelf in de hand hebben hoe ze bij overheidsinstanties bekend komen te staan (tenzij de overheid een wettelijke bevoegdheid heeft om iets over haar burgers aan de weet te komen). Dit veranderde met de komst van nieuwe vormen van dataverzameling, -analyse en -gebruik. De afgelopen jaren is de hoeveelheid informatie die beschikbaar en/of opvraagbaar is voor surveillance, opsporing en vervolging sterk gegroeid. De informatievergaring kan het resultaat zijn van gerichte dataverzameling ('directed'), van geautomatiseerde processen ('automated') of van vrijwillige verstrekking ('volunteered') (Kitchin 2014b: 87-98). Vooral de hoeveelheid data uit de laatste twee categorieën is exponentieel toegenomen met de komst van slimme apparaten, socialemediagebruik, geolokalisering en digitalisering van allerhande transacties en sociaal gedrag. In combinatie met goedkopere en flexibelere vormen van dataopslag, en computers die steeds ingewikkeldere gegevensbewerkingen kunnen uitvoeren, leidt dit ertoe dat overheden burgers steeds dichter op de huid kunnen zitten.

Wie de beschikbare informatiebrokstukjes statistisch bewerkt en analyseert, kan over anderen veel aan de weet komen, zonder dat zij daarover expliciet zijn geïnformeerd en/of toestemming hebben gegeven (Zuiderveen Borgesius 2015). Het produceren van nieuwe, praktisch bruikbare kennis is dé belofte van Big Data. Behalve partijen in het commerciële domein streven ook steeds vaker overheidsorganisaties die belofte na. Het object zijn van een data-analyse waarbij verschillende databronnen worden gecombineerd, gewogen, gemanipuleerd en geïnterpreteerd, doorkruist echter de beschermende werking van distantie en maakt mensen onvrij. Dat wordt sterker wanneer die geconstrueerde kennis onjuist is of alleen een statistische waarschijnlijkheid uitdrukt terwijl de betrokkene een uitzondering kan vormen. Deze mogelijkheden nemen onherroepelijk toe wanneer men benaderd wordt op basis van grootschalige statische data-analyse, zoals Big Data-processen.

Is de spanning tussen veiligheid en vrijheid onoplosbaar?

In de praktijk krijgt dit soort risico's vaak weinig aandacht. Bij de inzet van Big Data speelt namelijk steeds vaker de gedachte dat veiligheid en vrijheid onverenigbare grootheden zijn, die tegen elkaar afgewogen moeten worden (Neoclaus 2007). Zodra veiligheidsbelangen in het geding zijn lijken burgerlijke vrijheden – privacy in het bijzonder – steeds minder gewicht in de schaal te leggen (Raab 2014; WRR 2011). Privacy krijgt vooral betekenis als een 'tegenwaarde' (Vedder et al. 2007) en

het belang van veiligheid functioneert als 'tacit challenge' voor elk beleidsvoorstel (Lazarus en Goold 2007). Dit uit zich in een duidelijke trend naar verdergaande bevoegdheidsuitbreiding ten aanzien van dataverzameling, data-analyse en datagebruik op het terrein van veiligheid (Ball en Webster 2003; De Hert 2005; Schneier 2015; Lyon 2015).

De Twin Tower-aanslagen van 9 september 2001 en de aanslagen in Europa in de jaren die daarop volgden – de moord op Pim Fortuyn (2002) en Theo van Gogh (2004), Madrid (2004), Londen (2005), Parijs (2015, bij herhaling) en Brussel (2016) – hebben de uitbreiding van overheidsbevoegdheden versneld. Dit was het geval omdat de aanslagen het beeld bevestigden dat sprake was van een (informele) 'uitzonderingstoestand' (Lazarus en Goold 2007) waarin burgerlijke vrijheden – al dan niet tijdelijk – ingeperkt zouden moeten worden (vgl. Solove 2011). Dit proces van 'securitisatie', waarbij beleidsproblemen versneld behandeld worden op manieren die gangbare wettelijke en sociale regels schenden (Balzacq 2015; WRR 2015: 25-27), versterkt de in het veiligheidsbeleid reeds aanwezige tendens om bij afwegingen over het gebruik van Big Data veiligheid boven vrijheid te stellen. Bovendien werkt het in de hand dat bij de inzet van Big Data-toepassingen weinig aandacht wordt besteed aan de inrichting van de juiste 'checks and balances'. In tijden van crisis wantrouwt men deze waarborgen juist, omdat ze de effectiviteit van het veiligheidsbeleid zouden belemmeren (Solove 2011: hoofdstukken 6-9).

1.3 DE INVALSHOEK VAN DE WRR IN DIT RAPPORT

Het is niet mogelijk noch wenselijk om Big Data tegen te houden. Het uitgangspunt van dit rapport is daarom het gebruik van Big Data in het veiligheidsdomein op een verantwoorde manier in te kaderen. Hiertoe is het allereerst nodig geen afweging te hoeven maken tussen veiligheid en vrijheid. Bij het gebruik van Big Data zal de overheid beide waarden *tegelijkertijd* moeten waarborgen. Ten tweede moeten beleidsmakers bij het vaststellen van nieuwe regelgeving onderkennen dat Big Data-processen nieuwe gegevens over personen kunnen genereren die niet beschermd worden door de regulering van het *verzamelen* van persoonsgegevens. Om deze reden adviseert de WRR ten derde om de kaders voor privacy en gegevensbescherming uit te breiden: behalve de regulering van de gegevensverzameling gaat het ook om de regulering van de analysefase en het gebruik van de uitkomsten van Big Data-processen. Juist in deze twee 'fasen' van gegevensverwerking liggen namelijk de belangrijkste kansen en risico's van Big Data.

1.3.1 VOORBIJ HET AFWEGINGSMODEL TUSSEN VEILIGHEID EN PRIVACY

In dit rapport streeft de raad ernaar het beeld van een vrije samenleving niet te compartimenteren. Veiligheid en privacy zijn fundamentele rechten, waarvoor wetgeving en regeringsbeleid als geheel moeten instaan, ook al is dat niet altijd even gemakkelijk. Bij het streven naar meer veiligheid gaat het derhalve niet alleen

om het beschermen van nationale, vanuit een statelijk perspectief gedefinieerde belangen, maar ook om de veiligheid van de samenleving en van de daarin levende individuen. Naast de statelijke veiligheid draait het tevens om het welzijn en de kwetsbaarheid van 'gewone' mensen (Paris 2001; Kaldor 2007; Fukuda-Parr en Messineo 2011).

Deze individuele veiligheid of *human security* sluit aan bij de rechten van de mens en beschermt de persoonlijke rechten van burgers, ook tegen de uitwassen van het beleid van een overheid die maatschappelijke veiligheid nastreeft (Hirsch Ballin 2014). Dezelfde opvatting ligt ten grondslag aan het rapport dat Clarke et al. (2013: 14-15) in opdracht van president Obama over de Amerikaanse inlichtingendiensten schreven. Volgens de auteurs moet de Amerikaanse regering "protect, at once, two different forms of security: national security and personal privacy". Het idee van een afweging tussen beide vormen van veiligheid is volgens hen misleidend en potentieel schadelijk: in een vrije samenleving zijn sommige principes niet onderhandelbaar. Bij de inzet van Big Data zal met beide rechten rekening moeten worden gehouden. Dat is met name van belang gezien het feit dat de combinatie van veiligheid met technologie vele supporters heeft en nauwelijks steun in de rug behoeft (WRR 2011). Het zijn vooral de wettelijke waarborgen voor gegevensverwerking die bij de inzet van Big Data extra aandacht vereisen.

1.3.2 BIG DATA CREËERT NIEUWE PERSOONSGEGEVENS

De Nederlandse en de Europese wetgeving is in hoofdzaak gericht op het reguleren van het verzamelen en verwerken van persoonsgegevens die aan een individueel belang raken. Daarvoor is toestemming vereist (in de private sector, waarvan onder omstandigheden door de overheid gegevens kunnen worden gevorderd) of er is een bijzondere wettelijke machtiging verplicht. Dit is de basisnorm van de geldende wetten en verdragen. Veel gegevens zijn op zichzelf wellicht onschuldig, maar juist het in eindeloos veel combinaties kunnen analyseren van al die gegevens leidt tot het informatiepotentieel van Big Data. Gevoelige informatie kan worden afgeleid uit 'onschuldige' gegevens en levert weer nieuwe, soms gevoelige gegevens op (Hildebrandt 2008). Big Data-analyses bewerken dus niet alleen persoonsgegevens, ze genereren ook nieuwe. Dit is bijvoorbeeld het geval wanneer een statistische waarschijnlijkheid van het begaan van misdrijven wordt gekoppeld aan een persoon, dan wel aan een groep waarin individuele personen de gevolgen van die koppeling ervaren, mogelijk in de vorm van geïntensiveerde controle. Wet- en regelgeving die zich uitsluitend richt op het punt van verzameling schiet bij deze Big Data-ontwikkeling tekort. Big Data doet evident afbreuk aan normeringen die primair aangrijpen bij de verzameling en registratie van gegevens. Persoonlijke rechten voor de correctie van foutieve gegevens blijven van grote betekenis, maar behoeven aanvulling in het licht van nieuwe risico's die verbonden zijn aan de bewerking van op zichzelf 'onschuldige' gegevens, die nieuwe persoonsgegevens kunnen creëren.

1.3.3 REGULERING UITBREIDEN VAN VERZAMELEN NAAR ANALYSE EN GEBRUIK

De WRR constateert een achterstand in de wet- en regelgeving rond de *analyse* en het *gebruik* van data. Juist hierin ligt de voornaamste betekenis en waarde van Big Data. De raad adviseert daarom om naast de verzameling van persoonsgegevens ook de analyse en het gebruik daarvan nader te reguleren. De WRR zet niet in op een verdere verzwaring van de regels rondom de verzameling van gegevens, maar richt zijn pijlen op de regulering van analyse en gebruik. De verschillende fasen van Big Data-processen zijn daarmee van groot belang (zie figuur 1.1).

Figuur 1.1 De drie fasen van Big Data-processen

Deze drie fasen van Big Data-processen vereisen andere, deels nieuwe waarborgen, die in de loop van dit rapport verder worden uitgewerkt. In de rest van deze paragraaf schetsen we zeer kort de voornaamste overwegingen en aanbevelingen van de raad, die in hoofdstuk 6 in meer detail aan de orde komen.

Allereerst vereist Big Data een grotere verantwoordelijkheid van de gegevensverwerkende partijen voor de uitkomsten van Big Data-analyses. Een deel van de risico's die hiermee samenhangen, is met de huidige juridische kaders onvoldoende afgedekt. De gegevensverwerkende partijen zullen periodiek aan hun toezichthouders en het bredere publiek verantwoording moeten afleggen over de wijze waarop zij gegevens verwerken en over de effectiviteit van de door hen gehanteerde analysemethoden. De bewijslast voor de geldigheid van de uitkomsten van Big Data-analyses blijft hierbij te allen tijde bij de gegevensverwerkende partij liggen.

Bij gegevensanalyses in Big Data-processen zijn *profiling* en *datamining* nooit 100 procent accuraat. De kracht van Big Data-analyses ligt voornamelijk in algemene conclusies en structurele patronen; bij de toepassing van deze algemene beelden op concrete situaties en specifieke individuen bestaat altijd een mismatch, omdat een geconstrueerd profiel altijd zowel over- als onder-inclusief is. Er zullen dan ook nadere regels moeten worden gesteld over de toelaatbare foutmarge in het toepassingsbereik. Hiernaast is er een verklaarbare, maar onwenselijke neiging om de profielen en patronen relatief kritiekloos te volgen en computeranalyses als

quasi-objectief te beschouwen. Het is daarom nodig om nadere regels over toelaatbare foutmarges te stellen, het verbod op geautomatiseerde besluitvorming strikt te handhaven en alert te zijn op semi-automatische besluitvorming.

Goed gebruik van Big Data vereist een substantiële verzwaring van het toezicht. Toezichthouders zullen hun (technische) kennis omtrent de werking van Big Data-analyses op peil moeten brengen en houden. Dit toezicht veronderstelt, om effectief en vertrouwenwekkend te kunnen zijn, tevens een grotere mate van transparantie van dataverwerkingsprocessen. Door geheimhoudingsbepalingen binnen het veiligheidsdomein raken Big Data-analyses al snel aan het zicht van de samenleving onttrokken, een ontwikkeling die door overclassificatie steeds meer voorkomt. Dit levert spanning op met het democratisch en juridisch toezicht op veiligheidsactoren (zie bijvoorbeeld Curtin 2011; Horn 2011; Glennon 2014 en meer specifiek in het licht van data-analyse Lyon 2014; Schneier 2015; Pasquale 2015). De WRR adviseert de transparantie over de gegevensverwerking te vergroten en een beter evenwicht te zoeken tussen het vereiste van geheimhouding en het belang van openbaarheid over de uitvoering van staatstaken die aan fundamentele vrijheden raken.

Tot slot moeten burgers en organisaties meer aanknopingspunten krijgen om beslissingen gericht op en gebruikmakend van data-analyse door overheidsinstanties onafhankelijk, dat wil zeggen door toezichthoudende instanties en de rechter, te doen beoordelen. Big Data vergroot de machtsongelijkheid tussen burgers en overheid. Het is dus van belang de positie van de burger te versterken. Dit gebeurt deels door meer bevoegdheden en controlemogelijkheden aan de toezichthoudende organisaties toe te kennen en de transparantie over de gegevensverwerking te vergroten. Het is daarnaast belangrijk de stem van burger zelf in de ontwikkeling van Big Data te versterken. Dit kan gebeuren door de mogelijkheden voor algemeenbelangacties in het kader van privacy en databescherming in het recht te verruimen. Het is immers op collectief niveau dat een belangrijk deel van de consequenties van Big Data-analyses zich laat voelen.

Dit alles laat overigens onverlet dat een zorgvuldige afweging vereist is over de vraag waar en wanneer Big Data-analyses passend en legitiem zijn. Bij een positieve uitkomst is de vervolgvraag welke vormen daarvan het beste tegemoet komen aan de doeleinden binnen het veiligheidsdomein. Big Data-analyses kunnen in potentie een waardevolle bijdrage leveren aan de maatschappelijke veiligheid en vrijheid. Ze moeten echter betrouwbare, robuuste kennis opleveren, die bruikbaar is voor de besluitvorming van overheidsorganisaties. En ze moeten rusten op een stevige wettelijke basis, die helder is over de risico's die Big Data met zich meebrengt, en maatregelen bevat voor de omgang daarmee.

1.4 LEESWIJZER

De opzet van het rapport is als volgt. Hoofdstuk 2 bevat een verdere uitwerking van de hiervoor genoemde hoofdkenmerken van Big Data. Deze drie hoofdkenmerken, samengebracht onder de noemers data, analyse en gebruik, lopen als een rode draad door het verdere rapport en vormen de ruggengraat van de aanbevelingen in het slothoofdstuk.

Hoofdstuk 3 bespreekt in hoeverre overheidsorganisaties binnen het veiligheidsdomein met Big Data werken. Het gaat daarbij onder meer over de politie, inlichtingendiensten, inspecties, de Belastingdienst, transnationale databanken en samenwerkingsverbanden en/of systemen op het terrein van criminaliteits- en fraudebestrijding. De conclusie is dat Big Data in de volle breedte van de 'definitie' nog zeldzaam is, maar dat de aanzetten daartoe al duidelijk zichtbaar zijn en de ontwikkelingen de komende jaren waarschijnlijk een hoge vlucht zullen nemen.

Hoofdstuk 4 gaat in op de vraag wat de beloften en beperkingen van Big Data zijn. Wat kan er wel en niet mee en welke veiligheidsvraagstukken lenen zich het beste voor Big Data-analyses? Vervolgens komen enkele randvoorwaarden voor het gebruik van Big Data-analyses aan bod, zoals voldoende expertise, een goede inbedding in de organisaties en een adequate beveiliging van gegevensbestanden. Het hoofdstuk sluit af met een schets van de risico's van Big Data, die onder meer te maken hebben met structurele discriminatie, privacy en nieuwe vormen van machtsongelijkheid en geheimhouding, die Big Data-processen voor burgers onnavolgbaar en buitengewoon lastig te corrigeren maken.

Hoofdstuk 5 bespreekt de juridische kaders voor gegevensverwerking. Daarbij gaat het enerzijds om fundamentele rechten en algemene wet- en regelgeving voor gegevensverzameling en -verwerking. Anderzijds betreft het specifieke wet- en regelgeving die zich richt op het terrein van de veiligheidszorg, omdat daarbinnen vaak ruimere mogelijkheden bestaan voor gegevensverwerking en geheimhouding. Big Data staat op gespannen voet met een aantal bepalingen in deze kaders, waaronder doelbinding en proportionaliteit. Ook schiet het toezicht tekort en kunnen burgers tegen de uitkomsten van Big Data-analyses nauwelijks klachten indienen, omdat de nadruk te zeer ligt op (de schending van) individuele rechten en persoonsgegevens.

Hoofdstuk 6 bevat de conclusies en aanbevelingen. De belangrijkste conclusie van het rapport is dat de focus van de huidige wet- en regelgeving op de fase van gegevensverzameling in een Big Data-tijdperk niet langer volstaat. De voornaamste aanbeveling is dan ook om het regulerend kader voor Big Data te verbreden van de regulering van het verzamelen van gegevens naar de regulering van de analyse

en het gebruik van gegevens. Deze verbreding is noodzakelijk om te voorkomen dat het gebruik van Big Data een risico in plaats van een hulpbron voor een vrije samenleving gaat vormen.

NOOT

1 Het fenomeen 'Big Data' behandelen we in het voetspoor van het Engelse taaleigen als een grammaticaal enkelvoudig woord, dat overeenkomstig het Nederlandse taaleigen een veelheid van *data* (meervoud van *datum*, gegeven) omvat.

2 DE ONTWIKKELING VAN BIG DATA

2.1 INLEIDING

Dit hoofdstuk gaat nader in op het fenomeen Big Data. De nadruk ligt daarbij op een inkadering van wat er in de literatuur onder Big Data verstaan wordt en hoe het begrip in dit rapport wordt gehanteerd. Welke technologie en methoden gaan er achter Big Data schuil en in hoeverre moeten we hier spreken van een revolutie en in hoeverre van een voortzetting en versnelling van ontwikkelingen in de technologie en de statistiek die al langer gaande zijn? Om deze vragen te duiden wordt Big Data in dit hoofdstuk in de volle breedte bekeken, dus niet specifiek binnen het domein van de veiligheidszorg dat in de volgende hoofdstukken centraal zal staan. We kiezen ervoor een analytisch onderscheid aan te brengen tussen drie fasen van Big Data-processen: dataverzameling, -analyse en -gebruik. Die fasen roepen verschillende vragen op die in het werken met Big Data, en in de regulering daarvan, beantwoord moeten worden.

2.2 WAT IS BIG DATA?

2.2.1 EEN VEELHEID AAN DEFINITIES

Big Data is een wijdverspreide en veelgebruikte term. Toch bestaat er geen consensus over de betekenis en is er geen breed gedeelde definitie van Big Data (Floridi 2012; Ekbia et al. 2015). Op de vraag wat Big Data precies is, geven verschillende auteurs verschillende antwoorden. De meeste auteurs focussen op de hoeveelheid data, anderen noemen de verscheidenheid en complexiteit van de data. Weer anderen leggen de nadruk op een revolutie in methoden om met de data om te gaan (King 2016) of op de nieuwe maatschappelijke, economische en beleidsmatige mogelijkheden die ontstaan door het gebruik van Big Data.

Om een beter overzicht en begrip te krijgen van wat in de literatuur onder Big Data wordt verstaan, heeft de WRR een aantal definities van Big Data uit zowel de academische literatuur, de private sector, denktanks, private onderzoeksinstanties als overheidsinstellingen geanalyseerd.[1] Uit de analyse blijkt dat er in deze literatuur het meest wordt verwezen naar de '3 Vs' (Laney 2001). De eerste V staat voor *Volume*, het gebruik van grote hoeveelheden data. De tweede V staat voor *Variety*, het gebruik van verschillende databronnen die in verschillende structuren of zelfs ongestructureerd (zoals beeld en geluid of grote tekstbestanden) zijn opgeslagen. Ten derde gaat het om *Velocity*, de snelheid: er worden continu, soms realtime, gegevensstromen geanalyseerd. Een aantal auteurs voegt nog meer V's aan dit drietal toe, zoals *Value* (Dijcks 2012; Dumbill 2013), *Variability* (Hopkins en Evelson 2011; Tech America Foundation 2012), *Veracity* (IBM 2015) en *Virtual* (Zikopoulos en Eaton 2011; Akerkar et al. 2015). In de meeste definities ligt de nadruk op de data.

Het McKinsey Global Institute (2011: 1) hanteert de volgende definitie: "Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." Deze definitie neemt de grootte van de dataset als uitgangspunt. Wat als Big Data geldt, is volgens deze definitie gerelateerd aan tijd en technische vooruitgang en kan ook per sector verschillen. De *state-of-the-art* technieken in een bepaalde sector bepalen wat geldt als te groot om mee om te gaan. In de financiële sector bijvoorbeeld is men al veel langer gewend om met grote hoeveelheden data om te gaan en is de technische standaard veel hoger dan in bijvoorbeeld de toerismesector.

Naast definities die zich op de data oriënteren, zijn er ook definities die zich richten op de methoden en statistische analysetechnieken. Big Data impliceert een data-gedreven aanpak in plaats van de hypothese-gedreven aanpak van de traditionele wetenschappelijke methode, waarbij causaliteit vervangen wordt door correlaties en de onderzoeksvraag 'waarom' iets gebeurt vervangen wordt door 'wat' er gebeurt (Mayer-Schönberger en Cukier 2013: 7). Volgens Rubenstein (2013) verwijst Big Data naar nieuwe manieren waarop organisaties verschillende digitale datasets combineren, daar vervolgens statistische technieken als datamining en profiling op toepassen, om er uiteindelijk verborgen en verrassende correlaties uit af te leiden. Andere auteurs benadrukken de complexiteit en relationaliteit van de data. Boyd en Crawford (2012) zien Big Data als "notable not because of its size, but because of its relationality to other data". Volgens Hillard (2012) verwijst 'big' niet naar grootte maar naar complexiteit. Volgens hem kan Big Data ook klein zijn en zijn niet alle grote datasets automatisch Big Data. 'Big' verwijst dan naar het grote aantal mogelijke manieren waarop de data(bronnen) gecombineerd kunnen worden.

Enkele definities focussen op het gebruik van de data. Laney (2012) verwijst bijvoorbeeld naar verbeterde besluitvorming, de ontdekking van nieuwe inzichten en procesoptimalisatie. Anderen wijzen erop dat Big Data leidt tot een toegenomen begrip van de maatschappij en het leven (ADMA 2013) en nieuwe kennis en inzichten (Boyd en Crawford 2012; Mayer-Schönberger en Cukier 2013). Weer andere auteurs focussen op de waarde (Dijcks 2012; Dumbill 2013) of toegenomen objectiviteit, 'waarheid' en nauwkeurigheid (Anderson 2008) die Big Data kan genereren. Ten slotte zijn er auteurs die Big Data benaderen als een fenomeen of sociale beweging, een samenspel van technologie, analyse en mythologie (Boyd en Crawford 2012: 663; Ekbia et al. 2015).

2.2.2 BIG DATA ZOALS IN DIT RAPPORT OPGEVAT

Een eenduidige definitie van Big Data is dus niet gemakkelijk te geven. Er is geen duidelijke scheidslijn tussen wat wel en wat niet Big Data is. Uit de literatuur heeft de WRR enkele belangrijke kenmerken van Big Data gedestilleerd (tabel 2.1).

Tabel 2.1 Kenmerken van Big Data

Data	• Omvang van de data: het gaat om grote hoeveelheden gegevens. • Structuur van de data: het kan gaan om gestructureerde of ongestructureerde gegevens of een combinatie van beide. • Variëteit van de data: het gaat om de combinatie van verschillende databronnen.
Analyse	• Methode van analyse: de analyse is *data-gedreven*, er wordt dus gezocht naar patronen in de data zonder vooraf opgestelde hypothesen. • Oriëntatie van de analyse: hoewel Big Data-analyses ook inzicht kunnen geven in het verleden (retrospectieve analyses), zijn het met name de analyses van het heden (*realtimeanalyses / nowcasting*) en de toekomst (*predictive analyses / forecasting*) die de aandacht trekken.
Gebruik	• Ontschotting van domeinen: data uit het ene domein worden gebruikt voor beslissingen in het andere domein. • *Actionable knowledge**: conclusies op geaggregeerd niveau kunnen worden toegepast voor beslissingen op groeps- of individueel niveau (persoon of object).

* De term is van Degli Esposti (2014).

De WRR ziet Big Data als een samenspel van ontwikkelingen en niet zozeer als een vastomlijnd en definieerbaar gegeven. Er zijn niet veel praktijkvoorbeelden van toepassingen waarbij alle kenmerken uit de tabel tegelijkertijd aanwezig zijn. In de meeste gevallen gaat het om een combinatie van een beperkt aantal kenmerken. In sommige gevallen omschrijven we bepaalde datasystemen als Big Data-systemen door de mogelijkheden die in de nabije toekomst binnen handbereik liggen. Daarbij moet ook in gedachten worden gehouden dat Big Data niet enkel gaat over technologische ontwikkelingen. Het is ook een cultureel en sociaal fenomeen. Sommigen zien Big Data als een fenomeen dat grote sociale veranderingen teweeg gaat brengen, zowel positief als negatief.

2.3 WAT IS ER NIEUW AAN BIG DATA?

De grote aandacht die er momenteel voor Big Data is, wekt soms de indruk dat het fenomeen uit het niets is verschenen. Er is echter een lange ontwikkeling aan voorafgegaan. We schetsen hierna – heel kort – de historische context van Big Data en kijken vervolgens wat er nieuw is aan Big Data. Wat is evolutie en wat is revolutie?

2.3.1 VROEGE VOORLOPERS

Big Data is een samenspel van veel verschillende ontwikkelingen. Er is dan ook niet één geschiedenis van Big Data te schetsen. Zij hangt samen met de geschiedenis van automatisering, reclame, statistiek, sociale wetenschappen, artificiële intelligentie en dataverzameling (Barnes en Wilson 2014: 1; Bongers et al. 2015). De hype van Big Data gaat vaak voorbij aan deze voorgeschiedenissen, zoals treffend is verwoord in een artikel van Barnes (2013: *Big Data, little history*). De historische context kan echter wel helpen om de huidige ontwikkelingen te kunnen plaatsen.

Het verzamelen van grote hoeveelheden data is niet van gisteren. Informatie was altijd al macht en overheden en kerken waren dan ook de eerste grote verzamelaars van data (Hacking 1990). Rond 1800 voerde Napoleon de Burgerlijke Stand in voor doop-, trouw- en overlijdensakten, dienstplicht en belastingheffing. In de Verenigde Staten werd rond dezelfde tijd een tienjaarlijkse volkstelling ingevoerd (Bongers et al. 2015: 11-12). Ook in het Verenigd Koninkrijk werd begonnen met het bijhouden van bevolkingsstatistieken en andere gegevens zoals de aanwezigheid en verspreiding van ziekten in steden. Deze ontwikkelingen brachten in de loop der decennia een enorme explosie van gegevens met zich mee. Ambrose en Leta (2014) spreken van een 'avalanche of numbers'. Daarmee is ook de situatie waarin de hoeveelheid beschikbare gegevens groter is dan de techniek aankan, niet uniek voor het Big Data-tijdperk. Na de Tweede Wereldoorlog leidde de grote hoeveelheid gegevens tot de ontwikkeling van een nieuwe technologie om daarmee om te gaan, namelijk ponskaarten (Heide 2009). Elke nieuwe sprong in de omvang van data genereert nieuwe technieken en methoden.

De analysemethoden die bij Big Data gebruikt worden, wortelen in de statistiek en *machine learning*. De prille start van statistiek wordt toegeschreven aan John Graunt, die in 1662 de eerste systematische studie van sociale getallen publiceerde (Hald 2003). In de achttiende eeuw duiken de Duitse term *Statistik* en het Engelse *political arithmetic* al op, en aan het begin van de negentiende eeuw breekt statistiek in Europa door als een 'wetenschappelijke methode' (Porter 1986; Hacking 1990; Desrosières 2001). Verdere ontwikkelingen in de statistiek, maar ook doorbraken in de cryptologie tijdens de Tweede Wereldoorlog[2], hebben in de jaren vijftig en zestig geleid tot de ontwikkeling van *machine learning*: algoritmes die zelf kunnen leren. Hieruit is in de jaren negentig datamining ontstaan, een methode voor het vinden van anomalieën, patronen en correlaties in grote datasets (Fayyad et al. 1996).

Ook in het domein van de veiligheidszorg is de wens om sociaal gedrag met behulp van nieuwe technologieën in wetmatigheden te vangen ouder dan Big Data. De Belgische statisticus Quetelet, die aan de wieg stond van de doctrine van de statistische wetmatigheden, schreef in 1829 al over statistische wetmatigheden in verband met misdaad (Porter 1986). Quetelet (1835) geloofde dat het mogelijk was om op basis van onderliggende wetmatigheden normaal gedrag van abnormaal gedrag te onderscheiden. Door het verzamelen van data over bijvoorbeeld criminaliteit kon het gedrag van individuen vergeleken worden met hoe *l'homme moyen* zich zou gedragen. *La physique sociale* – zoals Quetelet het noemde – zou het mogelijk maken om niet alleen de sociale realiteit te beschrijven, maar ook de oorzaken van sociaal gedrag te identificeren (Stehr en Adolf 2015).

Verschillende aspecten van Big Data kennen dus een lange geschiedenis en ook de term Big Data *zelf* is niet helemaal nieuw. Wetenschappers van NASA gebruikten deze term al in een paper in 1997 (Cox en Elssworth 1997). Zij beschreven het probleem dat zij toen hadden met visualisaties omdat de bestaande hardware (computergeheugen, rekentijd etc.) de hele grote datasets van de NASA niet aankon (Cox en Elssworth 1997: 235; Crawford et al. 2014).

2.3.2 EVOLUTIE OF REVOLUTIE?

Big Data kan dus tot op zekere hoogte worden beschouwd als een voortzetting van langer lopende ontwikkelingen. Het is een verdere uitbreiding van de hoeveelheid beschikbare data, een voortzetting van de ontwikkeling van steeds meer geavanceerde technieken om deze data te verwerken en te analyseren, en een volgende stap in de wens om met analyses beter inzicht te krijgen in de samenleving. Het gaat echter niet alleen om een evolutie, in een aantal opzichten is Big Data ook een revolutie.

Exponentiële groei van opgeslagen data

Hoewel de afgelopen twee eeuwen steeds meer data beschikbaar kwamen, is die groei nauwelijks te vergelijken met de explosie van data op dit moment. De totale hoeveelheid beschikbare data in de wereld is moeilijk te bepalen, hoewel er verschillende speculatieve inschattingen zijn die veel worden aangehaald. IMC (2014) schat bijvoorbeeld dat de hoeveelheid opgeslagen data tussen 2013 en 2020 groeit van 4,4 zettabytes naar 44 zettabytes. Om deze hoeveelheid data enigszins aanschouwelijk te maken: als 1,2 zettabyte in boekvorm gedrukt zou worden, zou daarmee de totale oppervlakte van de Verenigde Staten in 52 lagen kunnen worden bedekt (Mayer-Schönberger en Cukier 2013: 9).

Ook de verbeterde mogelijkheden om ongestructureerde en ongelijksoortige data (zoals video's, foto's en boeken) in één analyse te verenigen dragen bij aan het revolutionaire karakter van Big Data. Data worden bovendien steeds vaker zonder vooropgezet doel verzameld, maar met het oog op mogelijke toepassingen in de toekomst. Soms vloeien gegevens voort uit andere activiteiten als een soort van digitaal 'bijproduct'. De nu al sterke groei van deze *born digital data* zal nog verder versterkt worden door de ontwikkeling van het Internet of Things (zie 2.4.2).
De alomvattendheid van de dataverzameling betekent dat er voor burgers steeds minder handelingen zijn die niet op een of andere manier worden vastgelegd. Er vindt een dataficatie van alledaagse handelingen plaats (Stehr en Adolf 2015) die gepaard gaat met een zekere onzichtbaarheid van de dataverzameling zelf. De dataverzameling vindt bijna ongemerkt plaats, buiten de controle en toestemming en vaak ook buiten het zicht van het individu om (PCAST 2014: 5; Stehr en Adolf 2015; Schneier 2015).

Een scherper beeld

De toegenomen hoeveelheid data (kwantiteit) zorgt voor een kwalitatieve sprong in de analyse. Dit wordt wel *the unreasonable effectiveness of data* genoemd (Halevy et al. 2009). Matig presterende algoritmen hebben met hele grote hoeveelheden data betere uitkomsten dan betere algoritmen met kleinere hoeveelheden data (Banko en Brill 2001; Leek 2015; Halevy et al. 2009). Big Data verbetert dus de werking van algoritmen, zonder dat de algoritmen zelf verbeterd worden. Waar bij steekproeven vaak alleen over het geheel of grote subgroepen van de populatie uitspraken kunnen worden gedaan – omdat bij verdere doorsneden de groep snel te klein wordt om nog zinvolle uitspraken te doen – kan Big Data een veel gedetailleerder beeld leveren. Een gedetailleerd beeld betekent overigens niet per definitie dat dat beeld ook tot in detail correct is. De kracht van Big Data ligt in de nauwkeurigheid van het beeld op geaggregeerd niveau. Het vertalen van die inzichten naar kleinere groepen of zelfs individuen – iets dat in het veiligheidsbeleid vaak gewenst is – gaat met haken en ogen gepaard.

Data-gedreven analyse

Een belangrijk kenmerk van Big Data is de data-gedreven analyse, een heel andere benadering dan de traditionele statistische methode. Het doel van de analyse is niet het toetsen van hypotheses maar het vinden van interessante verbanden en patronen, waarvan de gedachte is dat ze relevant zijn voor een sterk uitdijend aantal maatschappelijke domeinen. Dergelijke analyses kunnen interessante en onverwachte correlaties en inzichten opleveren, maar de causaliteit is – zeker in eerste aanleg – niet duidelijk en het risico bestaat dat correlaties tot causaliteiten verheven worden. Of met deze aanpak sprake is van een epistemologische revolutie of dat het slechts een voortzetting is van eerdere ontwikkelingen, is onderwerp van discussie (zie bijvoorbeeld Anderson 2008).

Wegvallen van grenzen tussen domeinen

Het laatste 'nieuwe' aspect aan de huidige ontwikkelingen is het wegvallen van de grenzen tussen domeinen. Zo kan informatie die verzameld is tijdens het winkelen van mensen gebruikt worden om te bepalen of mensen wel of niet een verzekering krijgen. DNA-data kunnen aangewend worden om te bepalen of iemand een baan krijgt (PCAST 2014). De laatste generatie slimme meters voor elektriciteit en gas leest niet alleen het energieverbruik van de verschillende huishoudelijke apparaten uit, maar ook het merk van de desbetreffende apparaten. Zo kan de energieleverancier ook iets over het koopgedrag van dezelfde consument te weten komen (Stehr en Adolf 2015). De kracht van Big Data ligt in het wegvallen van deze grenzen, want dat levert een rijker beeld van de werkelijkheid op. Maar juist op dit punt dienen zich ook risico's aan. Het wegvallen van grenzen tussen domeinen is het resultaat van de toename van technische mogelijkheden voor datakoppeling, politieke bereidheid om data te koppelen en een groeiende markt voor de uitwisseling

en verkoop van data (Kitchin 2014a). De eerdere data-explosie in de negentiende eeuw ging niet gepaard met een dergelijke commodificatie van data (Ambrose en Leta 2014).

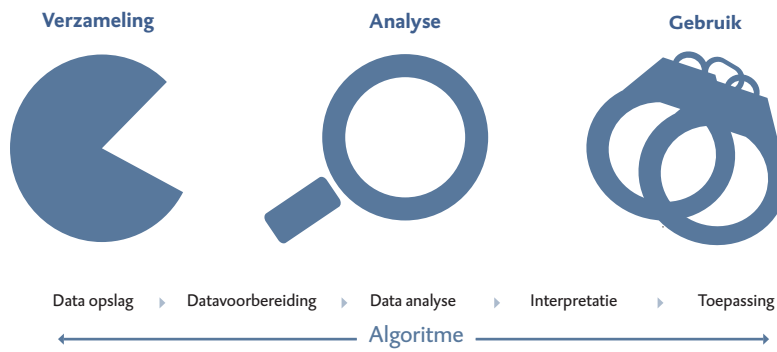
De technologische ontwikkelingen zijn zeker revolutionair te noemen. Of Big Data daadwerkelijk een revolutie zal blijken, hangt sterk af van wat wel de 'domesticatie' van technologie wordt genoemd: het doorbreken van deze technologie in de dagelijkse praktijk van de vele commerciële en publieke organisaties die het fenomeen nu verkennen.

2.4 BIG DATA-TECHNOLOGIE: VERZAMELING, ANALYSE, GEBRUIK

2.4.1 ANALYTISCH ONDERSCHEID TUSSEN VERZAMELING, ANALYSE EN GEBRUIK

Achter de term Big Data gaan meerdere stappen schuil: dataverzameling, datavoorbereiding, data-analyse, interpretatie, en het gebruik van deze inzichten (Jagadish et al. 2014; zie ook Akerkar et al. 2015). We brengen die fasen in dit rapport terug tot een analytisch onderscheid tussen verzameling, analyse en gebruik (figuur 2.1).

Figuur 2.1 Big Data-proces in fasen



De figuur toont het proces van dataverzameling, analyse en gebruik als een opeenvolging in fasen, maar in de praktijk komen er korte en lange *loops* voor in het proces. Een analyse kan bijvoorbeeld leiden tot het kiezen van andere databronnen en de interpretatie van de analyses kan leiden tot aanpassingen van de dataverzameling, opslag en datavoorbereiding. Door het hele Big Data-proces heen spelen algoritmen, een geautomatiseerde reeks stappen die inputdata in outputdata omzet, een vitale rol. Omdat het om zulke grote hoeveelheden data gaat, is deze automatisering onmisbaar. Algoritmen kunnen een enkele stap automatiseren, of zelfs alle verschillende stappen automatiseren en zo snel achter elkaar laten

doorlopen, dat zij in tijd nauwelijks meer te onderscheiden zijn. Het is in de praktijk dus vaak een meer iteratief proces dan deze driedeling suggereert. Toch is het nuttig om het analytische onderscheid hieronder verder uit te werken.

2.4.2 DATAVERZAMELING

Nieuwe databronnen…

De hoeveelheid beschikbare data is enorm toegenomen. Waar data een aantal decennia geleden nog moesten worden verzameld door onderzoek en tellingen en handmatig moesten worden ingevoerd, worden data nu automatisch geproduceerd en zijn vaak zelfs het bijproduct geworden van veel dagelijkse handelingen. De meeste data zijn nu *born digital data*, dat wil zeggen dat de data digitaal ontstaan zijn en niet voortkomen uit analoge data (PCAST 2014). Het dagelijks gebruik van internet, sociale media, mobiele telefoons en verschillende applicaties genereert enorme hoeveelheden data. Met de ontwikkeling van GPS, zendpalen van mobiele telefoons en wifinetwerken is het bijna permanent mogelijk om de locatie van telefoons en apparaten te bepalen (Mayer-Schönberger en Cukier 2013: 88). Surfgedrag op het internet wordt door middel van cookies en meer geavanceerde technieken verzameld. Zoektermen, uitingen op sociale media, e-mails, sms'en, muisklikken en onlineaankopen worden bewaard om eventueel op een later tijdstip te worden geanalyseerd op geaggregeerd en soms ook op individueel niveau.

Naast de computer en de telefoon zijn er steeds meer apparaten die met het internet verbonden kunnen worden en data produceren die realtime verzonden en uitgelezen kunnen worden. Dit wordt het Internet of Things genoemd. De smart tv verzamelt informatie over kijkgedrag, de smart meter over energiegebruik en de verschillende digitale boek-readers over leesgedrag. De verwachting is dat deze ontwikkeling zich de komende jaren verder zal uitbreiden (Moessner et al. 2015). Doordat technologie steeds geavanceerder en goedkoper wordt, is het mogelijk om in steeds meer apparaten software en draadloze communicatie en radiotransmitters in te bedden en ze te verbinden met de algemene internetstructuur (Regalado 2014). Ook gerelateerde ontwikkelingen als de vooruitgang in sensortechnologie, die fysieke fenomenen naar digitale data vertaalt, levert nieuwe data, zeker wanneer deze sensoren met de internetinfrastructuur zijn verbonden (PCAST 2014). Daarnaast zijn er ook steeds meer mogelijkheden om informatie te halen uit data die als ongestructureerd te boek staan. Door middel van nieuwe algoritmen worden bijvoorbeeld teksten (*natural language processing*) en foto's omgezet in digitale data.

…in combinatie met 'oude data'

Naast nieuwe bronnen van data zijn er andere ontwikkelingen die bijdragen aan de beschikbaarheid van data. Ten eerste worden de data van de overheid binnen de overheid steeds vaker gedeeld en voor verschillende overheidsdiensten beschik-

baar gemaakt. Data vloeien steeds vaker en rijkelijker over de grenzen van departementen, diensten, registraties en soms zelfs wettelijke bevoegdheden heen (WRR 2011). In de tweede plaats worden meer en meer databanken die traditioneel in handen zijn van de overheid of andere grote partijen toegankelijk gemaakt voor het grote publiek. Zo bieden openbaarvervoerbedrijven realtime-informatie aan over de planning van treinen, trams, metro's en bussen en biedt de overheid op data.overheid.nl 7015 verschillende datasets aan.[3] Deze beweging wordt ook wel met 'open data' aangeduid. Bedrijven maken daar gebruik van door deze data te combineren met andere (eigen) data, voor eigen gebruik of om deze weer door te verkopen of te verhuren (Kitchin 2014b).

Dataopslag

Tegelijkertijd nemen de mogelijkheden om data op te slaan toe en nemen de kosten van deze dataopslag af. Er wordt gezegd dat de capaciteit om data op te slaan ongeveer iedere drie jaar verdubbelt (Akerkar et al. 2015). Overigens neemt de dataproductie veel sneller toe dan de opslagcapaciteit, dus het opslaan en bewaren van data blijft een belangrijke technische uitdaging. Door het ontwikkelen van technieken waarmee data op meerdere plekken parallel kunnen worden geanalyseerd (zoals MapReduce en het daarop gebaseerde Hadoop) is het mogelijk geworden om data op verschillende plaatsen op te slaan (Madden 2012). Hierdoor is dataopslag goedkoper en kunnen ook data die te groot zijn voor één machine worden opgeslagen en geanalyseerd. Analyses kunnen dan op meerdere locaties tegelijkertijd worden uitgevoerd en samengevoegd (Rusitschka en Ramirez 2014). Een belangrijke ontwikkeling hierbij is het opslaan van data in de *cloud*. Hoewel de naam een niet-fysieke opslagplaats suggereert, gaat het om het opslaan van data in grote datacentra die worden beheerd door een internetbedrijf. Microsoft, Apple en Google bouwen alle grote datacentra in Nederland. Google bouwt bijvoorbeeld een datacentrum in Delfzijl dat meer dan veertig voetbalvelden groot is (Ministerie van Economische Zaken 2014). De kosten van het opslaan in zo'n datacentrum zijn lager dan het zelf opslaan van een grote hoeveelheid data. Wel komt hierbij een ander soort beveiligingsvragen kijken, omdat de beveiliging de facto is uitbesteed aan de aanbieder van de clouddienst (Rusitschka en Ramirez 2014).

Databeveiliging

De waarde van data – in het bijzonder persoonlijke data – maakt databeveiliging een belangrijke technische uitdaging. Deze beveiliging gebeurt in het algemeen door controle op de toegang tot data en netwerken en door encryptie van de data, zowel in transit als in rust. Dat wil zeggen dat de data versleuteld worden en alleen met een bepaalde *key* leesbaar zijn (Kitchin 2014b: 175; Gürses en Preneel 2016). De enorme groei van data en de vele apparaten die met het internet verbonden zijn maken beveiliging een permanente opgave. Criminelen en de beveiligingssector zijn in een wedloop verwikkeld (Kitchin 2014b: 175). Big Data is in die zin ook een potentiële bron van onveiligheid: het genereert nieuwe kwetsbaarheden voor

bedrijven en organisaties (grootschalig verlies van data en reputatieschade) en individuen (openbaring van persoonlijke informatie, reputatieschade en kwetsbaarheid voor chantage). De meldplicht voor datalekken bij de Autoriteit Persoonsgegevens die per 1 januari 2016 is ingegaan is een reflectie en een eerste antwoord op deze nieuwe kwetsbaarheden.[4] Van razendsnelle toekomstige kwantumcomputers wordt verwacht dat zij de volgende grote stap zijn in deze wedloop tussen beveiliging en (criminele) aanvallen. Een groot deel van de reguliere encryptiesleutels wordt door kwantumcomputers volstrekt onveilig. In eerste instantie zijn met name Public Key-algoritmen kwetsbaar terwijl symmetrische versleuteling de dans ontspringt. Naarmate kwantumtechnologie zou vorderen kan echter niet uitgesloten worden dat ook symmetrische versleuteling in de gevarenzone komt. Dus tegelijkertijd met het ontwikkelen van de kwantumcomputers wordt ook gewerkt aan nieuwe cryptografische algoritmes die bestand zijn tegen kwantumcomputers (Gurses en Preneel 2016). Het is overigens nog niet duidelijk wanneer de eerste kwantumcomputer er zal zijn en van wie deze zal zijn. Voor de hand liggende kanshebbers zijn grote internetbedrijven als Google, academische consortia en inlichtingendiensten als de NSA.

Datavoorbereiding

Om data te kunnen analyseren, moeten de data eerst daartoe voorbereid worden. Hoewel de technieken om ad hoc data te integreren verbeteren, blijven hier belangrijke technische uitdagingen bestaan. Voordat datamining begint, wordt er vaak een *data warehouse* gevormd waarin verschillende databases uit verschillende bronnen worden geaggregeerd. Hierna maakt men de data schoon en filtert men onbetrouwbare informatie eruit (Zarsky 2003). Deze stap is cruciaal voor het kunnen analyseren van grote hoeveelheden data en voor de validiteit van de analyse. Als de data niet op de juiste manier worden gekoppeld, data onterecht met elkaar in verband worden gebracht of belangrijke metadata verloren gaan, gaat ook de waarde van de analyse verloren. Ondanks het vele werk op het gebied van standaardisatie en interoperabiliteit zijn er nog steeds veel datasets die niet gecombineerd kunnen worden (Kitchin 2014b: 156).

Het anonimiseren van gegevens is een belangrijke stap in het voorbereiden van de data. Veel verschillende methoden worden gebruikt om data te de-identificeren zoals anonimisatie, pseudonimisatie, encryptie, *key-coding* en *data sharding* (Tene en Polonetsky 2013). Hierbij worden beschrijvende kenmerken eruit gehaald – zoals bijvoorbeeld de naam of het exacte adres – en worden ze soms vervangen door een lange code of andere sleutel. Er is momenteel nogal wat discussie over de vraag in hoeverre anonimiseren van data in deze tijd nog effectief is. Een veel aangehaalde studie is *Unique in the Crowd* (De Montjoye et al. 2013), waarin de auteurs laten zien dat vier willekeurige tijd- en locatiedatapunten voldoende zijn om 95 procent van de individuen te identificeren in een database met een half miljoen mobiele datarecords. Tegelijkertijd tonen andere auteurs aan dat de mogelijkheden

om te de-anonimiseren overschat worden (Narayanan en Shmatikov 2008; Ohm 2010; Tene en Polonetsky 2013; Article 29 DP Working Party 2014; Cavoukian en Castro 2014).

2.4.3 DATA-ANALYSE

De tweede stap in Big Data-processen is het analyseren van data. Analyse is de cruciale stap tussen grote hoeveelheden data en het gebruik en de waarde daarvan (Degli Esposti 2014). In de analysefase bepalen de algoritmen en de geselecteerde databronnen welke informatie uit de analyse komt. Deze stap wordt ook wel omschreven als *Knowledge Discovery in Databases (KDD)* of datamining. Dit zijn min of meer inwisselbare termen, hoewel *Knowledge Discovery in Databases* soms wordt beschouwd als een bredere term die bijvoorbeeld ook de datavoorbereiding omvat (Fayyad et al. 1996; Hand et al. 2001).

Datamining bestaat vaak uit een combinatie van kleinere en grote analyses. Deze analyses zijn vaak voor een groot gedeelte geautomatiseerd, maar ook mensenwerk speelt een rol van betekenis. De analist/opdrachtgever/*computer scientist* blijft belangrijk voor het beoordelen en begeleiden van het proces en het interpreteren van de eerste resultaten (Kitchin 2014b: 103). Het probleem of de vraag die voorligt, moet immers vertaald worden in een formele taal die de computer kan begrijpen. Dit betekent dat zowel domeinkennis als kennis van de formele computertaal (code) van groot belang is. Deze kennis is lang niet altijd in dezelfde persoon verenigd en vereist dus ook communicatie en een gemeenschappelijke taal. Het voert te ver om in dit hoofdstuk alle verschillende data-analysetechnieken zoals verschillende typen *neural networks* en lineaire regressie te bespreken. Om de essentie van Big Data-analyses te begrijpen zijn vooral twee aspecten van belang: patroonherkenning en voorspellende analyse.

Patroonherkenning en profiling

De analysemethoden voor archetypische Big Data-analyses zijn data-gedreven. Dat wil zeggen dat men zoekt naar de patronen die uit de data oplichten, in plaats van dat men een van tevoren geformuleerde hypothese test (Mayer-Schönberger en Cukier 2013). Het gaat bij patroonherkenning om het herkennen van correlaties in de data. Het zegt niets over causaliteit (PCAST 2014: 25). De twee belangrijkste typen van patroonherkenning zijn clusteren en associatieregels. Bij *clusteren* scannen algoritmen de dataset en zoeken ze naar gelijkenissen tussen de variabelen met als doel om gelijksoortige variabelen te groeperen. Op deze manier maken algoritmen verschillende initiële indelingen en geven daarbij de sterkte van de correlatie en mogelijke overlappingen tussen clusters aan. Deze methode wordt bijvoorbeeld gebruikt om categorieën van consumenten te maken uit grote hoeveelheden ongestructureerde data (Kitchin 2014b:103). Bij *associatieregels* doorzoeken algoritmen de data om patronen te laten zien van variabelen die gecorreleerd zijn. De algoritmen controleren of er regels zijn die de relatie tussen de verschillende variabelen

kunnen beschrijven. Bijvoorbeeld: als A en B aanwezig zijn, dan is de kans dat C aanwezig is 32 procent (Zarsky 2013). Deze vorm wordt gebruikt door winkels om op basis van een grote database van aankopen te bepalen welke producten vaak samen worden gekocht (McKinsey Global Institute 2011: 28). Wanneer patroonherkenning op personen wordt toegepast spreekt men van *profiling*. Zo kan een grote database worden doorzocht op identificeerbare karakteristieken van verdachte criminelen of terroristen. Wanneer de hoeveelheid data groot is, kunnen er hele specifieke profielen worden gemaakt. Zo biedt Twitter adverteerders de mogelijkheid om te adverteren aan meer dan 1000 verschillende doelgroepen (Boston 2015).

Voorspellende analyse

Niet iedere data-analyse beoogt hetzelfde doel. Er kan grofweg een onderscheid worden gemaakt tussen een beschrijvende (descriptieve) analyse, een verklarende (diagnostische) analyse en een voorspellende (predictieve) analyse (Rusitschka en Ramirez 2014; zie ook Kitchin 2014b: 101). Een beschrijvende analyse wil een kenmerk van de dataset beschrijven zonder verdere interpretatie. Een verklarende analyse beschrijft welke factoren van invloed zijn op een bepaalde uitkomst. Een voorspellende analyse wil op basis van bepaalde kenmerken een voorspelling doen over andere kenmerken van een specifiek item in de data, bijvoorbeeld over een plaats of een individu. Soms wordt hier nog een vierde categorie aan toegevoegd: prescriptieve analyse. Deze analyse gaat nog een stap verder dan de voorspellende analyse. Bij een prescriptieve analyse wordt op basis van een voorspelling een bepaald handelen voorgeschreven (Rusitschka en Ramirez 2014).

Big Data biedt mogelijkheden voor de verschillende typen analyses. Het CBS onderzoekt bijvoorbeeld of het uitlatingen op sociale media kan gebruiken om consumentenvertrouwen te meten (Van Sandijk 2013). De meeste aandacht gaat echter uit naar de voorspellende – en eventueel zelfs prescriptieve – analyses die met behulp van Big Data te doen zijn. Bij een voorspellende analyse bouwt men eerst een model. Een belangrijke bouwsteen voor zo'n model vormen de patronen die door middel van automatische data-analyse herkend worden. Het is ook mogelijk een model te bouwen op basis van *expert judgement*. Zo'n *expert driven* model kan wel aan de hand van data getest worden. Zo heeft de politie op basis van een kwalitatief onderzoek naar drugshandel in Maastricht een model ontwikkeld waarmee voorspeld kan worden welke auto's op de weg drugs brengen naar Maastricht.

Het zijn echter de data-gedreven modellen waar bij Big Data de meeste aandacht naar uitgaat. Op basis van de gedestilleerde patronen uit de data wordt een model gebouwd dat bepaalde eigenschappen van nieuwe items in de data kan voorspellen (Han et al. 2011, geciteerd in Kitchin 2014b: 103). Het voorspellen kan betrekking hebben op een bepaalde categorie (*classificiation*), bijvoorbeeld of een creditcard-

transactie een bonafide betaling of een frauduleuze actie is. Ook kan de voorspelling gaan over het toekennen van een bepaalde waarde, bijvoorbeeld een kwalificatie van de kredietwaardigheid van een persoon. Het nieuwe van Big Data-analyse zit dus in de combinatie van patroonherkenning in grote hoeveelheden data en het bouwen van voorspellende modellen op basis van die analyse. Hoewel dit analytisch twee te onderscheiden stappen zijn, vinden ze in de praktijk regelmatig gecombineerd plaats. Als de feedbackloop door het algoritme zelf wordt gedaan, spreekt men wel van *machine learning*.

2.4.4 GEBRUIK

Big Data-analyses worden zelden ondernomen zonder een vooropgezet doel. De uitkomsten mogen dan soms verrassend zijn – onverwachte correlaties – maar de zoektocht zelf dient een doel, zoals het behalen van winst of het vergroten van de veiligheid. In de regel is het de bedoeling dat de analyse leidt tot *actionable knowledge* en dus tot actie. Het gebruik is daarmee het minst technische onderdeel van Big Data-analyses, maar vanuit maatschappelijk perspectief wellicht de meest belangrijke fase. In het gebruik komen een aantal van de voorafgaande onderscheiden weer terug. Drie aspecten zijn in ieder geval van belang: de mate van toekomstgerichtheid, de mate van automatisering van het gebruik en de potentiële consequenties, oftewel de impact, van het gebruik van Big Data op groepen en individuen.

Toekomstgerichtheid

Een data-analyse kan worden gebruikt om het verleden of het heden te begrijpen. Zo'n analyse die terugkijkt is meestal verklarend van aard, maar kan ook nieuwe input opleveren voor verdergaande analyses. Een data-analyse kan ook worden aangewend om concreet te interveniëren in het heden op basis van een realtime-analyse of op basis van een analyse van een verwachte toekomst. Big Data-toepassingen die zich op realtime verkregen informatie baseren, richten zich in de regel ook op realtime-actie. De analyse van verkeersstromen leidt per ommegaande tot adviezen aan verkeersdeelnemers om het wegennet te ontlasten. Maar ook hier geldt dat deze analyses input kunnen zijn voor verdere analyses en profielen. De grote belofte van Big Data is namelijk zijn toekomstgerichtheid. Met slimme analyses van data kunnen verwachtingen over de toekomst in het heden al geadresseerd worden. Voorspellende modellen zijn uiteraard niet nieuw – denk aan verkeersmodellen – maar worden door grotere datasets rijker en scherper, hetgeen ten goede kan komen aan de interventie die erop gebaseerd is. Afgezien van het feit dat de toekomst nooit echt voorspelbaar is, hangt de inschatting van wat een goed en legitiem gebruik is van voorspellende analyses sterk af van andere factoren.

Automatisering

Sommige Big Data-toepassingen worden automatisch omgezet in beslissingen. Amazon raadt boeken aan op basis van algoritmes over het eigen gebruik en dat van anderen, daar zit verder geen menselijke beslisser meer tussen. Andere toepassingen nemen op basis van analyses min of meer automatische besluiten. Bij praktijken van *redlining*, waarbij bepaalde profielen leiden tot negatieve adviezen of het ontzeggen van hypotheken of leningen, wordt vaak niet van het profiel en de analyse afgeweken. De persoon die het besluit mededeelt, heeft vaak geen echte mogelijkheden om af te wijken van dit – illegale – advies. Computer says no. Overigens zijn er ook zorgen over het gebruik van Big Data in de verzekeringsbranche: door de rijke inzichten voortvloeiend uit Big Data-analyses kunnen verzekeringsmaatschappijen veel scherper de goede van de slechte risico's onderscheiden hetgeen de solidariteitsgedachte van verzekeringen onder druk zet (Timmer et al. 2015). Zeker waar de impact op de levenskansen van mensen groot is, is het nemen van automatische besluiten verboden, conform artikel 15 van de Databeschermingsrichtlijn. Dat geldt ook voor de overheid over de volle breedte.

Impact van besluiten

De (potentiële) impact van het gebruik van Big Data op het leven van mensen is bepalend voor hoe we naar dat gebruik kijken. Hoe groter de impact, hoe strenger er naar Big Data-processen gekeken moet worden. De meeste ruimte ligt daarmee bij gebruik dat relatief weinig impact heeft, maar waar door de macht van het getal toch belangrijke winst – financieel of anderszins – te behalen valt. De suggesties van Amazon kunnen zonder probleem automatisch gegenereerd worden omdat de impact op levenskansen verwaarloosbaar is. Het ergste dat je kan gebeuren is de aanschaf van een slecht boek. Ook voor Amazon is het geen bezwaar dat men er vaak naast zit met een aanbeveling; de extra inkomsten van klanten die de aanbevelingen wel volgen, wegen op tegen de investering in de technologie. Naarmate de impact van het gebruik van Big Data-analyses op de levenskansen van mensen groter wordt, veronderstelt dat hogere eisen aan de analyse, zeker als de analyse uitspraken doet over verwacht, toekomstig gedrag. Als bijvoorbeeld een analyse verwachtingen genereert over inbraken in een bepaalde buurt, dan is een voorlichtingscampagne in die buurt over hang- en sluitwerk minder problematisch dan het onder politietoezicht stellen van mensen die aan een bepaald daderprofiel voldoen.

2.5 CONCLUSIE

Big Data is een relatief nieuwe term die vele definities kent waarover nog weinig consensus bestaat. De WRR kiest er in dit rapport voor om een drietal elementen van Big Data-processen centraal te stellen. Big Data gaat over (1) de omvang, structuur, variëteit en *verzameling* van data; (2) de methoden en de oriëntatie van de *analyse*; en (3) het *gebruik* van/op basis van Big Data-analyse, oftewel hoe de ana-

lyse wordt omgezet in interventies in de dagelijkse leefomgeving van klanten en burgers. Zoals de meeste technologische innovaties heeft ook Big Data lange historische wortels in eerdere ontwikkelingen. Dat geldt zowel voor de hard- en software als voor de mathematische modellen die aan Big Data-analyses ten grondslag liggen. Big Data is daarmee deels evolutie en deels revolutie. Hoe die verhouding binnen de verschillende elementen van Big Data ligt, is nog niet overal duidelijk. Binnen het domein van de dataverzameling tellen de verschillende kwantitatieve sprongen – de verzameling en opslag van veel, gevarieerde en ongestructureerde data – waarschijnlijk op tot een kwalitatieve sprong. Binnen het domein van de analyse komen onder de vleugels van langer bestaande technieken en algoritmen, nieuwe methoden van analyse op zoals zelflerende algoritmen en *machine learning*. Deze kunnen wellicht als *game changers* beschouwd worden, zeker als deze in de toekomst de norm voor veel toepassingen worden. In het domein van het gebruik van Big Data hangt veel af van hoe Big Data-analyses worden gebruikt (om het verleden te begrijpen of om de toekomst te voorspellen) en van wat de impact van beslissingen op basis van Big Data-analyses op het dagelijks leven van burgers en consumenten is. Deze vragen zijn niet zinvol in abstracto te beantwoorden en vereisen casuïstiek om uitspraken te kunnen doen. In dit hoofdstuk is gekeken naar Big Data als fenomeen dat zich zowel in het commerciële domein als in het overheidsdomein afspeelt. In de hierna volgende hoofdstukken ligt de focus veel sterker op de overheid, meer specifiek op het gebruik van Big Data-oplossingen in het domein van de veiligheidszorg.

NOTEN

1 De analyse is gebaseerd op definities uit de volgende bronnen: ADMA (2013), Akerkar et al. (2015), Andrejevic en Gates (2014), Article 29 DP Working Party (2014), Boyd en Crawford (2012), Crawford en Schultz (2013), Diebold (2000), Dijcks (2012), Ekbia et al. (2015), European Commission (2014), European Data Protection Supervisor (2014), Hillard (2012), IBM (2015), Jacobs (2009), Information Commissioner's Office (2014), King (geciteerd in Shaw 2014), Kitchin (2013), Kraska (2013), Laney (2001; 2012), Mayer-Schönberger en Cukier (2013), McKinsey Global Institute (2011), Microsoft (2012), Oxford English Dictionary, Richards en King (2014), Rubenstein (2013), SAP (2014), De Swart (geciteerd in Bakker 2014), TechAmerica Foundation (2012), White House (2014) en Wikipedia (2016).

2 Alan Turing was een wiskundige en cryptoloog die tijdens de Tweede Wereldoorlog een doorslaggevende rol speelde in het breken van de Duitse militaire codetaal. Hij deed dit werk voor de Britse Government Code & Cypher School, oftewel Bletchly Park, waar hij voor dat doel een van de eerste computers bouwde, een zogenoemde Turing Machine.

3 Zie data.overheid.nl, geraadpleegd op 26 oktober 2015.

4 Deze meldplicht houdt in dat organisaties (zowel bedrijven als overheden) direct een melding moeten doen bij de Autoriteit Persoonsgegevens zodra zij een ernstig datalek hebben. En soms moeten zij het datalek ook melden aan de betrokkenen (de mensen van wie de persoonsgegevens zijn gelekt). Zie: https://autoriteitpersoonsgegevens.nl/nl/melden/meldplicht-datalekken.

3 BIG DATA IN HET VEILIGHEIDSDOMEIN

3.1 INLEIDING

In veel maatschappelijke domeinen groeit de belangstelling voor het gebruik van Big Data. González Fuster en Scherrer (2015: 14) spreken in een onderzoek naar Big Data, slimme apparaten en privacy van een 'variety of expanding domains'. Het veiligheidsdomein vormt hierop geen uitzondering. Naast particuliere organisaties die druk bezig zijn om Big Data-toepassingen te ontwikkelen, zijn het binnen dit domein vooral overheidsorganisaties en samenwerkingsverbanden van overheidsorganisaties die het voortouw nemen. Het veiligheidsdomein omvat in dit hoofdstuk hoofdzakelijk nationale veiligheid (*national security*), criminaliteits- en fraudebestrijding (*security*) en openbareordevraagstukken (*social security*). Om zicht te krijgen op de systemen, organisaties en organisatieverbanden die zich in Nederland met Big Data-processen bezighouden, heeft de WRR diverse cases verkend. Af en toe komen in dit hoofdstuk ook buitenlandse voorbeelden aan de orde, omdat sommige Nederlandse initiatieven zijn geïnspireerd op experimenten in andere landen. Dit geldt bijvoorbeeld voor de politie, die een systeem gebruikt dat is afgeleid van Big Data-toepassingen in de Verenigde Staten en het Verenigd Koninkrijk. In andere gevallen hebben databestanden bijna als vanzelf een grensoverschrijdend karakter, zoals data aangaande internationale mobiliteit, en worden nieuwe toepassingen dientengevolge ook deels op een bovennationale schaal ontwikkeld en toegepast.

Het eerste en meest omvangrijke deel van dit hoofdstuk bestaat uit een reeks casusbeschrijvingen van organisaties die op enigerlei wijze met Big Data experimenteren. Zij werken met zeer omvangrijke en vaak gekoppelde datasets, zoeken naar nieuwe, onverwachte verbanden en patronen, en gebruiken data-analyses voor het doen van voorspellingen over toekomstig gedrag. Het beeld dat hieruit oprijst is vanzelfsprekend niet meer dan een momentopname, temeer omdat Big Data-technologieën zich in hoog tempo ontwikkelen. Van volbloed Big Data-processen blijkt in veel gevallen nog geen sprake te zijn, maar de eerste aanzetten daartoe zijn al duidelijk zichtbaar. De ontwikkelingen gaan echter snel. Wat vandaag nog onmogelijk is – zowel in politieke als in technologische zin – kan morgen uitvoerbaar blijken, omdat steeds meer data worden gedigitaliseerd, dataopslag steeds goedkoper wordt en snellere en krachtigere computers steeds geavanceerdere bewerkingen kunnen uitvoeren. Het tweede deel van dit hoofdstuk bevat daarom een vooruitblik naar hoe de situatie er over een jaar of vijf uit zou kunnen zien. Die vooruitblik is naar zijn aard speculatief en schetsmatig en dient vooral om scherp te krijgen welke veranderingen het gebruik van Big Data met zich mee-

brengt. De slotparagraaf bevat een eerste inventarisatie van meer algemene kwesties die zijn verbonden aan het gebruik van Big Data-toepassingen in het veiligheidsdomein.

3.2 BIG DATA IN HET VEILIGHEIDSDOMEIN: 7 CASES

De WikiLeaks-onthullingen van Edward Snowden hebben duidelijk gemaakt dat sommige nationale inlichtingendiensten op grote schaal data verzamelen, analyseren en uitwisselen, met als doel de nationale veiligheid te bewaken. Voor de Nederlandse inlichtingen- en veiligheidsdiensten geldt dat zij technisch in staat zijn om grote hoeveelheden data te verwerken, met als doel daaruit betekenisvolle patronen af te leiden. Volgens de thans geldende wetgeving mogen zij echter niet ongericht grote hoeveelheden kabelgebonden data onderscheppen, hoewel er een wetswijziging in de maak is om dit in de toekomst mogelijk te maken (Jacobs 2016). Daarnaast zijn er binnen het veiligheidsdomein vele andere organisaties en samenwerkingsverbanden van organisaties die soortgelijke activiteiten ontplooien. De onderstaande zeven cases maken duidelijk dat data-analyses de afgelopen jaren een hoge vlucht hebben genomen, in het bijzonder op het terrein van misdaad- en fraudebestrijding.

3.2.1 CRIMINALITEITS ANTICIPATIE SYSTEEM (CAS): MISDAAD VOORSPELLEN

Het CAS is een tot de verbeelding sprekend systeem om misdaad te voorspellen, waarover bij direct betrokkenen hoge verwachtingen leven. Het CAS is de Nederlandse pendant van Big Data-toepassingen waarmee politieorganisaties in de Verenigde Staten en het Verenigd Koninkrijk al geruime tijd experimenteren. In beide landen wordt al sinds 2009 melding gemaakt van de ontwikkeling van *predictive policing software* (Beck en McCue 2009; Johnson et al. 2009; Perry et al. 2013). De meest gebruikte software, Predpol, werd ontwikkeld in samenwerking met de politiedepartementen van Los Angeles en Santa Cruz en wordt nu in verschillende Amerikaanse steden gebruikt (Perry et al. 2013; Predpol 2014). Sinds 2013 wordt met soortgelijke software ook in het Verenigd Koninkrijk geëxperimenteerd (Kent Online 2013). Het Nederlandse CAS is ontwikkeld door de politie van Amsterdam. De twee voornaamste doelstellingen van het CAS zijn criminaliteitspreventie en optimalisatie van politie-inzet (Willems en Doeleman 2014). Na een succesvol pilotproject is eind 2015 ook een nationaal pilotproject van start gegaan (Van Brakel 2016b). Het CAS zal geïmplementeerd worden in vier regionale eenheden: Noord-Nederland, Noord-Holland, Oost-Nederland en Den Haag.

Heat maps en hotspots

Het CAS deelt de stad Amsterdam op in vakjes van 125 bij 125 meter, waarvan een grote hoeveelheid gegevens wordt verzameld. Hierbij gaat het onder meer om:

– specifieke locatiegegevens van elk vakje, zoals de afstand tot de woonplaats van bekende verdachten, de afstand tot de dichtstbijzijnde snelwegoprit, soort en aantal bedrijven (bijvoorbeeld gokhallen, cafés), en daarnaast demografische en sociaaleconomische gegevens zoals huishoudomvang en gemiddeld buurtinkomen;
– historische data over woninginbraken en andere criminele incidenten;
– welke misdaden plaatsvonden binnen twee jaar voor het peilmoment (Willems en Doeleman 2014).

De keuze voor de gegevens is gebaseerd op criminologisch onderzoek en misdaadtheorieën (Van Brakel 2016b). De gegevensbestanden die voor het CAS worden geconsulteerd zijn onder meer afkomstig van de Basisvoorziening Handhaving van de politie, het CBS en Geo-informatie.

Het gegevensraster dat zo ontstaat stelt de politie in staat om zogenoemde *heat maps* te construeren. Op deze *heat maps* staan de plekken aangegeven waar de kans op bepaalde misdaden het grootst is. De identificatie van deze hotspots kan vervolgens behulpzaam zijn bij de besluitvorming over de inzet en briefing van politieagenten, zoals de Amsterdamse Flexteams (Willems en Doeleman 2014), het materieel dat zij gebruiken – bijvoorbeeld helikopters – en de tijd en plek van surveillance. Het CAS is momenteel gericht op *high impact crimes* en beperkt zich in de praktijk voornamelijk tot het voorkomen en bestrijden van inbraken. De resultaten zijn volgens de betrokkenen zeer positief (Rienks 2014). De claim is dat het CAS binnen de top-3 van de vakjes 40 procent van de woninginbraken en 60 procent van de straatroven helpt voorkomen (Willems en Doeleman 2014: 42). Deze auteurs merken overigens wel op dat dit succes niet uitsluitend aan het CAS valt toe te schrijven; de Eenheid Amsterdam heeft ook andere inspanningen geleverd om het aantal woningbraken terug te dringen. Ook internationaal blijken betrouwbare resultaten vooralsnog buitengewoon schaars te zijn (voor een overzicht zie Van Brakel 2016b).

Voorlopige balans

Predictive policing lijkt een kansrijk middel te zijn om woninginbraken beter te voorkomen. Bovendien is het idee dat dit instrument ook op andere soorten misdaden toepasbaar is, zoals nieuwe soorten inbraakcategorieën en georganiseerde misdaad. Hiernaast is er een tendens waarneembaar om de toepassing van *predictive policing* realtime te maken. Zo loopt er bij verschillende politiediensten in de Verenigde Staten een experiment met software die politieagenten informatie geeft over potentieel gevaarlijke situaties of individuen na een noodoproep via 911

(Intrado 2012; Skorup 2014). Hierbij worden behalve politiedata ook allerlei zachtere data gebruikt, onder meer afkomstig uit sociale media. Tegelijkertijd zijn er nog vele praktische bezwaren en technische problemen te overwinnen. Zo vereist het gebruik van *predictive policing* een stevige inbedding in de dagelijks politiepraktijk. Ook is de toegang tot relevante data die de politie in datawarehouses verzamelt belangrijk. Tegelijkertijd zijn er ook krachten die de andere kant opwerken. Bezuinigingen en de wens om meer kostenefficiënt te gaan werken zijn binnen het veiligheidsdomein vaak belangrijke drijfveren voor het gebruik van (Big) data-analyses (Galdon Clavell 2016). Specifiek voor de politie wijst Rienks (2014: 177) nog op twee andere factoren, namelijk het besluit tot de vorming van de Nationale Politie en het centraliseren van de datavoorziening. Door het Aanvalsprogramma Nationale Politie wordt de informatievoorziening van de politie momenteel grondig herzien. Oude en onderling verschillende regionale systemen maken plaats voor nieuwe nationale en gestandaardiseerde functionaliteiten waarmee data sneller beschikbaar zijn en meer gespecialiseerde analyses mogelijk worden (Politie 2013: 7). De verbeteringen aan de ICT-infrastructuur van de politie hebben echter de nodige vertraging opgelopen.[1]

3.2.2 BELASTINGDIENST: KOPLOPER EN SPIL IN SAMENWERKINGSVERBANDEN VOOR DATA-ANALYSES

De Belastingdienst bezit een ongeëvenaarde hoeveelheid data over personen in Nederland.[2] Deze data heeft de dienst nodig voor de uitvoering van bij wet opgedragen taken en verplichtingen. De dienst loopt voorop in de ontwikkeling en toepassing van data-analyses en vervult vanwege de grote hoeveelheid gegevens die hij bezit, en de structurering daarvan rond onder meer het burgerservicenummer (BSN), tevens een spilfunctie in veel samenwerkingsverbanden, waaronder de 'Infobox Crimineel en onverklaard vermogen' (zie 3.2.4) en de Landelijke Stuurgroep Interventieteams die van het 'Systeem Risico Indicatie' gebruikmaakt (zie 3.2.5). 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. Kostenverlaging en stijging van vorderingen zijn daarbij belangrijke doeleinden.

Data

De Belastingdienst heeft zelf veel gegevens in huis en kan daarnaast met vele instanties gegevens uitwisselen (zie onder meer de infographic in WRR 2011: 28). Zo kent de Belastingdienst 19 convenanten voor gegevensuitwisseling.[3] Ook mag hij bij diverse partijen direct gegevens opvragen. Voor opsporingsdoeleinden maakt de dienst daarnaast gebruik van openbare bronnen, zoals registers van de Kamer van Koophandel, en onderneemt de dienst speuracties op het internet. Belangrijk is ook dat de Belastingdienst beschikt over de uitkomsten van steekproeven. Deze vormen traditioneel de basis voor de opsporing van nalevings-

tekorten maar worden nu tevens gebruikt om statistische voorspelmodellen te ontwikkelen. Deze historische data zijn bij uitstek geschikt voor patroonherkenning. Hoewel de Belastingdienst een registratieplicht kent voor de gegevens die hij opvraagt en gebruikt, bleek de afgelopen jaren uit enkele rechtszaken dat dit niet in alle gevallen gebeurd was. Zo was niet bekend dat de Belastingdienst op grote schaal kentekenregistraties en parkeergegevens binnenhaalde en doorzocht (Olsthoorn 2016). Ook ontbreekt een duidelijke bewaartermijn voor gecombineerde bestanden en 'hits' op basis van risicoanalyses.[4]

Analysetechnieken

De Belastingdienst richtte in 2013 een aparte eenheid op om te onderzoeken welke nieuwe toepassingen de beschikbare data mogelijk maken. Deze eenheid (inmiddels opgegaan in de afdeling Data & Analytics) werkt daartoe met zelflerende modellen, die op basis van nieuwe praktijkresultaten steeds verder worden verfijnd. Er worden profielen geconstrueerd op basis van datamining van de gegevensbestanden van de Belastingdienst, met name eerder geïnspecteerde aangiften. "(D)aar gaan we volkomen blanco in", zo typeert Hans Blokpoel, algemeen directeur van de Belastingdienst, deze onderzoeksmethode in een interview met *De Correspondent*.[5] Deze 'research en development' gebeurt deels in samenwerking met buitenlandse belastingdiensten, vooral de Belgische en Britse, die binnen Europa als koplopers op het terrein van data-analyses gelden (Olsthoorn 2016).

Hiernaast heeft de Belastingdienst ook technieken ontwikkeld (het zogenoemde dynamisch monitoren) om automatisch verhaalsmogelijkheden te signaleren bij openstaande vorderingen waar beslag op gelegd mag worden. Ook het Landelijk Incassocentrum maakt intussen gebruik van deze 'innovatie'.[6] Hierdoor halveerde de tijd die nodig was voor het afhandelen van een loonbeslag, en verdubbelde de succesratio van beslagleggingen. De bestrijding van carrouselfraude vindt plaats aan de hand van netwerkanalyses, die aan de hand van 150 patronen de relaties van criminele netwerkorganisaties eerder en omvangrijker in beeld kunnen brengen. Behalve voor opsporing en controle gebruikt de Belastingdienst data-analyses tot slot ook voor preventief handelen. Daarbij gaat het behalve om het vooraf identificeren van risicovolle posten, waardoor bijvoorbeeld ten onrechte teruggevraagde inkomensbelasting niet hoeft te worden uitbetaald, ook om het voorkomen van onjuiste of onterechte aanvragen bij toeslagen. Dit leverde de Belastingdienst naar eigen zeggen respectievelijk 180 miljoen en 52 miljoen euro op.[7]

Balans

De handelwijze van de Belastingdienst is niet onomstreden. Er is door zowel individuele burgers als 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 *fishing expedition* beticht, maar mocht deze handelwijze na hoger beroep alsnog voortzetten.[8] Het probleem zit hem niet alleen in de ruime bevoegdheden van de Belastingdienst op het terrein van de gegevensverwerking. Ook de aard van de data, die steeds 'rijker' zijn waardoor er steeds meer uit valt af te leiden (bijvoorbeeld betaalgeschiedenissen in plaats van uitsluitend aankoopgegevens), en de sterk toegenomen mogelijkheden om databestanden te koppelen, roepen vragen op. De Belastingdienst wil het gebruik van grote databestanden overigens de komende jaren nog verder opvoeren en op termijn een datagestuurde dienst realiseren. Dit vergroot de kans dat de gegevensverwerking door de Belastingdienst ook in de toekomst regelmatig op de grenzen zal stuiten van wat maatschappelijk aanvaardbaar wordt geacht.

3.2.3 INFOBOX CRIMINEEL EN ONVERKLAARBAAR VERMOGEN (ICOV): OP DE DREMPEL VAN BIG DATA?

De iCOV is een samenwerkingsverband dat in 2013 werd opgericht door het Openbaar Ministerie, de Nationale Politie, de Belastingdienst (Belastingen, Toeslagen), de Douane, de FIOD en de Financial Intelligence Unit (Staatscourant 2013, 24607). Het aantal deelnemende organisaties is sinds 2013 uitgebreid met de opsporingsdienst van de Nederlandse Voedsel- en Warenautoriteit (NVWA), de Inspectie Sociale Zaken en Werkgelegenheid (ISZW) en de Inspectie Leefomgeving en Transport (ILT). Ook andere toezichthouders en inspecties hebben aangegeven tot de iCOV toe te willen treden (Olsthoorn 2016). Het bestaansrecht van de iCOV ligt in de maatschappelijke en financiële belangen die gemoeid zijn met criminele of onverklaarbare financiële of economische geldstromen in of via Nederland.

Doelen

De voornaamste doelstellingen van de iCOV zijn: gegevens verwerken en rapportages vervaardigen om crimineel en onverklaarbaar vermogen in kaart te brengen, witwas- of fraudeconstructies blootleggen, belastingontduiking tegengaan en de partners helpen om overheidsvorderingen alsnog te innen. Naast deze operationele doelen, gericht op toezicht, handhaving, opsporing en incasso, heeft de iCOV als meer structureel doel om de deelnemende partijen in staat te stellen 'betekenisvol te interveniëren' dan wel 'gedrag te beïnvloeden' (Staatscourant 2013, 24607). De gegevensverwerking heeft tot doel:

– onderzoek te doen met behulp van statistische bewerkingen en wetenschappelijke methoden, met als oogmerk 'indicatoren' en 'groepsprofielen' te ontwikkelen;

– 'criminaliteitsbeeldanalyses' te vervaardigen en gelijksoortige analyses met betrekking tot (criminele of onverklaarbare) vermogensbestanddelen uit te voeren;
– het ontwikkelen van 'methodieken' en 'rapportagevormen' ten behoeve van het in kaart brengen van vermogensbestanddelen of aan het licht brengen van witwas- of fraudeconstructies (Staatscourant 2013, 24607).

Werkwijze

De iCOV heeft de beschikking over een groot aantal relevante datasets van de deelnemende organisaties die in een eigen beveiligde omgeving zijn opgeslagen. De iCOV maakt zijn analyses louter op verzoek van deelnemende partijen en de wettelijke grondslag van de aanvrager bepaalt welke datasets in de analyse betrokken mogen worden. De iCOV produceert drie soorten rapportages (Convenant iCOV):

– De 'rapportage vermogen en inkomen', die aan te vragen is voor een subject, een object, of voor een grote hoeveelheid subjecten. In een '360 gradenanalyse' komen in deze rapportage inkomens, vermogens, transacties enz. naar boven.
– Het in kaart brengen van een relatienetwerk (zakelijke contacten, vastgoed enz.) rondom of tussen personen en/of organisaties. Hiermee kan met hoge snelheid de diepte en de breedte van een netwerk in kaart worden gebracht.
– Door bestandsvergelijking in het kader van een bepaald thema kan met behulp van risico-indicatoren casuïstiek in kaart worden gebracht. Dit betreft bijvoorbeeld onderzoek naar groepen, sectoren of een geografisch gebied. Aan de hand van scenario's en verschijningsvormen worden bijbehorende risico-indicatoren toegepast op de toegestane bronnen (open, gesloten en privaat), alleen voor zover de aanvrager die meebrengt. Hieruit volgt een beeld dat aanknopingspunten kan bieden voor nadere analyse en het bepalen van interventiestrategieën.

Op een aantal punten behaalt de iCOV goede resultaten met de data-analyse. Met name de snelheid van de analyse is een groot voordeel: zij levert resultaten die anders weken of zelfs maanden van traditioneel rechercheren zouden vergen (Olsthoorn 2016). De producten van de iCOV zijn meestal te beschouwen als 'sturingsinformatie' en gelden niet als bewijsvoering in een rechtszaak. Ze geven dus eerder 'richting' in de opmaat naar of in de vervolgstappen binnen een strafrechtelijk of fiscaal onderzoek dan dat zij een deel uitmaken van het (proces)dossier.

Ambities en onzekerheden

De iCOV en de organisaties die daarin samenwerken zien grote mogelijkheden om de aanpak van financiële fraude en criminaliteit te versterken via data-analyse. Daarover vindt een discussie plaats. In het bijzonder gaat het om de vraag in hoeverre

de huidige aanpak – analyse op basis van een vooraf bekende verdachte – uitgebreid kan en moet worden met een aanpak op basis van risicoprofielen en datamining. De indicatoren en profielen die de iCOV hier momenteel voor opstelt, vallen onder het juridisch regime dat ook geldt voor wetenschap en statistiek. Door deze aanpak zouden vervolgens opmerkelijke transacties, personen en organisaties uit de data kunnen worden gedestilleerd, voordat sprake is van daadwerkelijke verdenking. Deze stap zou de iCOV – in combinatie met de beschikbare grote datasets – veel meer in het hart van Big Data-processen brengen. Deze aanpak stuit echter ook op de nodige bezwaren. Om te beginnen zijn de doelstellingen van de samenwerkende organisaties niet eensluidend. Het combineren van gegevens stuit hierdoor op het principe van doelbinding. Datamining – het ongericht analyseren van een grote hoeveelheid data – staat daarnaast op gespannen voet met dataminimalisatie. Ook bestaat voor dergelijk 'verkennend' onderzoek geen juridische titel, hoewel er wel van bestanden gebruikt wordt gemaakt die niet zonder bijzondere toestemming, bijvoorbeeld van de Officier van Justitie, opgevraagd en ingezien mogen worden. Binnen de iCOV zou men voor deze nieuwe analysemethoden dan ook graag een heldere wettelijke grondslag zien (Olsthoorn 2016). Bovendien is men beducht voor negatieve beeldvorming in de buitenwereld en het risico om in een 'big brother'-achtig frame te belanden zoals gebeurde bij de maatschappelijke discussie over het Systeem Risico Indicatie.

3.2.4 SYSTEEM RISICO INDICATIE (SYRI): EEN ONTWIKKELD MAAR OMSTREDEN ANALYSESYSTEEM

Het Systeem Risico Indicatie (SyRI) is ontworpen om gegevens te verwerken voor het uitvoeren van risicoanalyses ten behoeve van "de voorkoming en bestrijding van onrechtmatig gebruik van overheidsgelden en overheidsvoorzieningen op het terrein van de sociale zekerheid en de inkomensafhankelijke regelingen, de voorkoming en bestrijding van belasting- en premiefraude en het niet naleven van de arbeidswetten" (Wet SUWI, art. 64, lid 1). Het systeem wordt gebruikt door een samenwerkingsverband van gemeenten, UWV, Sociale Verzekeringsbank, Inspectie SZW en Belastingdienst. Deze samenwerking vindt plaats in het kader van de (in 2014 gewijzigde) Wet Structuur Uitvoeringsorganisaties Werk en Inkomen (Wet SUWI) en is verder uitgewerkt in het Besluit SyRI, dat op 12 september 2014 in werking trad.[9] Gebruikte gegevens zijn:

- bestuursrechtelijke boetes en sancties;
- fiscale verplichtingen;
- onroerende goederen;
- uitsluitingsgronden van bijstand en uitkeringen;
- handelsgegevens;
- huisvestingsgegevens;
- persoons- en of bedrijfsgegevens;
- inburgeringsgegevens;

– nalevingsgegevens;
– onderwijsgegevens;
– pensioengegevens;
– re-integratiegegevens;
– schuldenlastgegevens;
– uitkerings-, toeslagen- en subsidiegegevens, vergunningen en ontheffingen;
– zorgverzekeringsgegevens (Besluit SyRI, artikel 5a.1).

Werkwijze

Gemeenten, UWV, Sociale Verzekeringsbank, Inspectie SZW en Belastingdienst werken al geruime tijd samen op het terrein van fraudebestrijding in de Landelijke Stuurgroep Interventieteams (LSI), waarin ook het Openbaar Ministerie, de Vereniging Divosa, de politie en de ministeries van Financiën en van SZW zitting hebben. Binnen dit samenwerkingsverband is ook de analysetechniek ontwikkeld waarvan het SyRI gebruikmaakt (Olsthoorn 2016). Deze techniek, de Black Box, kwam in 2006 tot stand na kritiek van het College Bescherming Persoonsgegevens (nu Autoriteit Persoonsgegevens). Het CBP bekritiseerde de rechtmatigheid van de koppeling van gehele bestanden, zoals de koppeling van gegevens over waterverbruik, gemeten door nutsbedrijven, aan adressen van uitkeringsgerechtigden in het project Waterproef.[10] Daarom werd de Black Box ontwikkeld om data anoniem te kunnen koppelen en analyseren.

Het SyRI werkt als volgt.[11] Twee of meer gemeentebesturen, bestuursorganen of toezichthouders op het terrein van sociale zekerheid of belastingen sluiten een samenwerkingsverband en verzoeken het ministerie van SZW om een risicoanalyse uit te voeren. Het zogenoemde 'Inlichtingenbureau'[12] koppelt en versleutelt de door deze organen aangeleverde gegevensbestanden, voert ze in het SyRI in, en ontsleutelt vervolgens de resultaten die op een verhoogd risico duiden. Deze potentiële 'hits' worden aan het ministerie (in casu de afdeling Onderzoek & Analyse van de Inspectie SZW) teruggeleverd. Die afdeling analyseert ze, bepaalt welke gegevens voor een risicomelding in aanmerking komen en verstrekt ze daarna aan de vragende organisaties (of politie of OM). Deze organisaties zijn verplicht nader te onderzoeken of het echt om fraude gaat alvorens een sanctie te treffen. Risicomeldingen worden opgenomen in een register, dat burgers op aanvraag kunnen inzien. Zij worden dus niet automatisch op de hoogte gesteld van het onderzoek. Het Inlichtingenbureau vernietigt direct alle resultaten die niet op een verhoogd risico duiden, het ministerie van SZW bewaart risicomeldingen maximaal twee jaar.

Opbrengst

Tussen 2004 en begin 2015 zijn er door de LSI 160 fraudeopsporingsprojecten goedgekeurd, waarvan er 22 van de Black Box gebruik hebben gemaakt (Ministerie van Sociale Zaken en Werkgelegenheid 2015a). Het ging daarbij om wijk- en

gebiedsgerichte projecten. Het ministerie van SZW claimt dat deze projecten een opbrengst hebben van zo'n 21 miljoen euro, door opgelegde belastingaanslagen en boetes, teruggevorderde en beëindigde uitkeringen, correcties op toeslagen, en boetes wegens overtreding van de Wet arbeid vreemdelingen en/of de Wet minimumloon- en minimumvakantietoeslag (Ministerie van Sociale Zaken en Werkgelegenheid 2015a; 2015b). Mede vanwege deze 'successen' fungeert het SyRI als inspiratiebron voor verdere stappen op het terrein van gegevensuitwisseling en -analyse, op een breder terrein dan alleen fraudebestrijding (Werkgroep Verkenning kaderwet gegevensuitwisseling 2014).

Een omstreden systeem

Het Besluit SyRI ontmoette echter ook veel kritiek, zowel tijdens de totstandkoming ervan als daarna. Volgens de Raad van State (2014, paragraaf 2a) was er "nauwelijks een persoonsgegeven te bedenken dat niet voor verwerking in aanmerking komt". Het koppelen van bestanden zou volgens de Raad bovendien tot "een 'fishing expedition' en zelfs tot willekeur" kunnen leiden. Het College Bescherming Persoonsgegevens (CBP 2014: 3) benadrukte om dezelfde reden dat risico-indicatoren een goede onderbouwing vereisen: ze moeten 'objectiveerbaar' zijn. Het College wees er daarnaast op dat profielen vooral op negatieve persoonskenmerken worden gebaseerd (bijvoorbeeld schulden of overtredingen), wat de privacybelangen van bepaalde groepen buitenproportioneel zou kunnen schaden. Het ministerie van SZW nam beide adviezen slechts gedeeltelijk over. Eind 2014 kwam het SyRI bovendien onder vuur te liggen na een artikel in *de Volkskrant*, met als steen des aanstoots de grootschalige koppeling van data en het 'profilen' van onschuldige burgers.[13] In de berichtgeving doemde het beeld op van een almachtige staat die door middel van ondoorzichtige en deels geautomatiseerde processen burgers nauwlettend in de gaten houdt, ongeacht of zij iets op hun kerfstok hebben. Het ministerie van SZW kwam met een uitgebreide repliek, maar slaagde er niet in om de kritiek op het systeem geheel te ontzenuwen.[14]

3.2.5 DE INLICHTINGENDIENSTEN: ONDERWEG NAAR DATAMINING?

De aandacht voor de inlichtingendiensten is nooit geheel weggeweest, maar sinds de onthullingen van Edward Snowden staan ze meer dan ooit in de schijnwerpers. Het is vooral de schaal van de dataverzameling en -analyse waarnaar de aandacht uitgaat. Nederland behoort met Duitsland, Zweden, het Verenigd Koninkrijk en Frankrijk tot de weinige landen binnen de Europese Unie met gedetailleerde wetgeving voor de verzameling van zowel gerichte als ongerichte gegevens (FRA 2015). De Wet op de inlichtingen- en veiligheidsdiensten 2002 (Wiv 2002) die het werk van de Algemene Inlichtingen- en Veiligheidsdienst (AIVD) en de Militaire Inlichtingen- en Veiligheidsdienst (MIVD) inkadert, maakt onderscheid tussen kabelgebonden communicatie en niet-kabelgebonden communicatie. Voor kabelgebonden communicatie (veelal internetverkeer) mag alleen gericht inlichtingenonderzoek worden gebruikt, onderzoek dus met een vooraf vastgesteld doel. Bij

niet-kabelgebonden communicatie, zoals communicatie die via satellieten of radiogolven verloopt, is dat anders. Die mag behalve gericht ook ongericht worden doorzocht. In het nieuwe conceptvoorstel voor de Wiv dat de regering medio 2015 presenteerde, verdwijnt dit onderscheid en kunnen de diensten ook ongericht het kabelgebonden internetverkeer gaan onderscheppen. De Ontwerp toelichting van het wetsvoorstel spreekt van het binnenhalen van 'bulk-communicatie', die vervolgens is te verrijken, correleren en combineren met gegevens die met andere middelen zijn verzameld.[15] Op deze wijze kunnen de diensten in feite de soort gegevensbestanden binnenhalen die ook bij Big Data-toepassingen gangbaar zijn.

Data

De AIVD en MIVD kunnen gegevensverzamelingen aanleggen door reeds verwerkte gegevens te combineren of door gegevens te verzamelen uit open bronnen, die vrijwillig zijn afgestaan door derde partijen. Die externe gegevensverzamelingen zijn zelf vaak al geautomatiseerd, waardoor ze relatief makkelijk te analyseren zijn en te koppelen zijn met andere gegevens. Gegevensverzamelingen kunnen hiernaast tot stand komen door de inzet van een bijzondere bevoegdheid (bijvoorbeeld taps, interceptie, hacken of vorderen van gegevens van derden) of door samenwerking met buitenlandse inlichtingendiensten. Er zijn geen cijfers beschikbaar over de frequentie en schaal waarmee de AIVD en de MIVD bijzondere bevoegdheden inzetten, met als argument dat deze gegevens onder het staatsgeheim vallen. In de ons omringende landen wordt hier overigens wel over gerapporteerd (CTIVD 2015a). Een voorbeeld is België, waar de toezichthouder vermeldt hoe vaak toestemming is verleend om bepaalde bevoegdheden in te zetten en voor welke doeleinden dat gebeurde. Wel rapporteert de MIVD (2015: 74) over 2014 grootschaliger en meer complexe digitale onderzoeken, onder meer om inlichtingen over terrorisme en extremisme te verkrijgen. De AIVD (2015: 16-17) richt zich in toenemende mate op sociale media en webfora. De dienst heeft hierin de afgelopen jaren veel geïnvesteerd. Onderzoek op internet is een vast en, naar eigen zeggen, effectief onderdeel geworden van het instrumentarium van de AIVD, en het is aannemelijk dat ook hier veel gegevensverzameling plaatsvindt. Een aantal van deze internetoperaties, waarbij heimelijk data van een aantal grotere algemene webfora werd binnengehaald, was volgens de toezichthouder onrechtmatig, mede omdat een disproportionele hoeveelheid data werd vergaard (CTIVD 2015a: 14).

Samenwerking en gegevensuitwisseling

De AIVD en de MIVD zetten bovendien onderling en met andere partijen veel samenwerkingsverbanden op, waarbinnen gegevensuitwisseling plaatsvindt (Van Schendel 2016). Bijvoorbeeld de Joint Sigint Cyber Unit, een gezamenlijk platform van AIVD en MIVD, dat zorg draagt voor de verzameling van metadata uit het radio- en etherverkeer. Een ander voorbeeld is de Contra Terrorisme Infobox, met als deelnemers naast AIVD en MIVD ook de Landelijke Eenheid van de Nationale Politie, de Immigratie- en Naturalisatiedienst, de Fiscale inlichtingen- en

opsporingsdienst, de Financial Intelligence Unit, het ministerie van SZW, de Koninklijke Marechaussee en het Openbaar Ministerie. Bij de laatstgenoemde samenwerking worden gegevens over personen en netwerken die zijn betrokken bij terrorisme, bij elkaar gebracht en daarna geanalyseerd. Er is ook samenwerking met buitenlandse diensten, waarbij metagegevens van ongerichte interceptie worden gedeeld (CTIVD 2015b). Hoewel uit metadata steeds preciezere gevolgtrekkingen zijn af te leiden over het leven van degenen van wie deze data zijn afgeleid, wordt deze gegevensuitwisseling als een vorm van 'technische ondersteuning' gezien, en is de toestemming van de minister daarvoor (vooralsnog) geen wettelijke vereiste.[16]

(Big) Data-analyse

De Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten (CTIVD) concludeert dat beide inlichtingendiensten gebruikmaken van "grotere hoeveelheden data en zij verwerken deze steeds vaker ook geautomatiseerd" (CTIVD 2015a: 11). Deze ontwikkeling is al geruime tijd gaande. Na de aanslagen in New York, Londen en Madrid presenteerde de toenmalige Nederlandse regering een voorstel om data-analyses beter in de Wiv 2002 te verankeren. Volgens dit wetsvoorstel, dat uiteindelijk geen doorgang vond, hanteerden de diensten vormen van data-analyse waarbij zij profielen gebruikten om geautomatiseerde gegevensverzamelingen te doorzoeken. Ook analyseerden zij deze verzamelingen met als doel daar patronen in te ontdekken. Big Data-achtige analysetechnieken behoorden dus al ruim tien jaar geleden tot de gereedschapskist van de inlichtingendiensten. Intussen is de technologie voor data-interceptie steeds geavanceerder geworden en is ook de hoeveelheid gegevens die de inlichtingendiensten kunnen gebruiken en/of onderscheppen exponentieel gegroeid doordat een steeds groter aandeel van onze communicatie via digitale kanalen en/of apparaten verloopt. De CTIVD wil mede vanwege deze ontwikkeling zelf meer investeren in kennis van verwerkingssystemen en de automatische gegevensverwerking pareren door 'geautomatiseerd toezicht' te ontwikkelen, bijvoorbeeld door de toegang tot bepaalde gegevens te limiteren en mechanismen voor automatische gegevensverwijdering aan te brengen (CTIVD 2015a: 11).

Zorgen

Al in 2005 uitten diverse partijen hun zorgen over de impact van data-analyses van grootschalige gegevensverzamelingen op de vrijheden van burgers. Het toenmalige wetsvoorstel bevatte een artikel dat het mogelijk maakte persoonsgegevens te verwerken van personen die niet eerder in beeld van de inlichtingendiensten waren, mits dat nodig was voor profilering en/of datamining. Het wetsvoorstel sneuvelde na kritiek in de Eerste Kamer in 2008, mede naar aanleiding van diverse bezwaren van het toenmalige College Bescherming Persoonsgegevens (EK 2008-2009, 30 553, E). Intussen zijn dergelijke analysemethoden staande praktijk en constateert de CTIVD (2015a: 59) dat de gegevensverzamelingen van de

diensten ook gegevens bevatten van personen die vanuit de taakstelling van de diensten geen object van onderzoek zijn. Ook vereisen deze bestanden vormen van geautomatiseerde data-analyse waarvoor de huidige Wiv 2002 geen juridische grondslag biedt. Het nieuwe conceptvoorstel voor de Wiv biedt hier vooralsnog geen oplossing voor. De commissie die de Wiv 2002 evalueerde, de commissie Dessens, maakte een sterke koppeling tussen de verruiming van de bevoegdheden van de diensten en een versterking van de wettelijke privacywaarborgen en verzwaring van het onafhankelijke toezicht. De regering legde deze laatste aanbeveling echter grotendeels naast zich neer. In de internetconsultatie over het wetsvoorstel hebben vele partijen het belang van extra waarborgen en controle onderstreept[17] (zie ook Eskens et al. 2015).

3.2.6 BIG DATA, OPENBARE ORDE EN SMART CITIES

Big Data wordt ook steeds vaker voor openbareordevraagstukken gebruikt. Dat gebeurt meestal op het niveau van grote steden, stadswijken, publieke ruimtes of grootschalige publieke of semi-publieke evenementen. De data zijn daarbij voornamelijk afkomstig van databases van de (lokale) overheid, mobiele telefoons van burgers en 'slimme' apparaten zoals verkeers- en beveiligingscamera's, parkeermeters en lantaarnpalen voorzien van intelligente sensoren. De term 'smart city' belichaamt de samenkomst van Big Data en deze 'slimme' objecten en/of apparaten en het gebruik van digitale technologie om het functioneren van steden te verbeteren (González Fuster en Scherrer 2015: 14). Vanuit het perspectief van Big Data is het vooral deze 'everyware' (Greenfield 2006 in Kitchin 2014a) van apparaten, netwerken en digitaal gecontroleerde infrastructuren waarop de moderne steden zijn gebaseerd, die een enorme potentie heeft voor datamining.

Data

Een van de meest bekende en vroege voorbeelden van realtime-data-analyses is het gebruik van camerabeelden om verkeersstromen te monitoren en reguleren (Kitchin 2014a; Yin et al. 2015). De simpele reden hiervoor is dat verkeerssystemen belangrijke stedelijke systemen zijn, en dat openbaarvervoerskaarten, GPS-systemen, verkeerscamera's en andere sensoren een enorme hoeveelheid data produceren. Smart-citytechnologieën zijn daarom vaak het eerst toegepast in dit domein. Een tweede belangrijke bron van gegevens zijn de beelden van het cameratoezicht. Enkele jaren geleden werd elke persoon die door het centrum van Londen – de internationale hoofdstad van het cameratoezicht of CCTV[18] – wandelde per dag al honderden keren door camera's geregistreerd (Craig en Ludoff 2011). Ook data van mobiele telefoons – locatiegegevens in het bijzonder – zijn al in een vroeg stadium (rond 2009) gebruikt om de temporele en geografische activiteiten van het publiek te analyseren (Yin et al. 2015). Intussen is de aandacht verruimd naar de analyse van sociale media, zoals Twitter, Facebook en Instagram.

Toepassingen

In Nederland is het gebruik van Big Data door gemeenten over het algemeen nog beperkt, de vele enthousiaste verhalen over slimme steden ten spijt (Daalhuijsen et al. 2015). De gemeenten die hiermee bezig zijn, zoals Almere, Assen, Eindhoven, Tilburg en Utrecht, richten zich hoofdzakelijk op veiligheidsanalyses en verkeersmanagement. Amsterdam experimenteert – in samenwerking met universiteiten en private dataorganisaties – met het realtime volgen en voorspellen van omvangrijke bezoekersstromen, bijvoorbeeld rond een evenement als Sail Amsterdam 2015.[19] In de internationale zone van Den Haag heeft een pilot plaatsgevonden met een systeem dat data kan combineren en analyseren van bijvoorbeeld meldpanelen, alarmsystemen, hekwerkdetectoren en beveiligingscamera's, maar ook van openbare bronnen zoals verkeersinformatiediensten, weercomputers en internetdata.[20] Kool et al. (2015: 18-19) beschrijven hoe de gemeente Eindhoven Big Data gebruikt om de veiligheid in het uitgaansgebied Stratumseind te verbeteren. Dat gebeurt door realtime-data-analyses van grote aantallen slimme camera's te combineren met gegevens van sociale media. De uitkomsten van deze informatie gaan vervolgens naar bestuur en politie, die bepalen of specifieke acties nodig zijn. Het systeem is volgens de auteurs niet foutloos; ondanks betere software valt niet uit te sluiten dat het systeem onschuldige burgers als verdachten aanmerkt bij afwijkend gedrag.

Potentie en vragen

Realtime-data-analyses kunnen bijdragen aan een verhoogd omgevingsbewustzijn, doordat ze een snellere identificatie van ontwikkelingen en stemmingen mogelijk maken. Dit blijkt bijvoorbeeld uit de analyse van Twitterberichten (Omand et al. 2012: 25; vgl. Vis et al. 2013). Nadat de media over bepaalde gebeurtenissen berichten, intensiveert vaak het Twitterverkeer. Voorafgaand aan deze gebeurtenissen zijn echter vaak al korte oprispingen van tweets waarneembaar, die, mits goed geanalyseerd, op de conventionele berichtgeving vooruitlopen. Gecombineerd met geo-locatiegegevens kan dit een snellere en meer effectieve interventie mogelijk maken. Ook kunnen socialemediagegevens gedetailleerd inzicht geven in wanneer bepaalde groepen demonstraties en opstootjes plannen of hoe en wanneer rivaliserende voetbalfans een 'treffen' afspreken (Omand et al. 2012: 25). Snel handelen kan dan potentieel grote economische schade en menselijk leed voorkomen. De politie kan op basis van dit soort gegevens in snel veranderende situaties het soort inzet bepalen om de openbare orde te handhaven.

De inzet van dit soort middelen roept echter ook vele vragen op (Kitchin 2014a; Edwards 2016). Allereerst zijn er spanningen met de privacy van burgers wanneer zij op grote schaal gevolgd en gemonitord worden, zeker wanneer daarbij ook socialemediaberichten worden geanalyseerd (Puschmann en Burgess 2013). Ook is waarneembaar dat hiermee buurten, reisarrangementen en grote sport- en vrijetijdsevenementen langzaamaan het veiligheidsdomein worden binnengetrokken

(Lyon 2014). Daarbij dreigen overheden bovendien afhankelijk te worden van de (grote) bedrijven die hiervoor de technologieën leveren, en met wie zij publiek-private samenwerkingen aangaan. Een ander groot probleem is tot slot dat nieuwe vormen van onlinegedrag, normen en taalgebruik een betrouwbare analyse ingewikkeld maken. In tegenstelling tot de relatief 'harde' financiële gegevens die voor fraudebestrijding worden gebruikt, vragen de 'zachte' data afkomstig van sociale media veel duiding en interpretatie, met als groot risico dat verkeerde conclusies worden getrokken. De vertaling van grote, complexe en elkaar regelmatig tegensprekende databestanden in *actionable knowledge* is een uitdaging die vooralsnog overwonnen dient te worden (Omand et al. 2012).

3.2.7 INTERNATIONALE MOBILITEIT, SMART BORDERS EN BIG DATA

Mobiliteit gaat van oudsher met identificatie gepaard, zeker wanneer landsgrenzen worden overschreden. De identificatie van personen is daarbij steeds verfijnder geworden, en beperkt zich al lang niet meer tot controle aan de landsgrenzen (Dijstelbloem en Meijer 2009). In Nederland en Europa wordt gewerkt met grote datasets om verschillende aspecten van het immigratie-, terugkeer- en mobiliteitsbeleid in goede banen te leiden. Ook aan de externe grenzen van de EU zijn nieuwe manieren van informatieverwerking, koppeling en analyse sterk in opkomst. Hoewel het hierbij om grote dataverzamelingen gaat, is de variëteit van deze verzamelingen beperkt, waardoor in strikte zin van Big Data nog geen sprake is. Wel voldoen in sommige gevallen de methoden van analyse aan het 'profiel' van Big Data. De Nederlandse gebruikers van deze systemen zijn voornamelijk de IND, de Koninklijke Marechaussee en diensten als de Dienst Terugkeer en Vertrek.

Databanken en hun doelen

Grofweg zijn er vier doelen die met de verschillende databanken en systemen nagestreefd worden:

- de uitvoering van het (Europese) asiel- en migratiebeleid (inclusief terugkeerbeleid);
- het verbeteren van de informatiepositie van de grensbewaking;
- het 'slimmer' afhandelen van de enorme reizigersstromen die naar Europa komen, met name via de vliegvelden;
- gaandeweg is ook veiligheid, meer in het bijzonder het opsporen en tegenhouden van terroristen in de reizigersstromen, steeds belangrijker geworden.

Deze doelen worden behalve met behulp van nationale (biometrische) databanken ook nagestreefd met Europese biometrische databanken als het Schengen Informatie Systeem II (SIS II), Eurodac, het Visum Informatie Systeem (VIS), intelligence systemen als het Eurosurprogramma en zogenoemde 'smart border'-initiatieven.

Databanken als het VIS bevatten grote en rijke datasets over iedereen die een toeristenvisum aanvraagt om naar de EU af te reizen. Het VIS is gebouwd om de gegevens van 70 miljoen personen met biometrische gegevens en alle informatie die een visumaanvraag vereist digitaal te registreren. Eurodac registreert alle asielaanvragen in de EU-lidstaten – inclusief biometrische data – met als doel om dubbele of meerdere asielaanvragen in beeld krijgen en aan te wijzen welk land verantwoordelijk is voor een asielaanvraag. Beide systemen worden ook gebruikt ten behoeve van het terugkeerbeleid van personen die als illegaal zijn aangehouden, door anonieme ongedocumenteerde personen via vingerafdrukken te koppelen aan een asiel- of visumaanvraag die in het systeem is geregistreerd (Broeders 2007). Het SIS II fungeert als een verzameling van digitale zwarte lijsten van onder meer criminelen, gestolen paspoorten, en van personen die de toegang tot de EU geweigerd moeten worden (vaak mensen die de immigratiewet hebben overtreden door bijvoorbeeld hun visum te laten verlopen).

Smart-borderprogramma's zijn erop gericht de reizigersstromen zowel voor aankomst als op het vliegveld zelf in verschillende banen te leiden. Via Advanced Passenger Information (API) en Passenger Name Records (PNR) die van de vervoersmaatschappijen betrokken worden, en andere databronnen – zoals de zwarte lijsten in het SIS II – worden sommige reizigers voor vertrek al geweigerd. Risicoprofielen bepalen voorts welke reizigers aan een nader onderzoek onderworpen worden en welke juist een snellere doorgang zouden moeten krijgen (Broeders en Hampshire 2013). Aan de EU-buitengrenzen ten slotte, wordt onder verantwoordelijkheid van het EU-agentschap Frontex gewerkt aan het Eurosur-systeem. Dit programma probeert verschillende nationale grenssurveillancesystemen aan elkaar te koppelen en met vele andere databronnen te verrijken. Het doel is om de *situational awareness* van de grensbewakers te versterken en ze in hun dagelijkse praktijk een accuraat *situational picture* mee te geven (Broeders en Dijstelbloem 2016). Onder dit programma is er een duidelijke ambitie om meer – ook private – bronnen te combineren en te integreren.

Ambities en resultaten

Systemen als Eurodac functioneren goed in het licht van hun oorspronkelijke doelstelling. Eurodac brengt bijvoorbeeld effectief in kaart of mensen in meerdere lidstaten asiel aanvragen en welk eerste land uiteindelijk verantwoordelijk is voor de afhandeling van die aanvraag. Op andere punten is minder bekend over de effectiviteit van de verschillende systemen. Analyse van de data in deze systemen draagt bij aan het opstellen van risicoprofielen die gebruikt worden om de reizigersstromen naar Nederland en Europa onder te verdelen in groene (doorlaten), zwarte (tegenhouden of weigeren) en vooral grijze categorieën (nader bestuderen en/of volgen) (Broeders en Hampshire 2013). Bellanova en De Hert (2011) stellen dat dergelijke data het grofmazige signaleren op basis van nationaliteit hebben vervangen door het fijnmaziger signaleren op basis van individuele kenmerken. Hoe effectief

dat is, is echter niet na te gaan. Op het terrein van de internationale mobiliteit, waar het om zeer grote aantallen reizigers gaat – in 2014 reisden 55 miljoen personen via Schiphol – is de verleiding groot om reizigers door middel van automatische besluitvorming in categorieën onder te verdelen. Zo constateerde de Nationale Ombudsman (2010) dat de Koninklijke Marechaussee en de IND een relatief 'automatische procedure' hanteren om mensen in het SIS te signaleren. Met de toename van de grote getallen, zowel in de reizigersstromen als in de systemen die worden gebruikt, ligt het risico van semi-automatische besluitvorming op de loer. Hierbij heeft een mens formeel het laatste woord maar feitelijk nauwelijks ruimte om van het systeem af te wijken. Recentelijk gaven de Amerikaanse autoriteiten zelfs voor een rechtbank toe dat personen op *no-fly lists* geplaatst worden op basis van aannames en profielen, gestoeld op voorspellingen van wat ze in de toekomst wellicht gaan doen en niet op basis van wat ze in het verleden gedaan hebben.[21] Binnen het domein van de internationale mobiliteit heeft dus ook het gebruik van voorspellende analyses zijn intrede gedaan.

3.3 VOORUITBLIK: BIG DATA EN VEILIGHEID OVER VIJF JAAR

Veel van de in hoofdstuk 2 genoemde kenmerken van Big Data zijn in de bovenstaande cases nog maar beperkt waarneembaar. De variëteit van de data is niet erg groot, misschien afgezien van de data die de Belastingdienst en de inlichtingendiensten analyseren. De gebruikte data zijn overwegend afkomstig uit overheidsbronnen, maar de betrokken organisaties combineren ze steeds vaker ook met data uit openbare bronnen als de registers van het Kadaster en de Kamer van Koophandel, het internet en sociale media. De data zijn bovendien naar hun aard (het betreft data uit overheidsbronnen) sterk gestructureerd en vaak gekoppeld aan het BSN.

Ook zijn nog relatief weinig data-gedreven analyses waarneembaar. Bij de meeste analyses gaat het om tamelijk eenvoudige koppelingen van data en is er sprake van een gering aantal bewerkingen en stappen in het analyseproces. Op kleine schaal wordt desondanks met meer data-gedreven analysetechnieken geëxperimenteerd, met als doel om interessante patronen te ontdekken en op basis daarvan risicoprofielen te ontwikkelen en/of verder te verfijnen. Dit gebeurt bijvoorbeeld bij de Belastingdienst en de inlichtingendiensten. Hoewel daarover geen harde gegevens beschikbaar zijn, mag worden verondersteld dat ook andere organisaties dit pad inslaan of hiertoe plannen hebben. Zo vindt binnen de ICOV een discussie plaats over het gebruik van datamining, om opmerkelijke transacties, personen en organisaties uit de data te kunnen destilleren voordat sprake is van daadwerkelijke verdenking. Opvallend is tot slot ook het gebruik van realtime-analyses, zoals bij het volgen van grote groepen mensen, wanbetalers en internetgebruikers waarbij vaak ook private data en metadata in de analyses betrokken worden.

Automatische besluitvorming lijkt vooralsnog nauwelijks plaats te vinden, afgezien van het Amerikaanse voorbeeld waarbij computers personen op een *no-fly list* plaatsen zonder dat daarvoor een menselijk oordeel vereist is. Aan de andere kant, en zoals hetzelfde voorbeeld ook laat zien: veel besluitvorming is door geheimhouding aan het zicht onttrokken. Een tendens naar (gedeeltelijke) automatische besluitvorming, of semi-automatische besluitvorming, is bovendien onmiskenbaar waarneembaar, al is het maar doordat de grootte van de datasets het vrijwel onmogelijk maakt om per uitkomst elke stap in de analyse aan een zorgvuldige controle te onderwerpen. Immers, hoe groter de omvang en variëteit van de data, des te aantrekkelijker wordt het om (ten minste een deel van) het analyseproces te automatiseren.

Overheden in binnen- en buitenland investeren fors in onderzoek en ontwikkeling van Big Data-toepassingen, zo blijkt uit een internationale inventarisatie van overheidsuitgaven (Van der Sloot en Van Schendel 2016). Wanneer we deze ontwikkeling doortrekken naar de nabije toekomst, zeg vijf jaar verder, hoe ziet het gebruik van Big Data er dan uit? Wat kunnen we op dat vlak verwachten? Welke tendensen zijn waarneembaar en welke vragen roepen zij op?

3.3.1 DATA

De koppeling van gegevens neemt nu al een hoge vlucht. Het zijn immers niet de gegevens zelf die waardevol zijn, maar de koppeling daarvan, en dan het liefst de koppeling van grote hoeveelheden gegevens. Organisaties als SyRI en iCOV kunnen reeds grote hoeveelheden persoonsgegevens koppelen. De bevoegdheid van de inlichtingendiensten en de politie is nog vele malen ruimer, omdat zij breder geformuleerde doelstellingen hebben en niet gebonden zijn aan de Wet bescherming persoonsgegevens (zie hoofdstuk 5). De komende jaren zal deze koppeling naar verwachting sterk toenemen, door de enorme toename van data over menselijk gedrag en doordat overheden meer digitaal registreren en verzamelen. Ook worden gegevens steeds 'rijker', waardoor met dezelfde bevoegdheid een steeds groter palet aan informatie kan worden binnengehaald. Betalingsgegevens worden betalingsgeschiedenissen, naarmate de digitale sporen van betalingen gemakkelijker te bewaren zijn. Kort gezegd: we laten meer en diepere digitale sporen na. En daarmee groeit vanzelfsprekend ook het aantal analyse- en controlemogelijkheden.

Maar ook de herkomst van de gegevens zal veranderen. De binnen het veiligheidsdomein gebruikte data zijn overwegend afkomstig uit overheidsbronnen, maar worden in toenemende mate gecombineerd met data uit vrij opvraagbare openbare bronnen en data verkregen via het internet en sociale media. De rol die private data in de analyses spelen, zal in de nabije toekomst steeds groter worden. Temeer omdat private partijen steeds vaker de beschikking hebben over grote geautomatiseerde gegevensbestanden, die opvraagbaar zijn door organisaties binnen het

veiligheidsdomein. Met de opkomst van het internet, mobiele telefonie en binnenkort ook het Internet of Things ontstaat bovendien een nieuw soort data, die behulpzaam kan zijn voor toezicht en opsporing. Zo experimenteren diverse organisaties binnen het veiligheidsdomein met het monitoren van het klik- en surfgedrag van burgers op hun sites, bijvoorbeeld om na te gaan hoe zij hun belastingformulier invullen of om te controleren of zij voldoende investeren in het zoeken naar werk om een uitkering te krijgen (Olsthoorn 2016).

Behalve de herkomst verandert ook de aard van de data, die nu nog tamelijk hard is. Vaak gaat het om financiële gegevens en allerhande registraties, die al dan niet elektronisch worden gegenereerd. In de toekomst kunnen ook zachtere, meer sociale gegevens een rol in de analyses gaan spelen. Beide ontwikkelingen – het gebruik van private en zachtere data – worden ondersteund door de opkomst van software die steeds beter in staat is om data te analyseren die op verschillende wijze zijn opgeslagen en/of gestructureerd. De grenzen tussen gestructureerde en ongestructureerde en publieke en private data zullen daarmee vervagen. Ook het onderscheid tussen persoonsgegevens en niet-persoonsgegevens – zoals metadata – wordt in een Big Data-wereld steeds meer betekenisloos, als op basis van verschillende datapunten relatief eenvoudig een persoon te 'construeren' is. Beroemd is het onderzoek 'Unique in the crowd', waarin slechts vier punten in tijd en plaats genoeg waren om 95 procent van het sample van 'anonieme data' van mobiel telefoongebruik tot unieke personen te herleiden (De Montjoye et al. 2013).

Als gevolg van deze ontwikkelingen zal ook de organisatie van de gegevensverzameling en -uitwisseling sterk van karakter veranderen. Een toenemend aantal organisaties zal geneigd zijn zich te willen aansluiten bij de bestaande samenwerkingsverbanden, die nu gegevens uitwisselen en daarop analyses laten uitvoeren, zoals het Inlichtingenbureau. Of ze nemen zelf een nieuw initiatief, wat de overzichtelijkheid en kwaliteit van de gegevensuitwisseling waarschijnlijk niet ten goede zal komen (Kool et al. 2015: 26-27; Galdon Clavell 2016). Hoe dan ook zal de overheid meer afhankelijk van private partijen worden. Het rapport van de Werkgroep Verkenning kaderwet gegevensuitwisseling (2014) suggereert al om de gegevensuitwisseling met private partijen te intensiveren, bijvoorbeeld met banken in het kader van fraudebestrijding. Schneier (2015: 51-53) beschrijft voor de Verenigde Staten verschillende vormen van private gegevensverzameling, die vervolgens door kredietinstellingen en marketingbedrijven zijn gecombineerd tot databanken van een ongeëvenaarde reikwijdte en diepgang, die te koop en/of te huur zijn (zie ook Kitchin 2014a). Een aantal van deze 'gegevensmakelaars' is ook in Nederland actief, maar de aard en omvang van de gegevens die zij kunnen en mogen verzamelen is vooralsnog onvergelijkbaar met de Amerikaanse situatie. Naarmate overheidsorganisaties meer gebruik willen maken van private data zullen dergelijke gegevensmakelaars meer gewicht in de schaal gaan leggen bij het ver-

garen en voorbewerken van data. Voor dergelijke publiek-private samenwerkingen op het terrein van persoonsgegevensvergaring ontbreekt echter de benodigde wettelijke basis (Koning 2013).

3.3.2 ANALYSE

Uit de voorgaande voorbeelden blijkt dat data-analyses vaak een enorme tijdwinst opleveren en dus vele mensuren kunnen schelen. Dat is een aantrekkelijk gegeven in een tijd waarin organisaties steeds efficiënter moeten opereren en overheidsinstanties het vaak met minder middelen moeten doen. Deze ontwikkeling is dan ook een belangrijke stimulans voor innovaties op het terrein van data-analyse bij de overheid (WRR 2011). Bij het politiewerk speelt datatechnologie al veel langer een centrale rol. Bezuinigingen en groeiende zorgen over de veiligheid hebben deze trend versterkt, en stuwen in de richting van meer informatie-gedreven politiewerk. Een soortgelijke ontwikkeling zien we ook terug bij andere organisaties in het veiligheidsdomein (Galdon Clavell 2016; Van Brakel 2016b).

Tegelijkertijd is de toelaatbaarheid van het gebruik van datamining en profielen een terugkerend element in de maatschappelijke en politieke discussie over data-analyses in het veiligheidsdomein. Dat geldt zeker als het gaat om de opsporing van lichte vergrijpen, of bij socialefraudegevallen, die eerder optreden als gevolg van veranderde leefsituaties en onwetendheid dan dat er sprake is van doelbewust crimineel handelen. Een vraag die aan belang wint, is hoe profielen tot stand komen en welke indicatoren ten grondslag liggen aan de verwachting dat mensen bepaald gedrag zullen gaan vertonen (Zarsky 2013; Citron en Pasquale 2014; Pasquale 2015). Naarmate er meer 'zachte', sociale data aan de mix toegevoegd worden, stelt dat hogere eisen aan de kwaliteit van algoritmen en analyses en aan het gebruik van de uitkomsten in de praktijk. Als een systeem waarschijnlijkheden produceert, moeten die in de praktijk ook als waarschijnlijkheden – en niet als zekerheden – worden gebruikt. De vraag is of die technische en professionele kwaliteit in de praktijk wel voldoende geborgd is.

De opkomst van Big Data-processen in het veiligheidsdomein heeft bovendien als effect dat instrumenten en technieken die aanvankelijk vooral in het kader van nationale veiligheid werden ingezet, nu ook voor andere, vooral lichtere vormen van misdaadbestrijding en toezicht worden gebruikt. Zo wordt 'intelligence' steeds belangrijker, en daarmee ontwikkelen organisaties als de politie, organisaties die digitale surveillance uitoefenen en inspecties een werkwijze die overeenkomsten vertoont met die van de inlichtingendiensten (Završnik 2013). Volgens Galdon Clavell (2016) zijn de grenzen tussen de verschillende veiligheidsdomeinen sowieso al langer aan het vervagen. De dataficatie van de samenleving en vermenging van het gewone leven met het leven online draagt verder bij aan die vervagende grenzen (WRR 2015). Het uitwisselen en koppelen van data en de oprichting van organisaties die op verzoek van publieke organisaties data-analyses

kunnen uitvoeren, zal dit proces nog verder versterken. Zeker wanneer die analyses behulpzaam blijken te zijn bij de uitvoering van bepaalde veiligheidstaken, is de kans groot dat een groeiend aantal organisaties daar gebruik van zal willen maken. Dit is reeds waarneembaar bij een samenwerkingsverband als de ICOV, waarbij zich steeds nieuwe organisaties (willen) aansluiten, waarvan op het eerste gezicht niet duidelijk is hoe de opsporing van crimineel en onverklaard vermogen zich tot hun taakstelling verhoudt.

Data-analyses zullen ook steeds complexer worden en zullen een steeds belangrijkere rol gaan spelen bij het nemen van beslissingen in het veiligheidsbeleid. Dit betekent dat de inzet van algoritmen en *machine learning* groter zal worden. Correctieprocedures – in de software zelf of extern door middel van transparantie, audits of toezicht – zullen hierdoor een groter gewicht in de schaal leggen. Het is immers niet vanzelfsprekend dat computers minder fouten zullen maken, zeker niet in de beginfase van een project. Omdat mensen hun gedrag aanpassen in reactie op profielen, is het bovendien nodig om de uitkomsten van analyses regelmatig opnieuw te valideren.

Bij de opzet en uitvoering van data-analyses zijn vaak private partijen betrokken. Zij adviseren niet alleen over de bouw en het beheer van databanken, maar leveren soms ook de software om data-analyses uit te voeren. Soms wordt het analytische hart van een Big Data-toepassing gevormd door een commercieel algoritme dat als bedrijfsgeheim wordt gezien en dus niet inzichtelijk is voor de overheidsorganisatie die het systeem gebruikt. De black box van een Big Data-toepassing in het veiligheidsdomein herbergt dan nog een kleinere black box in zich van het commerciële algoritme. Hierdoor dreigen overheden afhankelijk te worden van de (grote) bedrijven met wie zij publiek-private samenwerkingen aangaan.

3.3.3 GEBRUIK

Het toekomstig gebruik van Big Data is nog het meest lastig te duiden, zeker omdat een aantal initiatieven en experimenten in de loop van de tijd omstreden is geraakt. Het is echter aannemelijk dat het gebruik van Big Data-processen de komende jaren sterk zal toenemen. De vraag is alleen hoe. Momenteel worden Big Data-processen vooral gebruikt als aanvulling op de bestaande methoden van onderzoek, opsporing en toezicht. Ook worden Big Data-processen aangewend om middelen efficiënter te kunnen inzetten en als manier om de dienstverlening aan burgers te verbeteren, al gebeurt dat laatste interessant genoeg opvallend weinig.

Binnen het veiligheidsdomein is preventie de grote belofte van Big Data. Maar in de praktijk ligt het accent vooral op repressie, mede gestuurd door informatie afkomstig uit Big Data-processen. Big Data-processen vinden plaats met als doel om opsporing mogelijk te maken. De belofte om burgers, organisaties en bedrijven

beter toe te rusten voor de omgang met mogelijke risico's en dreigingen komt echter niet of nauwelijks aan de orde. Als regel geldt dat hoe kleiner de vergrijpen zijn waarmee een organisatie te maken heeft des te belangrijker het is om het gebruik van Big Data-processen niet uitsluitend voor handhavingsdoeleinden te gebruiken. Hoewel dit van groot belang lijkt om het vertrouwen in Big Data te vergroten, wijst veel erop dat de inzet voornamelijk dienstbaar zal zijn aan een reeks van meer repressieve doeleinden binnen het veiligheidsdomein.

Ten tweede valt te verwachten dat Big Data een steeds prominentere plek in de besluitvorming zal gaan spelen. Een interessant voorbeeld is de eerdergenoemde software Intrado Beware waarmee momenteel een pilot loopt bij verschillende politiediensten in de Verenigde Staten. De maker claimt dat de software binnen een aantal seconden miljoenen data (uit commerciële databases, publieke databases en sociale media) kan vinden, sorteren en scoren en op deze manier potentieel gevaarlijke situaties of individuen kan melden aan een politieagent die onderweg is na een noodoproep via 911 (Intrado 2012; Skorup 2014). Het voorbeeld toont dat Big Data-processen steeds vaker een realtime-karakter krijgen. De vraag in welke mate Big Data een rol in de besluitvorming gaat spelen en interventies zal gaan sturen, is echter tegelijkertijd sterk afhankelijk van de vraag welke foutmarges we bereid zijn te accepteren. Zeker wanneer Big Data-analyses worden gecombineerd met realtime-waarnemingen, bijvoorbeeld door beveiligingscamera's, en de dreiging groot is, bestaat er weinig ruimte voor toetsing en menselijke afwegingen.

De juridische context en legitimiteit van het gebruik van Big Data roepen tegelijkertijd de nodige vragen op. Om te beginnen weten burgers vaak niet welke gegevens er van hen bewaard en/of verzameld worden, hoewel deze gegevens gebruikt worden om risicoprofielen te construeren en overtredingen en strafbare feiten te traceren. Vervolgens zijn de meeste analyses letterlijk een black box: alleen de direct betrokkenen weten welke gegevens erin gaan, welke indicatoren en wegingsfactoren worden gebruikt, en welke 'hits' dat oplevert. Tot slot is onduidelijk welke rol en welk gewicht de uitkomsten in de besluitvorming spelen. Zijn het enkel aanwijzingen voor nader onderzoek of worden ze al min of meer als bewijs beschouwd? Surveillance en vele vormen van (voor)onderzoek zijn in het veiligheidsdomein aan minder strikte regels gebonden dan het strafrecht. Zelfs wanneer Big Data-processen zich hiertoe beperken, heeft dit consequenties voor de schaal en intensiteit van de digitale en 'reguliere' surveillance en komt derhalve ook de presumptie van onschuld onder druk te staan. Niet alleen doordat de gegevens van een steeds groter aantal burgers steeds vaker in Big Data-processen zal worden opgenomen, maar ook omdat er al 'bewijs' wordt verzameld voordat er sprake is van een redelijke verdenking.

De grote vraag blijft dan ook in hoeverre de overheid bereid is om openheid van zaken te geven over de mate waarin en hoe zij Big Data gebruikt dan wel van plan is te gebruiken. Uit het overheidshandelen zelf valt dat immers vaak niet af te leiden. Dat komt soms omdat Big Data-processen verborgen zitten in de backoffice van overheidsinstellingen en soms omdat de hierboven beschreven praktijken achter een sluier van geheimhouding verborgen zijn. Deze geheimhouding dient om te voorkomen dat verdachten zich kunnen wapenen tegen de gehanteerde opsporingsmethoden en/of deze kunnen manipuleren. 'Gaming the system' is een reëel risico, maar kan ook als stoplap fungeren om enige transparantie over de werkzaamheden uit de weg te gaan. Uit het voorbeeld van de Inlichtingendiensten blijkt dat landen nogal kunnen verschillen in de mate waarin zij informatie over aantallen en soorten inlichtingenoperaties geven. Ook op andere terreinen zit er veel licht tussen totale geheimhouding en een volledig transparante werkwijze, die organisaties machteloos maakt tegenover degenen die zij bestrijden.

3.4 CONCLUSIE

Big Data speelt binnen het veiligheidsdomein nog een beperkte rol. Er zijn weinig praktijkvoorbeelden die voldoen aan alle kenmerken. Sommige onderdelen van het veiligheidsdomein, zoals het inlichtingenwerk en de fraudebestrijding, getuigen desondanks van een verregaande ontwikkeling op dit terrein. Bij diverse andere organisaties en samenwerkingsverbanden van organisaties zijn de voorwaarden aanwezig voor een kwalitatieve sprong in de methoden om data te verwerken en te analyseren. Ondanks de omvangrijke preparatie en koppeling van databases lijken veel organisaties binnen het veiligheidsdomein nog niet de keuze te hebben gemaakt om grootschalige data-gedreven analyses uit te voeren waarbij correlaties de uitkomsten bepalen.

Het gebruik van Big Data-achtige technieken levert vaak positieve resultaten op. Er zijn hoge opbrengsten in termen van fraudebestrijding doordat data-analyses een scherper beeld geven dan steekproeven over het geheel of grote subgroepen. Ook zijn ze vaak veel sneller en met minder menskracht uit te voeren, waardoor overheidsorganisaties hun schaarse middelen veel efficiënter kunnen inzetten. Bovendien vergroot dit de reactiesnelheid van opsporingseenheden en surveillanten, zeker wanneer analyses realtime plaatsvinden. Tegelijkertijd is de inzet van Big Data-processen nog onvoldoende grondig onderzocht om er harde conclusies aan te verbinden. Dat is niet alleen te wijten aan de kwaliteit van het onderzoek, maar heeft ook te maken met het feit dat de effecten van data-analyses moeilijk te scheiden zijn van andere beleidsinspanningen op het terrein van de misdaad- en fraudebestrijding. Op sommige terreinen, zoals het inlichtingenwerk, is het vanwege de geheimhouding sowieso lastig om de effectiviteit van operaties aan te tonen.

Hiertegenover staan maatschappelijke zorgen over de impact van grootschalige data-analyses. Door de nieuwe technologische mogelijkheden kunnen overheidsorganisaties met dezelfde bevoegdheden veel dieper inzicht krijgen in het persoonlijk leven van burgers dan voorheen. Voor een deel zit dat in de koppeling van steeds grotere en meer diverse gegevensbestanden, waardoor een steeds preciezer en volledig beeld van de handel en wandel van burgers en bedrijven is te construeren. Voor een ander deel heeft dat te maken met het gebruik van technieken als profiling en patroonherkenning, die vaak ook betrekking hebben op de gegevens van burgers die in strikte zin (nog) geen object van onderzoek en/of opsporing zijn. Het groeiend belang van automatische gegevensanalyse roept daarenboven de vraag op in hoeverre burgers zich tegen besluiten kunnen verweren, zeker wanneer de totstandkoming daarvan door geheimhouding aan het oog is onttrokken. De huidige juridische waarborgen lijken deze zorgen onvoldoende te kunnen wegnemen.

De inzet van Big Data in het veiligheidsdomein lijkt dus zowel positieve als negatieve effecten te kunnen hebben. Het nu volgende hoofdstuk bevat een nadere evaluatie van de belofte van Big Data voor het veiligheidsdomein en de risico's die daarmee samenhangen.

NOTEN

1 Zie www.volkskrant.nl/binnenland/van-der-steur-komt-belofte-over-betere-ict-politie-niet-na~a4239444.

2 De lijst met gegevens waarvan de Belastingdienst gebruik mag maken telt vele pagina's, zie http://download.belastingdienst.nl/belastingdienst/docs/meldingen_belastingdienst_2008_al5303z4fd.pdf, geraadpleegd 18 februari 2016.

3 Deze convenanten zijn te vinden via: www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/intermediairs/toezicht/convenanten/convenanten_met_afnemers_van_informatie/convenanten_met_afnemers_van_informatie_van_de_belastingdienst.

4 Volgens Olsthoorn (2016) is de bewaartermijn gebaseerd op de Archiefwet, maar die kent geen privacyprincipes. De gegevens mogen daarom worden bewaard zo lang ze als fiscaal relevant worden geacht, wat in de praktijk neerkomt op zo'n 5 tot 7 jaar. Voor deze problematiek, en hoe een aantal juristen hierover denkt, zie ook: https://decorrespondent.nl/1766/Vergeet-de-politiestaat-Welkom-in-de-belastingstaat/90525160-5a2c27c2.

5 Zie https://decorrespondent.nl/2720/Baas-Belastingdienst-over-Big-Data-Mijn-missie-is-gedragsverandering/83656320-f6e78aaf.

6 Dit zegt directeur Hans Blokpoel in een interview met de Correspondent, zie https://decorrespondent.nl/2720/Baas-Belastingdienst-over-Big-Data-Mijn-missie-is-gedragsverandering/83656320-f6e78aaf.

7 Zie 15e halfjaarrapportage Belastingdienst, 26 maart 2015, p. 20, www.eerstekamer.nl/overig/20150326/15e_halfjaarsrapportage_van_de/document, geraadpleegd 23 februari 2016.

8 De volledige uitspraak is te vinden via: http://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBOBR:2013:6553. De Belastingdienst is op eigen initiatief gestopt met het vorderen van grote bestanden met kentekengegevens bij parkeerbeheerders.

9 Staatsblad (2014) 'Besluit van 1 september 2014 tot wijziging van het Besluit SUWI in verband met regels voor fraudeaanpak door gegevensuitwisselingen en het effectief gebruik van binnen de overheid bekend zijnde gegevens met inzet van SyRI, jaargang 2014, nr. 320.

10 Voor een korte beschrijving van dit project zie behalve Olsthoorn (2016) ook Inspectie SZW (2012), Bestandskoppelingen bij fraude. Nota van bevindingen, Den Haag. Beschikbaar via: www.inspectieszw.nl/Images/Bestandskoppeling-bij-fraudebestrijding_tcm335-335618.pdf. Voor juridische details zie voorts H.S. Prins en A. Bok (2010: 149-160).

11 De uiteenzetting over de werkwijze van SyRI is gebaseerd op het Besluit SyRI en de daarbijbehorende Nota van Toelichting.

12 Voor verdere uitleg over het in 2001 door het ministerie van SZW, VNG en Divosa opgerichte Inlichtingenbureau, zie Olsthoorn (2016).

13 Zie www.volkskrant.nl/politiek/burger-wordt-straks-doorgelicht-zoals-profiel-van-crimineel-wordt-opgesteld~a3759563.

14 Voor dit verweer zie: www.rijksoverheid.nl/actueel/nieuws/2014/10/01/reactie-minister-op-artikel-in-de-volkskrant-over-fraudeaanpak.

15 Voorstel van Wet op de inlichtingen- en veiligheidsdiensten 20XX; memorie van toelichting (consultatieversie juni 2015), beschikbaar via: www.internetconsultatie.nl/wiv.

16 Voor deze problematiek zie: Commissie van Toezicht betreffende de inlichtingen- en veiligheidsdiensten, *Toezichtsrapport inzake gegevensverwerking op het gebied van telecommunicatie door de AIVD en de MIVD*, rapport nr. 38, 2014, p. 34 & 35; en *Kabinetsstandpunt herziening interceptiebestel Wiv 2002*, 21 november 2014, p. 6.

17 Zie: www.internetconsultatie.nl/wiv.

18 CCTV staat voor Closed Circuit Television, oftewel video surveillance.

19 Zie onder meer www.ams-institute.org/events/event/de-mobiele-stad-drukte-in-kaart/ en http://dat.nl/nl/cases/monitoring-sail-amsterdam.

20 Een beschrijving van dit systeem is te vinden in www.bssholland.com/uploads/products/BeWare-Brochure.pdf.

21 Zie hiervoor het Volkskrant-artikel beschikbaar op www.volkskrant.nl/buitenland/vs-geven-toe-reisbeperkingen-gebaseerd-op-aannames~a4117748.

4 EVALUATIE VAN BIG DATA IN HET VEILIGHEIDSDOMEIN

4.1 INLEIDING

Big Data zal een revolutie veroorzaken in de manier waarop we leven, werken en denken, zo stellen Mayer-Schönberger en Cukier (2013). Die claim is door velen overgenomen, zowel in het bedrijfsleven als binnen de overheid. Een veelgeciteerd voorbeeld is dat van de brandweer in New York, die voorheen bij de eerste 25 procent van de jaarlijkse inspectie van de brandveiligheid van 25.000 adressen 21 procent overtredingen constateerde. Door toepassing van Big Data-analyses, waarbij op basis van verschillende indicatoren vooraf eerst de meest risicovolle panden werden geïdentificeerd, steeg dat percentage tot 71 (Mayer-Schönberger en Cukier 2013). Er circuleren echter ook meer weidse vergezichten over het succes van Big Data: daders kunnen effectiever en eerder worden opgespoord, men kan fraudeleuze transacties makkelijker reconstrueren (Schneier 2015) en men kan betrouwbaarder voorspellen welke situaties een verhoogd veiligheidsrisico hebben en wie mogelijke daders en slachtoffers zijn (Perry et al. 2013).

De hoeveelheid informatie waarover overheden kunnen beschikken is enorm vergroot door vooruitgang op het terrein van datamining en -analyse, door de toegenomen computerkracht en door goedkopere dataopslag. Tegenover de vele beloften die deze ontwikkeling in zich draagt, staan echter ook vele zorgen, waarbij zorgen over privacy zonder enige twijfel bovenaan staan. Onderzoekers hebben de afgelopen jaren herhaaldelijk aangetoond dat zelfs geanonimiseerde data zonder veel moeite tot specifieke individuen te herleiden zijn (Tene en Polonetsky 2012: 65). Anderen wijzen op de zogenoemde transparantieparadox, die zegt dat burgers steeds transparanter worden voor de overheid, terwijl andersom overheidsorganisaties en de manier waarop zij analyses toepassen niet zo open zijn voor burgers (Richard en King 2013). Volgens deze paradox gaat het gebruik van Big Data gepaard met een (verdere) verschuiving in het machtsevenwicht tussen overheid en burgers. Overheden weten steeds meer van burgers, terwijl Big Data-processen het hoe en wat van de gegevensverzameling aan de openbaarheid onttrekken. Al langer is er bovendien de zorg dat het gebruik van algoritmes en automatische besluitvorming tot een toename van sociale stratificatie en uitsluiting kan leiden (Zarsky 2014).

Het lijkt dus nodig de beloftes van Big Data af te wegen tegen de zorgen die daarover bestaan. Dit hoofdstuk bevat daarom een evaluatie van de opkomst van Big Data in het veiligheidsdomein. Naast de beloftes van Big Data komen daarbij ook de beperkingen van Big Data aan de orde, evenals de randvoorwaarden voor een

goed gebruik daarvan. Tot slot komen de risico's van het gebruik van Big Data in het veiligheidsdomein aan bod. Niet alle mogelijkheden van Big Data zijn immers ook wenselijk.

4.2 BELOFTES

De private sector loopt voorop bij de introductie van Big Data-toepassingen. Bedrijven als Google, Facebook en vele andere datagiganten hebben de verzameling en analyse van data tot hun voornaamste inkomstenbron gemaakt. Fysieke goederen maken, in tegenstelling tot data en software, een zeer klein deel uit van hun bedrijfswaarde. Maar ook overheden bezitten van oudsher veel persoonsgegevens, die gebruikt kunnen worden voor Big Data-analyses. De overheidsorganisaties die binnen het veiligheidsdomein opereren, hebben de bevoegdheid om ook bij derde partijen gegevens op te vragen, mits dat binnen hun taakstelling past. Dit geeft hun bijzonder veel mogelijkheden om met Big Data aan de slag te gaan. En daar zijn ook goede redenen voor. Volgens de Amerikaanse veiligheidsexpert Bruce Schneier (2015) kan Big Data een positieve bijdrage aan het veiligheidsbeleid leveren. Sommige van deze toepassingen, zoals *predictive policing* en het gebruik van Big Data voor beslissingen over verlof en borgtocht van gedetineerden, hebben in de Verenigde Staten en elders hun nut reeds bewezen, zij het op kleine schaal (Bennet Moses en Chan 2014).

4.2.1 EFFICIËNTIE

Een omvangrijk deel van de literatuur over Big Data gaat over ontwikkelingen in de private sector en leest als een lofzang op de vele mogelijkheden die Big Data-toepassingen kunnen hebben voor innovatie en economische groei (Klous 2016). Volgens het McKinsey Global Institute (2011) kan het gebruik van Big Data leiden tot een half tot één procent productiviteitsgroei in de detailhandel in de Verenigde Staten en tot 300 miljoen dollar besparingen in de Amerikaanse gezondheidszorg. Een andere 'berekening' in dezelfde studie toont dat de wereldwijde beschikbaarheid van persoonlijke locatiedata meer dan 100 miljoen dollar extra inkomsten voor dienstverleners en 700 miljoen dollar waarde voor de eindgebruikers kan opleveren. IDC (2012) schatte de markt voor Big Data-technologie en dienstverlening wereldwijd in 2011 op 4,8 miljard dollar en voorspelde dat deze markt zou groeien naar 16,9 miljard dollar in 2015, een jaarlijkse groei van bijna 40 procent.

Deze claims worden ondersteund door enkele meer grondige empirische studies. Zo toonde onderzoek van het MIT aan dat bedrijven die data-gedreven besluitvorming toepassen een toename van 5 procent in productiviteit waarnemen (Brynjolfsson et al. 2011). Bakhshi et al. (2014) vinden in hun onderzoek naar het gebruik van klantgegevens bij 500 bedrijven een nog hoger percentage (namelijk 13% bij de top-25% van deze bedrijven). Een ander voorbeeld is het gebruik van

Big Data om de juist geschatte aankomsttijd (*estimated time of arrival*, ETA) van burgervliegtuigen te geven. Eerder werd hiervoor vertrouwd op de schattingen van piloten. Tegenwoordig wordt een combinatie van verschillende publieke en private databronnen (zoals weerberichten, vluchtschema's en radarinformatie) gebruikt om de aankomsttijd te voorspellen. Door deze toepassing – RightETA getiteld – is het verschil tussen geschatte en werkelijke aankomsttijden geminimaliseerd, hetgeen heeft geleid tot efficiëntere processen en kostenbesparingen op vliegvelden (McAfee en Brynjolfsson 2012). Inmiddels zijn legio voorbeelden van dit soort toepassingen voorhanden, zoals bij supermarktketens, bedrijven in de maakindustrie en zelfs relatief laagtechnologische sectoren als de textielindustrie en de landbouw (OECD 2014; Klous 2016).

Ook de publieke sector kan veel baat hebben bij het gebruik van Big Data. Toepassing ervan kan bijdragen aan een vermindering van overheidsuitgaven (dankzij hogere operationele efficiëntie), een meer effectieve belastinginning (door maatwerk) en minder fraudegevallen en fouten door automatische gegevensanalyse. Deze winst is het grootste in sectoren als onderwijs, gezondheidszorg en openbaar bestuur (OESO 2014: 19). Deze sectoren kennen namelijk een relatief groot aandeel beroepen waarvan de kerntaken met de verzameling en analyse van data en informatie te maken hebben. Big Data heeft in de zorg zijn intrede al gedaan en bedrijven zoals Apple, Google, Samsung en IBM beschouwen de zorg als een aantrekkelijke groeimarkt (Ottes 2016). Big Data kan een bijdrage leveren aan het opvullen van kennislacunes over de precieze oorzaak, of het samenspel van oorzaken, van ziekten en aandoeningen en van het effect van behandelingen. Een belangrijke voorwaarde hiervoor is de integratie (en standaardisering) van grote hoeveelheden gegevens vanuit bronnen als genenbanken, medisch dossiers en omgevingsfactoren (Ottes 2016). De verwachtingen op dit terrein zijn hooggespannen. Of zoals Mayer-Schönberger en Cukier (2013: 61) het kort en krachtig stellen: "big data saves lives".

Het veiligheidsdomein kent vele kennis- en informatie-intensieve terreinen, die door gebruik van Big Data-toepassingen efficiënter en effectiever kunnen functioneren. Zo blijkt bij 'risicogestuurd adresonderzoek' om rechtmatige bewoning te controleren het rendement van een Big Data-benadering tienmaal zo groot als bij steekproefsgewijs huisbezoek (ICTU 2014). Analyses die voorheen soms dagen, weken of maanden duurden, kunnen nu in enkele uren of minuten worden uitgevoerd, zoals de iCOV heeft laten zien (Perry et al. 2013). De grote hoeveelheid gebruikte gegevens maakt bovendien een scherpere risicoanalyse mogelijk. Behalve voor de in paragraaf 4.1 genoemde brandweer van New York geldt dat ook voor het Criminaliteits Anticipatie Systeem (CAS), dat soortgelijke opbrengsten lijkt op te leveren, maar dan op het terrein van criminaliteitsbestrijding. Ook inspecties blijken door het gebruik van Big Data gerichter te kunnen controleren en realiseren daardoor hogere opbrengsten (Custers 2014). Overeenkomstig deze

voorbeelden is het gebruik van risicoprofielen en bijbehorende risicoscores van invloed op de inzet van middelen. Door Big Data-analyses kunnen schaarse middelen, zoals ambulances, mobiele eenheden of politiehelikopters, worden ingezet wanneer en waar ze de grootste kans hebben om het meeste effect te sorteren.

4.2.2 TERUGKIJKEN, REALTIME-ANALYSES EN MISDAAD VOORSPELLEN

Het gebruik van Big Data biedt sectoren echter niet alleen mogelijkheden om zich anders te organiseren, maar ook om nieuwe producten en diensten te ontwikkelen en dingen te doen die voorheen niet mogelijk waren (Klous 2016). Daarvan zijn er binnen het veiligheidsdomein ten minste drie: het maken van historische reconstructies van misdaden, het realtime monitoren van risico's, en het voorspellen van het wie en wanneer van delicten.

Historische analyses

Volgens internetveiligheidsexpert Schneier (2015: 35-45) stelt Big Data de gebruikers ervan onmiskenbaar in staat 'ever more startling conclusions' uit grote datasets te trekken. Nieuw is volgens hem dat datamining het mogelijk maakt om gebeurtenissen uit het verleden te reconstrueren, simpelweg omdat databases steeds meer historische informatie bevatten. Hierdoor is het bijvoorbeeld mogelijk om na te gaan welke locaties een persoon heeft bezocht en welke telefoontjes hij of zij heeft gepleegd. Zo werd bij het onderzoek naar de daders van de bomaanslagen in Boston gebruikgemaakt van materiaal van camera's in de openbare ruimte, mobieletelefoongegevens en een gemeentelijk systeem voor gezichtsherkenning. Ook kunnen oude data met nieuwe methoden worden doorzocht, waardoor nieuw bewijs aan het licht is te brengen. Historische analyses zijn volgens Schneier bovendien bij uitstek geschikt om frauduleuze financiële transacties te traceren, ook over landsgrenzen heen. In het verleden was het vanzelfsprekend evenzeer mogelijk historische analyses te maken, maar de data zijn nu meer volledig, tegen lagere kosten verkrijgbaar en de kwaliteit van 'historische' analyses is sterk verbeterd.

Realtime-analyses

Big Data-analyses zijn daarnaast te gebruiken om realtime mensen, objecten en gebeurtenissen te volgen. De combinatie van camera's, gezichtsregistratiesoftware en grote bestanden met digitale foto's, van rijbewijzen, paspoorten, vervoersbewijzen, facebookpagina's of nieuwsberichten maken het bijvoorbeeld mogelijk om automatisch personen te identificeren (Omand et al. 2012). Zulke koppelingen zijn relatief eenvoudig te maken en komen in de praktijk al voor doordat bedrijven gegevens met elkaar uitwisselen, over zowel onlinegedrag als offlinegedrag dat digitaal geregistreerd wordt, zoals bij pinbetalingen. Dit soort technieken gebruikt de politie in de praktijk bij *crowdcontrol* van grootschalige evenementen. Op basis van locatiegegevens maken politiemedewerkers een schatting van het aantal mensen dat deelneemt aan demonstraties of risicowedstrijden bijwoont en kunnen zij

bewegingspatronen in kaart brengen. Aan de hand daarvan bepaalt men de inzet van middelen. Dit gebeurt in Nederland rond evenementen als Sail Amsterdam en in bepaalde straten in drukke uitgaansgebieden (zoals in Eindhoven, beschreven in Kool et al. 2015). Door dit soort gegevens te combineren met sociale media en/of risicoprofielen kan men bovendien de kans op escalatie beter inschatten. In het Verenigd Koninkrijk gebruikt de politie SOCMINT (Social Media Intelligence) om zich een meer accuraat beeld te vormen van het aantal agenten dat nodig is om de openbare veiligheid te waarborgen (Omand et al. 2012).

Misdaad voorspellen

De hoogste verwachtingen rond Big Data-analyses hebben te maken met het voorspellen van toekomstige overtredingen, misdaden en dreigingen. Big Data-voorspellingen kunnen in het algemeen verschillende doelen dienen (Kerr en Earle (2013). In het private domein dienen ze vooral om de preferenties van burgers op het spoor te komen (*preferential predictions*). Op basis van dergelijke voorspellingen krijgen consumenten bijvoorbeeld gericht muziek, boeken of films aangeboden. Voorspellingen zijn ook te gebruiken om bepaalde acties of handelingen van individuen of groepen te blokkeren (*pre-emptive predictions*). Het doel hierbij is om de handelingsvrijheid van mensen in te beperken. Een goed voorbeeld zijn *no-fly lists*. Voorspellingen met Big Data kunnen tot slot aangewend worden om in te schatten wat de consequenties van sommige gedragingen zijn (*consequential predictions*). Voorspellingen worden in dat geval gebruikt om mensen te helpen de keuzes te maken die waarschijnlijk goed voor ze zijn (Van Brakel 2016). Denk aan een ander slot op de deur als er in een nieuwbouwwijk met dezelfde woningen veel wordt ingebroken. Mensen kunnen met deze informatie een betere afweging maken over de situatie waarin zij zich bevinden. *Preferential predictions* zijn weinig relevant voor het veiligheidsbeleid maar *consequential predictions* kunnen daar een grote bijdrage aan leveren. Binnen het veligheidsdomein gaat de aandacht vooralsnog vooral uit naar *pre-emptive predictions*.

Perry et al. (2013) onderscheiden drie verschillende methoden die door de politie worden gebruikt om misdaad te voorkomen: (1) het voorspellen van de tijd en plaats waar een verhoogd risico op misdaad is; (2) het identificeren van toekomstige daders en slachtoffers; en (3) het ontwikkelen van daderprofielen voor specifieke misdaden. Bij dat laatste gaat het om het berekenen van de kans dat een persoon met bepaalde kenmerken een bepaalde misdaad zal begaan. De toepasbaarheid van deze methoden hangt overigens sterk samen met de combinatie van voldoende informatie over gebeurtenissen en duidelijke stijlen of patronen in crimineel gedrag. Pas dan is het optreden van deze gebeurtenissen daadwerkelijk voorspelbaar (Ratcliffe 2010). Voorspellingen kunnen worden gebruikt om de inzet van middelen te sturen. Daarnaast stellen *heat maps*, lijsten met hotspots en hot persons, de betrokken organisaties in staat hun omgevingsbewustzijn te

verbeteren en op tactisch en strategisch niveau strategieën te ontwikkelen om efficiënter en effectiever te werken (Perry et al. 2013). Anders gezegd: Big Data vergroot de pakkans.

De meest tot de verbeelding sprekende toepassing van Big Data-analyses heeft te maken met de identificatie van potentiële daders en – in mindere mate – potentiële slachtoffers. Deze toepassing is overigens ook op andere terreinen inzetbaar, bijvoorbeeld dat van de fysieke veiligheid. Men kan hier bijvoorbeeld denken aan het identificeren van schepen of containers met een verhoogd risico van illegale goederen. Deze toepassingen zijn specifiek gericht op misdaadbestrijding en het oplossen van een bepaald soort veiligheidsproblemen. Met andere woorden, meer nog dan bij het voorspellen van plekken met een verhoogd veiligheidsrisico dragen Big Data-analyses hier bij aan een proactief veiligheidsbeleid, dat zich behalve op repressie ook in toenemende mate richt op preventie. Het is vanzelfsprekend heel aantrekkelijk om onveilige situaties te voorkomen in plaats van de schuldigen aan te moeten houden nadat een strafbaar feit is gepleegd.

Potentiële daders zijn onder meer te vinden door hun interpersoonlijke relaties in kaart te brengen, waardoor men criminele netwerken op het spoor komt en die kan ontmaskeren. Vooral metadata van gevarieerde bronnen zijn hiervoor zeer geschikt, omdat ze verschillende sociale interactiepatronen blootleggen. Hiernaast zijn Big Data-analyses te gebruiken om personen te vinden en/of te volgen die een bepaald gedrag vertonen. Wanneer bekend is dat bepaald gedrag een verhoogd risico op misdaad met zich meebrengt, kan preventief actie worden ondernomen. Zo worden Big Data in de Verenigde Staten gebruikt bij besluiten over verloftoekenning aan gevangenen, hoewel daar ook problematische kanten aan zitten (Bennet Moses en Chan 2014; Harcourt 2007).

Al met al kunnen Big Data-analyses bijdragen aan een meer efficiënte en effectieve uitvoering van de veiligheidstaak, wat door een grotere nadruk op kostenefficiënt werken de afgelopen jaren steeds belangrijker is geworden (Galdon Clavell 2016). Ook kunnen Big Data-analyses kennis en informatie opleveren die de kwaliteit en efficiëntie van onderzoek en opsporing aanzienlijk verbeteren. De hoogste — maar tegelijkertijd ook meest lastig te realiseren — opbrengst van Big Data-analyses ligt in het vroegtijdig signaleren van potentiële daders en slachtoffers. Hierdoor kunnen veiligheidsfunctionarissen tijdig preventieve maatregelen treffen, zodat zich uiteindelijk minder (on)veiligheidsincidenten voordoen. Dit soort toepassingen is in de praktijk nog veelal toekomstmuziek.

4.3 BEPERKINGEN

Big Data is niet voor alle veiligheidsvraagstukken een even geschikt middel. Bovendien kent de toepassing Big Data-analyses – evenals veel andere vormen van statistische analyse – de nodige beperkingen. Neem Google Flu Trends, een applicatie die op basis 45 zoektermen de opkomst van griepgolven realtime beweerde te kunnen volgen en in kaart te kunnen brengen. Dat was twee weken sneller dan de officiële overheidskanalen. Toch bleek Google Flu Trends de griep structureel te overschatten en miste de applicatie de buiten het winterseizoen optredende A–H1N1 pandemie in 2009 volledig (Lazer et al. 2014). Het voorbeeld illustreert dat ook bij de analyse van grote hoeveelheden data de resultaten niet automatisch correct zijn. Maar er zijn nog andere beperkingen verbonden aan Big Data-analyses, zoals statistische tekortkomingen en beperkingen in het toepassingsbereik van de analysetechniek.

4.3.1 STATISTISCHE BEPERKINGEN

Een eerste statistisch probleem is bias. Data worden in een specifieke context verzameld. Daardoor zit in bijna iedere dataset een specifieke bias. Alles hangt dus af van de discrepantie tussen de dataset en de populatie waarover men een uitspraak wil doen. Met de nieuwste technieken is het bijvoorbeeld mogelijk om alle creditcardtransacties realtime te analyseren. Bij fraudedetectie binnen creditcardtransacties gaat het om data van de gehele populatie – er worden immers uitspraken gedaan over alle creditcardtransacties. Maar deze creditcardtransacties zijn onvoldoende als basis om uitspraken te doen over het koopgedrag van de inwoners van Nederland. Hooguit gaat het dan om een grote, maar niet aselecte steekproef. Hetzelfde geldt voor sociale media: het is een misverstand dat op basis van Twitterdata een uitspraak valt te doen over de gehele populatie in een land (Tufekci 2014). Er zijn altijd groepen mensen van wie de data niet regelmatig verzameld en geanalyseerd worden, omdat ze niet of onvoldoende participeren in het (online)gedrag dat als basis voor veel Big Data-analyses dient (Lerman 2013). Ook is Twitter geen willekeurige sample van de etnische samenstelling van de bevolking (Mislove et al. 2011). Vooral bij het combineren van databronnen en hergebruik kan het lastig zijn om te achterhalen hoe datasets tot stand zijn gekomen en dus wat de precieze bias is die in de data zit.

Onvermijdelijke vooronderstellingen

Anderson publiceerde in 2008 een veelbesproken artikel onder de titel ‘The End of Theory: The Data Deluge Makes the Scientific Method Obsolete’. Hierin stelt hij dat Big Data het einde van theorie en de wetenschappelijke methode betekent. Het traditionele proces van wetenschappelijke ontdekking heeft volgens Andersen zijn beste tijd gehad. Het is niet langer nodig om met behulp van een model een hypothese te toetsen aan de werkelijkheid. Wat volgens hem volstaat is een statistische analyse van zuivere correlaties zonder theorie. Hier valt veel op af te dingen. Algo-

ritmen vertellen niet uit zichzelf wat de data betekenen of hoe ze te interpreteren zijn. De onderzoeksvraag, de dataselectie en de interpretatie van de uitkomsten bevatten onvermijdelijk vele theoretische vooronderstellingen (Gillespi 2014).

Correlatie ≠ causaliteit

Volgens Anderson (2008) zijn ook causale verbanden in het tijdperk van Big Data niet langer relevant. Het kennen van correlaties is voldoende. Causaliteit betekent dat variabele A variabele B veroorzaakt. Correlatie betekent alleen dat variabele A en variabele B samenhangen. Een gevonden correlatie die statistisch significant is, is dan ook niet noodzakelijk causaal. Zonder theorie over het waarom van correlaties en zonder goede aanwijzingen dat deze causaal zijn, kunnen daarop gebaseerde interventies de plank volledig misslaan (Harford 2014). Een onschuldig voorbeeld is het boekenplan van de Amerikaanse staat Illinois (Levitt en Dubner 2005). Onderzoek liet zien dat de aanwezigheid van meer boeken in huis sterk verband hield met betere studieprestaties van kinderen. De staat had daarom bijna boeken naar ieder kind gestuurd. Later bleek echter dat kinderen uit huizen met veel boeken ook betere studieresultaten behaalden wanneer zij geen enkel boek lazen. Het causale verband was waarschijnlijk dat boeken in huis op een stimulerende leeromgeving wijzen. Of correlaties voldoende informatie bieden om er conclusies op te baseren of dat het toch zinvol is om meer over causale verbanden te weten te komen, hangt sterk af van de context, het niveau van de analyse en de soort interventie die daaraan wordt verbonden. Vooral binnen het veiligheidsdomein maakt het nogal uit of informatie aanleiding is voor verder onderzoek of dat ze gebruikt wordt voor handhavings- en/of opsporingsdoeleinden.

Fouten

Big Data-analyses zullen, net als andere statistische analyses, altijd fouten bevatten. Omdat het om waarschijnlijkheden gaat, kunnen er twee soorten fouten in de resultaten ontstaan: vals positieven en vals negatieven. In het geval van valse positieven geeft een uitkomst ten onrechte aan dat een bepaalde conditie aanwezig is. Dit kan bijvoorbeeld gaan om het onjuist constateren van een ziekte, het onterecht aanwijzen van iemand als terrorist of crimineel of een wettige creditcardtransactie als frauduleus bestempelen. Bij vals negatieven geeft de uitkomst aan dat een bepaalde conditie niet aanwezig is terwijl dat in de werkelijkheid wel zo is. Dit is bijvoorbeeld het geval wanneer een creditcardtransactie als legitiem wordt beoordeeld, terwijl het in feite om een frauduleuze transactie gaat. Hoe erg deze fouten zijn, hangt wederom van de toepassing af. Voor het aanraden van boeken op Amazon of het laten zien van een advertentie, is een fout niet heel problematisch. Het wordt al wat ongemakkelijker wanneer de creditcardbetaling in een winkel als frauduleus wordt bestempeld en de kaarthouder eerst met de bank moet bellen alvorens hij kan betalen. En helemaal anders is het wanneer op basis van analyses mensen opgepakt worden of op een *no-fly list* worden gezet.

Meerdere vergelijkingen probleem

Wie een patroon in de data vindt, vraagt zich allereerst af of dit patroon wel of geen toeval is. Als het onwaarschijnlijk is dat het geobserveerde patroon bij toeval is ontstaan, wordt het patroon 'statistisch significant' genoemd. Indien het om meerdere patronen gaat, kan het zogenoemde *multiple-comparisons problem* optreden. Dat wil zeggen dat bij het analyseren van grote hoeveelheden data altijd wel correlaties zijn te vinden. Dit wordt ook wel *data dredging* of *fishing expedition* genoemd. Het gevolg is dat men (sterke) correlaties tussen twee variabelen kan vinden die niets met elkaar te maken hebben. Sprekende voorbeelden zijn het 'verband' tussen hoeveel kaas mensen eten en hoeveel mensen overlijden doordat ze vast komen te zitten in hun dekbed, en de correlatie tussen films waarin Nicolas Cage speelt en het aantal mensen dat verdrinkt in een zwembad (Kool et al. 2015: 46-47). Deze zogeheten schijncorrelaties worden frequenter wanneer men hele grote hoeveelheden data analyseert (Taleb 2013). Een technische oplossing voor dit probleem is verhoging van de significantiedrempel. Een meer principiële oplossing is transparantie, zodat het mogelijk wordt voor onderzoekers om uit te zoeken hoeveel resultaten verkregen zijn die niet interessant genoeg waren om verder te onderzoeken of handelingen aan te verbinden (Ioannidis 2005; Harford 2014).

4.3.2 ONREGELMATIGE EN INCIDENTELE VEILIGHEIDSVRAAGSTUKKEN

Een andere beperking van Big Data-analyses is dat zij niet de meest geschikte oplossing zijn voor ieder veiligheidsvraagstuk. Big Data-analyses ontlenen hun waarde aan patroonherkenning in grote hoeveelheden data. Dit veronderstelt dat veiligheidsvraagstukken een regelmatig karakter hebben. Wanneer die regelmatigheid ontbreekt, is het lastig om ze op het spoor te komen. Een ander probleem kan zijn dat een bepaald vraagstuk zo weinig voorkomt dat er onvoldoende materiaal beschikbaar is om er een betekenisvol patroon uit af te leiden. Zo betoogt Bruce Schneier (2015: 136-139) dat datamining voor het zoeken naar terroristen een ineffectieve methode is ('locating a needle in a haystack'). Het foutpercentage is te hoog omdat er te weinig aanslagen zijn om een goed profiel te maken. Bovendien is elke terroristische aanslag uniek waardoor het lastig, zo niet onmogelijk is daar een patroon op te baseren. Volgens Schneier is datamining in dergelijke gevallen "simply the wrong tool for the job". Uit een beperkt aantal gegevens en qua karakter sterk uiteenlopende gebeurtenissen vallen niet of nauwelijks betekenisvolle patronen af te leiden. De inzet van Big Data-analyses zal derhalve telkens beoordeeld moeten worden in het kader van een breder palet aan onderzoeks- en opsporingsmethodes.

Ook de inzet van *predictive policing* is aan deze beperking onderhevig. In Nederland wordt dit middel momenteel vooral gebruikt om woninginbraken te voorspellen (Van Brakel 2016b). Dat heeft niet alleen met politieke prioriteiten te maken, maar ook met het karakter van dit type misdaad, dat een aantal kenmerken heeft dat betrouwbare voorspellingen mogelijk maakt. Zo vinden veel woningin-

braken rond dezelfde tijdstippen plaats, op specifieke plekken, bijvoorbeeld dichtbij goede uitvalswegen, en is er sprake van een relatief helder omlijnd daderprofiel. Ook straatroof, fietsendiefstallen en winkelovervallen laten zich redelijk tot goed voorspellen, maar voor andere misdrijven – bijvoorbeeld georganiseerde misdaad of mensensmokkel – is dat een stuk lastiger. Bovendien zullen misdadigers zoveel mogelijk proberen opsporing te ontlopen, door (persoons)gegevens te manipuleren en/of te verbergen. Meer in het algemeen zullen burgers hun gedrag aanpassen in reactie op – vooral grootschalige – observatiepraktijken. Het is daarom belangrijk om de uitkomsten van Big Data-analyses voortdurend te valideren en periodiek de gebruikte indicatoren te herzien.

4.4 RANDVOORWAARDEN

Lang niet alle bedrijven en organisaties slagen erin tot waardevolle toepassingen van Big Data te komen. Om het potentieel van Big Data daadwerkelijk te kunnen realiseren, moet aan een aantal voorwaarden zijn voldaan.

4.4.1 BESCHIKBAARHEID EN KWALITEIT VAN DATA

Een eerste randvoorwaarde is de beschikbaarheid van data. Hoewel de hoeveelheid data enorm is gegroeid, zijn daardoor niet meteen ook alle *relevante* data beschikbaar. Over bepaalde onderwerpen, zoals zoekgedrag op internet, activiteit op sociale media en mobieletelefoongebruik, is de hoeveelheid data enorm, terwijl die over andere onderwerpen veel geringer is. Ook geldt dat in bepaalde landen veel meer data beschikbaar zijn dan in andere landen (Graham 2010; Taylor en Broeders 2015). Evenzo bestaan er over de ene bevolkingsgroep meer gegevens dan over andere bevolkingsgroepen. Voor veel wereldwijde vragen zijn dan ook niet de juiste data beschikbaar. De Verenigde Naties beschrijft bijvoorbeeld in haar rapport *A world that counts*, dat gaat over het inzetten van Big Data voor duurzame ontwikkeling, dat voor de Millenium Development Goals slechts 60 procent van de data beschikbaar is die nodig zou zijn voor een betrouwbaar beeld (VN 2014: 12). Het is aannemelijk dat dit probleem zich ook in het veiligheidsdomein voordoet, temeer omdat sommige individuen, groepen en organisaties zich doelbewust aan registratie proberen te onttrekken, wat bestandskoppeling – ook internationaal – welhaast noodzakelijk maakt (AIV 2007).

De beschikbaarheid van data heeft daarnaast ook met bewaartermijnen te maken, die voor elke organisatie binnen het veiligheidsdomein weer anders liggen. Organisaties als de politie, de AIVD en de Belastingdienst kunnen data van andere organisaties opeisen. Dit is iets wat 'gewone' organisaties niet kunnen. Vooral om de beschikbaarheid van internet- en telefoongegevens voor de politie en veiligheidsdiensten te garanderen, is er een richtlijn die internet- en telefoniebedrijven verplicht de locatie- en verkeersgegevens van gebruikers voor een bepaalde tijd op te slaan. In 2014 is de deze richtlijn echter ongeldig verklaard door het Hof van

Justitie van de Europese Unie wegens een ernstige inmenging in de fundamentele rechten op eerbiediging van het privéleven en op bescherming van persoonsgegevens. Heldere afspraken over bewaartermijnen – voor zowel ruwe als bewerkte data en voor de resultaten van data-analyses, toegespitst per veiligheidsdomein – zijn aldus een belangrijke voorwaarde voor Big Data-analyses.

Tegelijkertijd zijn meer data niet altijd beter. Met kleine sets van kwalitatief hoogwaardige data zijn vaak betere resultaten te behalen. Deze sets bieden meer controle over het ontwerp van het onderzoek en maken de beantwoording van meer specifieke en toegespitste vragen mogelijk (Kitchin en Lauriault 2014). Het opbouwen van datasets van kwalitatief hoogwaardige data stuit echter op problemen wanneer data uit verschillende bronnen afkomstig zijn en in verschillende formats zijn opgeslagen (Kool et al. 2015: 23-24). Ze kunnen bovendien gefragmenteerd zijn over verschillende datasystemen, locaties en organisaties, waardoor het lastig is om er toegang toe te krijgen (WEF 2014: 44; Kool et al. 2015: 23). Zelfs één enkele organisatie kan nog de grootste moeite hebben om data uit verschillende organisatieonderdelen bijeen te voegen. De Economic Intelligence Unit (2012) beschouwt het integreren van verschillende datahuishoudingen dan ook als een van de belangrijkste uitdagingen voor het toepassen van Big Data-analyses.

4.4.2 EXPERTISE EN INBEDDING IN DE ORGANISATIE

Twee andere belangrijke voorwaarden voor het realiseren van de kansen van Big Data liggen op het niveau van de organisatie. Er is ten eerste voldoende expertise nodig om goede analyses te kunnen uitvoeren en ten tweede zullen de uitkomsten daarvan hun weg in de organisatie moeten vinden.

Expertise

Data zijn steeds gemakkelijker te verkrijgen en goedkoper op te slaan. Hierin wordt door bedrijven dan ook terecht veel geïnvesteerd. De investeringen in de analyse van die data zijn daar echter bij achtergebleven (WEF 2014; PWC 2013; EIU 2012). Het blijkt geen eenvoudige opgave om de expertise aan te boren die nodig is om relevante analyses uit te (laten) voeren. McKinsey Global Institute (2011) voorspelde al in 2011 dat in 2018 in de Verenigde Staten de vraag naar *data scientists* het aanbod met 140.000 tot 190.000 posities zou overstijgen. De Economist Intelligence Unit (2012) beschouwt het tekort aan geschoolde mensen als de op een na grootste belemmering voor het gebruik van Big Data. Ook recentere studies van het World Economic Forum (2014) en de OESO (2014) signaleren dit tekort.

Het woord *data scientist* wordt inmiddels veel gebruikt. De Harvard Business Review riep dat beroep in 2012 zelfs uit tot 'The Sexiest Job of the 21st Century' (Davenport en Patil 2012). Toch is niet geheel duidelijk wat hiermee bedoeld wordt. Volgens Davenport et al. (2012) gaat het om personen met zowel statistische kennis als kennis over programmeren. Ze kunnen dus meer dan de klassieke data-

analisten. Voor Big Data-analyses is bovendien een combinatie van technische en domeinkennis nodig om tot waardevolle toepassingen te komen. Ook data-gedreven analyses vereisen het vermogen om relevante kennisvragen te stellen (WEF 2014; McAfee en Brynjolfsson 2012). Het tekort aan 'data-savvy managers' – managers die data op een juiste manier kunnen inzetten – schijnt naar verwachting zelfs nog groter te zijn. De eerder aangehaalde studie van het McKinsey Global Institute (2011) rept van 1,5 miljoen posities in 2018.

Inbedding in de organisatie

Behalve expertise is ook een goede inbedding in de organisatie belangrijk om de potentie van Big Data ten volle te kunnen benutten (WEF 2014; OECD 2014; Sanders et al. 2015). Kort gezegd: de uitkomsten van Big Data-analyses dienen een rol te spelen in besluitvormingsprocedures. Bakhshi et al. (2014) constateren op basis van een analyse van 500 Britse bedrijven dat het grootschalig verzamelen van klantgegevens niet zonder meer resulteert in een positieve bijdrage aan de bedrijfsresultaten: "(F)irms need to introduce complementary changes in order to reap the full returns from their online data activity". Zij zullen hun organisatiestructuur en bedrijfsprocessen ingrijpend moeten hervormen om de uitkomsten van Big Data-analyses in klinkende munt om te zetten. Of in de woorden van Devlin et al. (2012: 20): "Technology is seldom a showstopper, but organizational issues often are".

Het hiervoor besproken CAS vertoont een vergelijkbare problematiek. Behalve de aanwezigheid van voldoende *data scientists* en een goede ontsluiting van data is belangrijk dat agenten zich ondersteund voelen door de informatie die het CAS oplevert. Ook is het nodig dat de bredere organisatie bereid is de eigen werkwijze aan te passen, bijvoorbeeld door surveillancerondes uit te voeren op tijden en plekken die volgens de analyses de meeste kans op misdaad hebben. Het hogere management moet dus het inhoudelijke belang van Big Data inzien (EIU 2012). Wanneer dit niet gebeurt, degradeert Big Data van een middel om intelligentere besluitvorming te realiseren naar een mechanisme om de huidige inzet van middelen te verantwoorden (EIU 2012: 23; WRR 2011). Een voorbeeld van dit laatste is een casestudie naar Intelligence Led Policing (ILP) in Canada, waar het gebruik van ILP onder de druk van dalende budgetten vooral de functie kreeg om de inzet van politiemiddelen te kunnen legitimeren (Sanders et al. 2015).

4.4.3 BEVEILIGING VAN DATA

De verspreiding van data over verschillende organisaties is gedeeltelijk te verhelpen door gebruik te maken van zogeheten *trusted data aggregators*. Hierbij is er één partij die de data van verschillende organisaties bijeenbrengt en zo de inzichten kan combineren. In de Verenigde Staten vervullen de zogenoemde commerciële *data brokers* samen deze rol, maar in Europa en Nederland is de markt voor dit soort partijen minder ontwikkeld. De afgelopen jaren zijn er binnen het veilig-

heidsdomein enkele organisaties opgericht die eenzelfde soort rol vervullen. Maar die organisaties opereren uitsluitend in opdracht van publieke diensten, die binnen datzelfde domein de taak en bevoegdheid hebben om gegevens te verzamelen en analyseren. Voorbeelden hiervan zijn het Inlichtingenbureau, dat onder meer analyses uitvoert voor het SyRI, en de iCOV, die alle data van onder andere het OM, de politie, de Belastingdienst, de Douane, de FIOD en de Financial Intelligence Unit in bezit heeft en analyseert op speciaal verzoek van de deelnemende organisaties. Het is daarnaast ook mogelijk om binnen een organisatie een speciaal datacentrum op te zetten waar gegevens uit de verschillende organisatieonderdelen wordt gecombineerd, wat onder meer gebeurt bij de Belastingdienst en de politie.

Met het verzamelen, opslaan, uitwisselen en koppelen van data wordt ook de beveiliging daarvan steeds belangrijker. Het risico van datalekken en beveiligingsincidenten neemt toe naarmate hierbij meer organisaties betrokken zijn. De beveiliging van data is net zo sterk als de zwakste schakel in de keten van gegevensuitwisseling. Ook goed beveiligde systemen kunnen echter op vele manieren aangevallen worden, door codes te kraken, hardware te manipuleren of de beveiliging te omzeilen door al dan niet doelbewust geïnstalleerde zwakheden in systemen en software uit te buiten (Gürses en Preneel 2016). Dat het hier geen theoretisch risico betreft, bewijzen de vele voorvallen van de laatste jaren. In de zomer van 2015 werden alle klantgegevens van Ashley Madison – een website voor mensen die op zoek zijn naar een affaire – gehackt en enkele weken later online gezet. In 2014 kregen hackers toegang tot de persoonsgegevens van 145 miljoen gebruikers van Ebay. In de winter van 2013/2014 werden de betaalgegevens van 40 miljoen klanten van Target – een grote Amerikaanse winkelketen – gehackt. Later werd bekend dat de aanvaller waarschijnlijk ook nog eens toegang had tot de persoonsgegevens van 70 miljoen klanten (Perez 2014).

Binnen het veiligheidsdomein is de gevoeligheid van data nog vele malen groter, omdat die gegevens bruikbaar zijn voor de meest uiteenlopende vormen van misdaad. Ook criminele netwerken is dit niet ontgaan. Zo voorziet Europol (2015) dat Big Data het aangezicht van de georganiseerde misdaad ingrijpend zal veranderen. Nu al zijn cybercriminelen in staat grote hoeveelheden data te vergaren door netwerken binnen te dringen en dataverkeer te onderscheppen. In de nabije toekomst zullen ook grote databases van overheidsorganisaties in toenemende mate het doelwit van georganiseerde misdaad worden, waardoor vooral identiteitsfraude een hoge vlucht zal nemen (Europol 2015: 19-20). Diefstal van biometrische data is interessant omdat biometrische gegevens constant zijn (een gestolen wachtwoord kun je vervangen, biometrische gegevens niet) en toegang kunnen bieden tot fysieke locaties en gevoelige informatie. De grote toename van data vergemakkelijkt sowieso illegale immigratie, mensenhandel en drugssmokkel. Criminelen

kunnen steeds rijkere en volledige persoonsgegevens verhandelen. Ook kunnen ze gebruikmaken van gegevens die naar alle waarschijnlijkheid niet binnen de risicocategorieën van opsporingsinstanties vallen.

Organisaties kunnen veel doen om hun data te beveiligen. Ze kunnen data verspreid opslaan, ze kunnen data classificeren naar gevoeligheid en op basis daarvan verschillende beveiligingsniveaus aanbrengen, en ze kunnen de toegang van (vooraf gescreende) medewerkers tot databestanden reguleren. Voor de beveiliging van creditcarddata geldt bijvoorbeeld een speciale industriestandaard. Op alle niveaus van data-uitwisseling, dataopslag en toegang daartoe is het mogelijk data te versleutelen. Daarvoor zijn zeer veel technieken beschikbaar (Gürses en Preneel 2016). Toch brengt de toegang van vaste en tijdelijke medewerkers tot vertrouwelijke gegevens risico's met zich mee, omdat compliance met geheimhoudingsbepalingen moeilijk te controleren is, gezien de enorme grootte van de organisatie. Dit is een probleem waarmee vele organisaties kampen, die per fase in de analyse van data vaak verschillende regimes kennen om de toegang tot informatie te reguleren. Bij het gebruik van steeds grotere databestanden en vanwege het groeiend aantal publieke én private partijen dat betrokken is bij Big Data-processen neemt het belang van adequate beveiligingsmaatregelen navenant toe.

4.5 RISICO'S

De inzet van Big Data binnen het veiligheidsdomein kent een aantal risico's. Deze risico's zullen geadresseerd moeten worden, wil het gebruik van Big Data maatschappelijk aanvaardbaar zijn. Sommige van deze risico's vloeien voort uit het niet adequaat omgaan met een aantal van de statistische beperkingen genoemd paragraaf 4.3. Andere hebben te maken met de verzameling en analyse van data: te veel data verzamelen en combineren heeft al snel implicaties voor de privacy van individuen en voor privacy als een maatschappelijke (collectieve) waarde. Ook kan de succesvolle introductie van een data-gedreven oplossing in het ene domein, er in de praktijk toe leiden dat deze ook in een ander domein wordt toegepast (waar ze niet voor bedoeld en ontwikkeld was) of dat er meer databronnen aan elkaar gekoppeld worden en meer organisaties toegang willen. Deze *function creep* zorgt ervoor dat we via allerlei incrementele en op zichzelf te verdedigen keuzes over het toevoegen van functies, data, koppelingen en deelnemers eindigen met systemen die als geheel, en in één keer, waarschijnlijk nooit goedgekeurd waren (WRR 2011). De laatste twee risico's hebben te maken met verschuivingen in de machtsbalans tussen overheid en burger als gevolg van de keuze voor Big Data-oplossingen. De kennis van de overheid over de eigen burgers kan – zeker in het domein van de veiligheidszorg – te groot worden en een intimiderend effect krijgen. Een andere zorg is dat de burgers zich geconfronteerd zien met de uitkomsten van Big Data-processen die voor hen onnavolgbaar zijn en lastig kunnen worden gecorrigeerd.

4.5.1 PROFILING EN DISCRIMINATIE

Data-gedreven oplossingen werken altijd toe naar *social sorting* (Lyon 2003). Dat gebeurt door categorieën in de data aan te brengen, profielen te maken en daar gevolgen aan te verbinden. Dit mondt uit in acties als wel of geen onderzoek wegens fraude, wel of niet een goedkope of juist een dure lening. Bij Big Data-processen gebeurt *social sorting* vaak op basis van correlaties die in de data te vinden zijn. Correlaties alleen zeggen uiteraard nog niets over causaliteit en zijn dus een wankele basis om conclusies aan te verbinden (zie paragaaf 4.3). De afbakening van groepen op basis van Big Data-analyses kan problematisch zijn als de bias die in elke dataset zit en de bias die zich in algoritmen verbergt niet goed geadresseerd worden. Als de data worden gezien als een perfecte afspiegeling van een bepaalde groep, terwijl deze dat niet zijn, zullen de conclusies die daaruit getrokken worden ook niet perfect passen op deze groep. De bias kan zich dan reproduceren, resulterend in een toename van sociale stratificatie en discriminatie (Gandy 1993; Lyon 2003; Zarsky 2014). Die discriminatie kan zich bovendien vertalen in een cumulatief nadeel voor bepaalde groepen in de maatschappij (Gandy 2009). In het uiterste geval kunnen Big Data-scoringmethoden leiden tot datadeterminisme. In dat geval worden individuen beoordeeld op basis van wat probabilistische kennis (correlaties en inferenties) aangeeft dat ze misschien zullen doen, in plaats van op basis van de dingen die ze echt hebben gedaan (Zarsky 2014). In dit extreme geval zou dat ook afbreuk doen aan de idee van de mens als een autonoom moreel wezen dat in staat is te veranderen en andere keuzes te maken dan die in het verleden (Custers et al. 2013). 'Eens een dief, altijd een dief' is geen onderdeel van ons rechtsstelsel.

Dit is geen denkbeeldig risico. Het ontstaan van cumulatieve voor- en nadelen is bijvoorbeeld waarneembaar bij het veel bestudeerde systeem van kredietbeoordeling in de Verenigde Staten (zie bijvoorbeeld Citron en Pasquale 2014; Pasquale 2015). Iemand die goed uit het proces van *credit rating* komt, krijgt een lening of een hypotheek en kan verder aan zijn of haar financiële toekomst bouwen. Iemand met een negatief resultaat krijgt niets, of ontvangt een lening met striktere aflossingsbepalingen, waardoor mogelijk problemen ontstaan bij de terugbetaling, wat de kans op een (tweede) slechte kredietbeoordeling vergroot. Het White House report over Big Data waarschuwt ook voor dit risico: "The increasing use of algorithms to make eligibility decisions must be carefully monitored for potential discriminatory outcomes for disadvantaged groups, even absent discriminatory intent" (White House 2014: 47). Het feit dat discriminatie niet intentioneel hoeft te zijn maar soms in een correlatie, dataset of algoritme 'verscholen' zit, maakt het alleen maar lastiger om deze op te sporen. Dit mechanisme kan op het niveau van personen spelen, maar ook op het niveau van wijken, zoals bijvoorbeeld bij *predictive policing*. Een wijk waarin veel gesurveilleerd wordt, zal prominenter terugkomen in de criminaliteitscijfers. De extra aandacht vergroot de bestaande problemen verder uit, hetgeen weer de basis voor nieuw beleid kan zijn, dat op zijn beurt het (negatieve) beeld verder versterkt. En een individu dat op basis van een

risicoscore intensief gevolgd wordt, zal uiteraard eerder op een strafbaar feit worden betrapt dan iemand die niet op die lijst staat. Aangezien de discriminatie in veel gevallen niet intentioneel is en niet met opzet in het algoritme ingeschreven wordt door de computerprogrammeurs, zal het zeer moeilijk zijn te achterhalen wie verantwoordelijk is voor het probleem en om dit te bewijzen in een rechtszaak. Maar zelfs in de gevallen waar de bias intentioneel is, bestaat het risico dat de technologie als verantwoordelijke aangewezen wordt.

4.5.2 SCHENDING VAN INDIVIDUELE EN COLLECTIEVE PRIVACY

Privacy en Big Data zijn niet zomaar met elkaar te verenigen. Privacyschendingen zijn een groot risico in het toepassen van Big Data-processen. Het is niet voor niets dat privacy een van de kernelementen in de adviesaanvraag van de regering is. Het risico van privacyschendingen is allereerst groot omdat de combinatie van volume en variëteit van data maakt dat Big Data-processen drijven op een overvloed van informatie, die op verschillende manieren aan personen gekoppeld of te koppelen is. Spanningen met privacy en het gegevensbeschermingsrecht zijn daardoor nooit ver weg. Daarnaast zijn voor de ontwikkeling van profielen altijd meer persoonsgegevens nodig dan in de feitelijke toepassing daarvan, die immers op een specifieke groep of zelfs een individu gericht kan zijn (De Hert en Lammerant 2016). Ook secundair gebruik – de grote kracht van Big Data – staat op gespannen voet met het beginsel van doelbinding, dat zegt dat informatie alleen mag worden gebruikt ten behoeve van welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden. Dit is een hoeksteen van de gegevensbescherming en privacywetgeving.

De overheid maakt op dit moment – ook in het domein van de veiligheidszorg – nog voornamelijk gebruik van eigen data. Het is bekend dat het combineren van verschillende data van verschillende overheden al duidelijke privacyrisico's mee zich meebrengt (WRR 2011). Het toenemend gebruik van data uit openbare bronnen en de private sfeer vergroot deze risico's niet alleen, maar roept ook nieuwe vraagstukken in het kader van privacy op. Privacy staat in de informatiesamenleving uiteraard al langer onder druk van zowel bedrijven als van overheden en Big Data is een intensivering van die druk. Op een aantal punten staan het karakter en de *unique selling points* van Big Data bijna diametraal tegenover de huidige maatschappelijke en juridische normen over privacy en gegevensbescherming.

Gezien de massaliteit van Big Data-processen en daarmee de massaliteit van de individuele schendingen van privacy, is er nog een element van privacy dat de aandacht verdient. Big Data-processen veroorzaken eigenlijk een dagelijkse massa van kleine individuele privacyschendingen door de verzameling en de analyse van grote hoeveelheden van (persoonlijke) data. Eigenlijk gaat het om miljoenen speldenprikjes, verdeeld over miljoenen mensen. De gangbare manier om daar naar te kijken, is deze speldenprikjes te zien als schendingen van het fundamentele recht

op privacy van een individueel persoon. Dit is het principe van *individual harm* dat in het huidige recht vooropstaat. Het wordt dagelijks massaal geschonden, maar zelden in die mate dat mensen ervoor naar de rechter stappen. Alle speldenprikken bij elkaar opgeteld is de schending echter zonder meer groot. Er gaan dan ook stemmen op om privacy ook als een collectief recht te definiëren (zie bijvoorbeeld Van der Sloot 2016a; Taylor et al. 2016), waarbij het niet gaat om de individuele schade van personen (paar prikjes per persoon) maar om schade aan het fundamentele recht zelf (alle prikken van een programma bij elkaar). Hoofdstuk 5 gaat uitgebreid in op de regulering van privacy en databescherming in het licht van de nieuwe mogelijkheden van Big Data.

4.5.3 FUNCTION CREEP

Het feit dat sommige Big Data-toepassingen succesvol zijn, kan risico's met zich meebrengen. Wat werkt – in functionele zin – wordt vaak uitgebreid of doorvertaald naar andere contexten. Soms is dat een uitstekend idee, maar soms worden daarbij data, organisaties en functies samengevoegd die vanwege goede (juridische) redenen voorheen gescheiden waren. Als dat laatste gebeurt noemen we het *function creep* of *surveillance creep*. Vaak gaat het om de verleiding die ontstaat als systemen eenmaal in gebruik zijn genomen voor een bepaald doel. Als er camera's boven de snelweg hangen om de doorstroom te monitoren en het verkeer in goede banen te leiden – met als functie verkeersveiligheid – dan kunnen die wellicht toch ook gebruikt worden om openstaande boetes te innen of gestolen auto's op te sporen – met als functie toezicht en opsporing (Griffioen 2011). Lang voordat de Nederlandse en Europese burgers vingerafdrukken moesten afgeven voor identificatie en het biometrische EU-paspoort werd biometrische identificatie uitgerold voor asielzoekers en (illegale) vreemdelingen op nationaal en Europees niveau (Broeders 2007; 2009). Bovendien duurde het niet lang voordat veiligheidsdiensten en opsporingsinstanties zich meldden om toegang te krijgen tot deze grootschalige biometrische datasystemen die in het kader van het immigratiebeleid waren opgezet (Balzacq 2008; Broeders 2011).

Tot op zekere hoogte ligt *function creep* al in de logica van Big Data-analyses besloten. Het idee is immers dat meer data, uit verschillende bronnen de beste informatie opleveren. Bovendien gaat men ervan uit dat secundair gebruik van data – oftewel gebruik anders dan het doel waarvoor de data zijn verzameld – een grote meerwaarde oplevert. Secundair gebruik kan men op zichzelf al zien als een vorm van *function creep*: de data worden immers voor iets anders ingezet dan voor de functie waarvoor ze verzameld zijn. Uiteraard wordt het risico bepaald door het soort data dat gebruikt wordt en door de gevolgen die aan die analyse verbonden worden. Het is van belang om dit risico serieus te nemen, zeker als het gaat om plannen van de overheid in het domein van de veiligheidszorg. *Function creep* gaat in de regel in kleine incrementele stapjes: er wordt een koppeling gemaakt tussen het ene en het andere systeem, er komt toegang voor een nieuwe organisatie tot

deze data, enzovoort. Voor elk van deze stapjes is op dat specifieke moment een politieke rechtvaardiging te geven – effectiever, betere dienstverlening, veiliger – maar het cumulatieve resultaat is vaak veel groter dan de som der delen (WRR 2011). Bovendien is het 'eindresultaat' iets waar nooit een serieus politiek debat over is gevoerd. In een Big Data-wereld zal dat zich alleen maar versterken.

4.5.4 CHILLING EFFECTS

De grootschalige verzameling en opslag van data door private partijen en overheden, niet in de laatste plaats door inlichtingen- en veiligheidsdiensten, hebben bredere maatschappelijke effecten. De dooddoener 'wie niets te verbergen heeft, heeft niets te vrezen' kan niet verhelpen dat er een negatief sociaal effect uitgaat van het constante (elektronische) toezicht door publieke en private partijen in het digitale tijdperk. Wanneer overheidsorganisaties ook private data voor veiligheidsdoeleinden gaan gebruiken, zal dit effect naar alle waarschijnlijkheid sterk worden uitvergroot. Met name in de (machts)relatie tussen burgers en de overheid geeft dit een aanzet tot wat bekend staat als *chilling effects* op het genieten en uitoefenen van bepaalde rechten. Mensen kunnen het gevoel krijgen dat hun recht op privacy en vrijheid van meningsuiting in gevaar is. Als deze effecten optreden bij journalisten, schrijvers, klokkenluiders, NGO's en advocaten, komt ook het functioneren van de bredere democratie in het geding. Harvard-hoogleraar Internationale Betrekkingen Stephen Walt vond de individuele privacyschendingen door de NSA dan ook minder belangrijk dan het effect dat surveillance op deze schaal en scherpte kan hebben op mensen en organisaties die tegengeluiden laten horen en vraagtekens plaatsen bij de keuzes van de overheid:

> "In short, the real reason you should be worried about these revelations of government surveillance is not that you're likely to be tracked, prosecuted, or exposed. You should be worried because it is another step in the process of making our vibrant, contentious, and most of all free-minded citizenry into a nation of sheep." (Walt 2013)

In de VS houden journalisten en advocaten dagelijks rekening met de surveillanceactiviteiten van de Amerikaanse overheid en vinden het steeds moeilijker om hun – soms bij wet beschermde – werk goed te doen (Human Rights Watch 2014). Ook in Nederland speelt dit. Zo behoudt de AIVD zich het recht voor om in bijzondere gevallen vertrouwelijke gesprekken tussen advocaten en hun cliënten af te luisteren. Advocatenkantoor Prakken d'Oliveira daagde de staat hiervoor in de zomer van 2015 voor de rechter. De staat verloor, maar ging tevens in beroep tegen de uitspraak van de voorzieningenrechter. Een identieke zorg speelt rond de herziening van de Wiv, die grootschalige ongerichte dataverzameling mogelijk maakt. In reactie op dit soort praktijken kunnen ook 'gewone' mensen zich als gevolg van deze continue dataverzameling op een andere manier gaan gedragen. Sommigen zullen hun gedrag aanpassen om zo normaal mogelijk te lijken, anderen zullen wellicht zoveel mogelijk buiten beeld proberen te geraken. In beide gevallen wordt hun vrijheid ingeperkt.

4.5.5 MACHTSEVENWICHT EN EEN TRANSPARANTIEPARADOX

Informatie is macht. Veel van de gereedschappen van Big Data zijn gedemocratiseerd, maar de data zelf zijn dat niet (Mayer-Schönberger en Cukier 2013: 16). Veel van onze data liggen opgeslagen bij enkele grote bedrijven, waartoe wij weinig toegang hebben. Dit is behalve vanuit economisch opzicht (bedrijven verdienen hun geld met onze data) ook relevant vanuit een democratisch perspectief, omdat het evenwicht tussen staat en burger, tussen bedrijven en burgers en bedrijven onderling aan verandering onderhevig is. Er kan hier een transparantieparadox worden waargenomen (Richards en King 2013): enerzijds worden burgers steeds transparanter voor overheid en bedrijfsleven, anderzijds zijn deze organisaties en de manier waarop zij analyses toepassen niet zo open voor burgers. Profielen en algoritmen zijn nauwelijks transparant of navolgbaar. Ook de verhouding tussen het dataproces en de acties en besluiten die op basis daarvan worden genomen, is moeilijk te beoordelen en controleren. Onderzoekers spreken van een *Black Box Society*, waarin systemen belangrijke besluiten nemen over het leven van 'gewone' mensen (Pasquale 2015; Schneier 2015; Gillespie 2014). Deze systemen zijn moeilijk te doorgronden en begrijpen, soms zelfs voor de mensen die er direct mee werken. Transparantie – het openen van de black box van Big Data – is daarom een veelgehoord pleidooi als het gaat om het gebruik van Big Data-processen (Hildebrandt en De Vries 2013; Zarsky 2016).

Transparantie is ook buitengewoon relevant voor het veiligheidsdomein. In de eerste plaats omdat organisaties die zich bezighouden met de (nationale) veiligheid uit dien hoofde meer bevoegdheden hebben gekregen en meer recht hebben om hun activiteiten af te schermen. Juist de overheidsorganisaties die op het terrein van (nationale) veiligheid werken mogen de meeste informatie verzamelen, daarbij de meeste inbreuk maken op fundamentele rechten en zij mogen die activiteiten het meest afschermen van de buitenwereld (geheimhouding). Big Data vergroot de mogelijkheden voor deze organisaties om data te verzamelen en te koppelen en om dieper in het leven van burgers te kijken. De vraag is of en in hoeverre dat gevolgen moet hebben voor (het gebrek aan) transparantie over hun werkzaamheden. Het is deze zorg die ten grondslag ligt aan de tweede hoofdvraag van de adviesaanvraag: hoe bij het gebruik van Big Data ervoor kan worden gezorgd dat het proces van profiling, datamining en andere analysetechnieken ten behoeve van de veiligheid voldoende transparant is (zie bijlage 1).

In de tweede plaats geldt dat in het domein van de veiligheidszorg de data vaak uit menselijk gedrag en intenties worden gedestilleerd. Veiligheid is een sociaal fenomeen en dat betekent dat de data vaak 'zacht' zijn – naast hardere data zoals financiële transacties en DNA – en dat de mogelijkheden voor vals positieven en vals negatieven groter zijn. Daarmee wordt geheimhouding van de ins en outs van Big Data-analyses in maatschappelijke zin ook problematischer. De geheimhouding kan Kafkaësk worden. Het probleem wordt dat van een "bureaucracy with

inscrutable purposes that uses people's information to make important decisions about them, yet denies the people the ability to participate in how their information is used" (Solove 2007: 756). Geheimhouding maakt de burger transparant en de overheid die over hem oordeelt wordt juist ondoorzichtig. Juist in het domein van de veiligheidszorg – waar de consequenties van fouten groot kunnen zijn – is het van belang dit risico te beperken.

4.6 CONCLUSIE

Big Data heeft de potentie om het veiligheidsdomein ingrijpend te veranderen. Door een betere informatievoorziening is het mogelijk de bestaande taken vele malen efficiënter uit te voeren. Belangrijker is dat Big Data organisaties in staat stelt om overtredingen en misdrijven sneller op het spoor te komen en daar preciezere reconstructies van te vervaardigen. Bovendien maakt Big Data betere voorspellingen mogelijk van potentieel risicovolle situaties, personen en netwerken. De belofte van Big Data voor het veiligheidsdomein is daarbij breder dan de huidige, overwegend op repressie gerichte doeleinden van Big Data-analyses. Het is tevens mogelijk burgers te helpen om beslissingen te nemen, die ten goede komen aan de individuele en maatschappelijke veiligheid.

Tegelijkertijd is de inzet van Big Data aan beperkingen gebonden. Big Data-analyses kennen deels dezelfde tekortkomingen als hun voorlopers in de statistiek, zoals bias, schijncorrelaties en het voorkomen van vals positieven en vals negatieven. Binnen het veiligheidsdomein, waar de gevolgen van overheidsinterventie voor burgers zeer ingrijpend kunnen zijn, is daarom een zeer terughoudende en prudente omgang met Big Data vereist. Een tweede beperking ligt in de aard van de betrokken veiligheidsvraagstukken. Sommige vormen van misdaad zijn eenvoudigweg te onregelmatig en incidenteel om er betrouwbare voorspellingen op te baseren. Big Data-toepassingen hebben de potentie om het bredere palet aan onderzoeks- en opsporingsmethoden ingrijpend te transformeren. Vanwege de inherente beperkingen van die toepassingen zullen zij dit palet echter nooit geheel kunnen vervangen. Het daadwerkelijk inlossen van de beloftes van Big Data vereist bovendien de nodige expertise en aanpassing in de organisatie. Vooralsnog ontbreekt het veel organisaties aan voldoende expertise, zowel op het vlak van de analyse zelf als op het niveau van de bestuurslagen daarboven, waar de uitkomsten idealiter in de besluitvorming worden geïntegreerd. De inbedding van Big Data-analyses in de bestaande organisatiestructuren en -routines blijkt echter een lastige opgave.

Ten slotte kent het gebruik van Big Data in het veiligheidsdomein ook enkele risico's, die van invloed zijn op de maatschappelijke aanvaardbaarheid van Big Data-toepassingen. Zonder voldoende besef van de bias in de data en algoritmen kunnen de gebruikte datacategorieën en profielen tot een toename van sociale

stratificatie en uitsluiting leiden. Het credo 'eerst verzamelen en daarna pas toepassingen verzinnen' heeft daarnaast al snel implicaties voor de privacy van individuen en voor privacy als maatschappelijke waarde. Ook tendeert Big Data naar *function creep*, juist omdat veel toepassingen pas later in het proces ontstaan, en dus plaatsvinden zonder toe- en/of instemming van de personen over wie de data zijn verzameld. Het gebruik van Big Data gaat bovendien gepaard met een verschuiving in het machtsevenwicht tussen overheid en burgers, omdat overheidsorganisaties steeds dieper in het leven van burgers kunnen kijken. Door grootschalige data-analyses kunnen burgers het gevoel krijgen dat zij permanent gevolgd worden, met als consequentie dat zij hun gedrag aanpassen, iets dat grote gevolgen heeft voor de vrijheid van burgers. Ten slotte neemt de mate van geheimhouding toe: terwijl overheden steeds meer over burgers te weten komen, blijft het hoe en waarom van Big Data-analyses en de besluiten die daarop worden gebaseerd voor diezelfde burgers een schimmige aangelegenheid. Het nu volgende hoofdstuk bespreekt in hoeverre de huidige juridische kaders de ruimte bieden om de belofte van Big Data te realiseren en de bijbehorende risico's adequaat te mitigeren.

5 BIG DATA, VEILIGHEID EN DE JURIDISCHE KADERS VOOR GEGEVENSVERWERKING

5.1 INLEIDING

De gegevensverwerking binnen het veiligheidsdomein is aan verschillende juridische kaders gebonden. Een deel van deze kaders geldt alleen specifiek voor enkele aparte onderdelen van het veiligheidsdomein. Zo kennen politie en justitie hun eigen regime voor gegevensverwerking, net als de inlichtingen- en veiligheidsdiensten. Andere organisaties in het veiligheidsdomein, zoals inspecties, de Belastingdienst en samenwerkingsorganen op het terrein van misdaad- en fraudebestrijding, vallen onder het algemene kader van de Wet bescherming persoonsgegevens (Wbp). Deze juridische kaders geven de betrokken organisaties de ruimte om gegevens te verzamelen die nodig zijn voor de uitvoering van hun taken binnen het veiligheidsdomein. Voor de inlichtingen- en veiligheidsdiensten is die ruimte aanzienlijk groter dan voor de organisaties die actief zijn op het terrein van de fraudebestrijding. De juridische inkadering richt zich niet alleen op organisaties op veiligheidsterrein, maar ook op burgers. De wetgeving heeft dan ook mede tot doel burgers te vrijwaren van te indringende, ongecontroleerde en/of arbitraire handelingen van de overheid (of andere machtige partijen). Hierbij geldt dat de waarborgen groter zijn naarmate overheidsorganisaties dieper kunnen ingrijpen in de rechten en vrijheden van burgers.

Big Data biedt organisaties binnen het veiligheidsdomein nieuwe mogelijkheden op het terrein van surveillance, toezicht en opsporing. Deze mogelijkheden zetten op verschillende manieren druk op de huidige juridische kaders. Een eerste vraag is in hoeverre het verzamelen van gegevens in het Big Data-tijdperk nog een adequaat aangrijpingspunt vormt voor juridische waarborgen om de vrijheden van burgers te beschermen (Koops 2013a; Hildebrandt 2015; Van der Sloot 2016a; Zarsky 2016). Met de huidige beschikbaarheid van grote hoeveelheden data is het niet ongewoon om eerst uitgebreid gegevens te verzamelen (of de wetenschap te koesteren dat anderen ze reeds verzamelen) en pas later een schifting te maken in gegevens die bruikbaar en onbruikbaar zijn. Bovendien kan de combinatie van meerdere databases met niet-identificerende gegevens door koppeling en analyse leiden tot profielen en nieuwe persoonsgegevens (op basis van een combinatie van informatiepunten). Belangrijke juridische principes als doelbinding en dataminimalisatie staan daarom op gespannen voet met het adagium van Big Data dat ‘meer beter is’. De waarde van Big Data-analyses zit immers in het secundair gebruik van data en het vinden van onverwachte – en dus niet per definitie aan een vooraf vastgesteld verzameldoel gebonden – toepassingen.

Een tweede vraag is in hoeverre de juridische kaders ruimte bieden voor het gebruik van Big Data-toepassingen. Daarbij gaat het erom of deze toepassingen voldoende zijn ingebed in de bestaande wet- en regelgeving. Als dat niet zo is, dan ontbeert het gebruik daarvan legitimiteit. Ook speelt hierbij de vraag of de technologische vooruitgang bepaalde wettelijke bevoegdheden de facto niet dusdanig oprekt, dat er opnieuw gekeken moet worden of de wetgeving nog aan haar doel beantwoordt. De beschikbare hoeveelheid data, de toegenomen analysetechnieken en de afgenomen kosten om een persoon en zijn sociale, professionele en eventueel criminele netwerken digitaal te volgen, geven de bestaande bevoegdheden voor onderzoek en vervolging een grote impuls. Dit geldt eveneens voor de rijkdom aan informatie die een mobiele telefoon of een computer over iemand kan onthullen. Wanneer bevoegdheden breed geformuleerd zijn, kan technologische vooruitgang deze oprekken voorbij de originele intentie van de wetgever. Of, zoals de Amerikaanse jurist Kerr (2011) het formuleert: "When new technologies expand law enforcement's capabilities, the law does (and should) respond by placing new limits on the government; when new technologies give criminals a leg up, the law does (and should) respond by loosening the government's reigns". Een soortgelijke dynamiek speelt ook bij de technieken waarmee Big Data werkt.

5.2 JURIDISCHE KADERS VOOR GEGEVENSVERWERKING BINNEN HET VEILIGHEIDSDOMEIN

Er zijn drie verschillende regimes die de belangrijkste organisaties in het veiligheidsdomein aansturen en inkaderen: die voor politie en justitie, die voor de inlichtingen- en veiligheidsdiensten en die voor andere overheidsinstanties, die onder de werking van de Wbp vallen.[1] Alle drie de regimes vertrekken vanuit de fundamentele rechten zoals vastgelegd in internationale verdragen, het Handvest van de grondrechten van de Europese Unie en de Grondwet. Figuur 5.1 biedt een schematische weergave van deze kaders.

Figuur 5.1 Juridische kaders voor gegevensverwerking binnen het veiligheidsdomein

5.2.1 HET GRONDRECHTELIJK KADER

Overheidsingrijpen in het kader van de (nationale) veiligheid kan burgerrechten op diverse manieren raken. Rechten die in het geding kunnen zijn, zijn onder andere het recht op vrijheid van meningsuiting, het recht op gelijke behandeling, het recht op een eerlijk proces en het recht op privacy. Deze rechten zijn vastgelegd in onder andere de Universele Verklaring van de Rechten van de Mens (UVRM), het Internationaal Verdrag inzake burgerrechten en politieke rechten (IVBPR), het Europees Verdrag tot bescherming van de rechten van de mens en de fundamentele vrijheden (EVRM), het Handvest van de grondrechten van de Europese Unie en de Nederlandse Grondwet. Voor de Nederlandse situatie zijn deze laatste drie in de praktijk het meest relevant.

Uitgangspunt van deze verdragen en de Grondwet is dat de overheid niet mag ingrijpen in de grondrechten van de burgers, tenzij dit noodzakelijk is ter bescherming van de nationale veiligheid, de opsporing van strafbare feiten en de bescherming van de openbare orde. Dergelijke beperkingen moeten bij wet zijn voorzien. In relatie tot Big Data en veiligheid is het recht op privacy (de bescherming van de persoonlijke levenssfeer) het meest relevant, waaronder ook de bescherming van persoonsgegevens valt (de informationele privacy). Uitgangspunt bij Big Data en privacy is dus dat de overheid geen persoonsgegevens voor veiligheidsdoeleinden mag verwerken wanneer dit een meer dan geringe inbreuk op de privacy oplevert, tenzij dit noodzakelijk is en wettelijk is geregeld.

Naast de bescherming van de persoonlijke levenssfeer is ook het verbod op discriminatie belangrijk. Aan de opslag en verwerking van gevoelige persoonsgegevens als 'geslacht, ras, kleur, taal, godsdienst, of politieke mening' zijn zwaardere eisen verbonden, omdat deze gegevens in het gebruik discriminatie tot gevolg kunnen hebben. Het gebruik van deze gegevens valt onder de reikwijdte van het in verdragen en Grondwet verankerde verbod van discriminatie. Alleen binnen wettelijke kaders mogen onder bepaalde voorwaarden gevoelige gegevens over personen worden vergaard. Wanneer de gegevensverzameling gebeurt in het kader van toezicht of in het kader van verkennende vooronderzoeken, raakt dit de persoonlijke levenssfeer van de betrokkenen. Er kan dan bovendien spanning ontstaan met de presumptie van onschuld. Dit is uiteraard zeer relevant in het licht van Big Data-processen, die via datamining en profiling nieuwe verbanden met persoonsgegevens trachten te ontdekken.

EVRM en Handvest van de grondrechten van de Europese Unie

In Nederland gelden het Europees Verdrag tot bescherming van de rechten van de mens en de fundamentele vrijheden (EVRM art. 8), de jurisprudentie daarover en het Handvest van de grondrechten van de Europese Unie (art. 8) als belangrijkste kaders voor de regulering van de verwerking van persoonsgegevens in het domein van openbare orde en veiligheid. Het uitgangspunt is dat gegevensverwerking

potentieel een ingreep vormt in de persoonlijke levenssfeer en daarom gereguleerd moet worden. Deze invalshoek is ook verankerd in de Nederlandse Grondwet (art. 10 lid 2): "De wet stelt regels ter bescherming van de persoonlijke levenssfeer in verband met het vastleggen en verstrekken van persoonsgegevens." Het derde lid van artikel 10 omschrijft een aantal rechten van de geregistreerde (de 'betrokkene'), zoals het recht kennis te nemen van de over hem of haar vastgelegde gegevens en het gebruik dat daarvan wordt gemaakt.

Artikel 8 EVRM beschrijft het recht op privacy en luidt:

"1. Eenieder heeft recht op respect voor zijn privéleven, zijn familie- en gezinsleven, zijn woning en zijn correspondentie.
2. Geen inmenging van enig openbaar gezag is toegestaan in de uitoefening van dit recht, dan voor zover bij de wet is voorzien en in een democratische samenleving noodzakelijk is in het belang van de nationale veiligheid, de openbare veiligheid of het economisch welzijn van het land, het voorkomen van wanordelijkheden en strafbare feiten, de bescherming van de gezondheid of de goede zeden of voor de bescherming van de rechten en vrijheden van anderen."

Het Handvest van de grondrechten van de Europese Unie spreekt in artikel 8 specifiek over de bescherming van persoonsgegevens. Het artikel luidt:

"1. Eenieder heeft recht op bescherming van de hem betreffende persoonsgegevens.
2. Deze gegevens moeten eerlijk worden verwerkt, voor bepaalde doeleinden en met toestemming van de betrokkene of op basis van een andere gerechtvaardigde grondslag waarin de wet voorziet. Eenieder heeft recht op toegang tot de over hem verzamelde gegevens en op rectificatie daarvan.
3. Een onafhankelijke autoriteit ziet toe op de naleving van deze regels."

In deze structurering is de invloed van het Duitse concept van de *informationelle Selbstbestimmung* te herkennen (Roßnagel en Richter 2016), omdat zij terugvoert op het uitgangspunt dat individuen zelf over de verstrekking en het gebruik van persoonsgegevens zouden moeten kunnen beschikken. Principes als het vereiste van een legitiem doel en het recht om geïnformeerd te worden, liggen dus op dit hoge niveau vast.

Binnen de Europese Unie zijn de lidstaten zelf exclusief verantwoordelijk voor de zogeheten nationale veiligheid (art. 4 lid 2 Verdrag betreffende de Europese Unie). Dit begrip nationale veiligheid valt bij benadering samen met de taken van de inlichtingen- en veiligheidsdiensten en van defensie. Andere onderwerpen van veiligheid vallen onder de EU-taken in het kader van de ruimte van vrijheid, veiligheid en recht. Dit is geregeld in Titel V van het Derde Deel van het Verdrag betreffende de werking van de Europese Unie. Het op die titel gebaseerde Kaderbesluit 2008/977/JBZ van de Raad van de Europese Unie over de bescherming van persoonsgegevens die worden verwerkt in het kader van de politiële en justitiële samenwerking in strafzaken kent in artikel 1 lid 4 eveneens een uitzondering voor het domein van de nationale veiligheid in deze strikte betekenis. Artikel 3 lid 2 van

de Europese privacyrichtlijn (Richtlijn 95/46 EG) bepaalt dat deze richtlijn niet van toepassing is op de materie van de ruimte van vrijheid, veiligheid en recht en evenmin op die van de nationale veiligheid (daar 'veiligheid van de Staat' genoemd).

Op basis van dit grondrechtelijk kader moeten ingrepen in de persoonlijke levenssfeer die meer dan gering zijn, bij wet zijn geregeld.[2] Nederland kent daarom diverse wetten die beschrijven onder welke omstandigheden de overheid beperkingen mag stellen aan de grondrechten van burgers (bijvoorbeeld het Wetboek van Strafvordering). In zoverre hierbij persoonsgegevens worden verwerkt, moeten deze zorgvuldig worden verwerkt. Daarom zijn in de wet ook concrete regels gesteld over de omgang met persoonsgegevens. Hierna bespreken we kort de voornaamste juridische kaders voor de bescherming van persoonsgegevens in het veiligheidsdomein aan de hand van de instanties waarvoor zij gelden.

5.2.2 POLITIE EN JUSTITIE

In het kader van haar taakuitvoering verzamelt en verwerkt de politie persoonsgegevens (politiegegevens). De politie kan gegevens verzamelen ter uitvoering van haar dagelijkse politietaak – de "handhaving van de rechtsorde en het verlenen van hulp aan hen die deze behoeven" (art. 3 Politiewet) – en in het kader van strafrechtelijke onderzoeken. Wanneer gegevens in het kader van een strafrechtelijk onderzoek worden verzameld met behulp van (bijzondere) opsporingsbevoegdheden, dan zijn daar de specifieke bepalingen uit het Wetboek van Strafvordering (WvSv) op van toepassing.

De persoonsgegevens die de politie verzamelt zijn vaak privacygevoelig en de personen hebben voor het gebruik daarvan geen toestemming gegeven. Samen met de bijzondere verhouding tussen de overheid en burger – de politie kan geweld gebruiken en vrijheid ontnemen – resulteert dit in specifieke wetgeving voor het gebruik en de analyse van politiegegevens. De Wet politiegegevens (Wpg) regelt de bescherming van persoonsgegevens bij de Nationale Politie, de bijzondere opsporingsdiensten, de Koninklijke Marechaussee en de Rijksrecherche, en is tevens van toepassing op andere taken die de politie uitvoert voor justitie, zoals de Vreemdelingenwet.

Politiegegevens mogen volgens de Wpg (art. 8-10) worden verzameld en verder verwerkt wanneer dat noodzakelijk is voor:

- de uitvoering van de dagelijkse politietaak;
- onderzoek in verband met de handhaving van de rechtsorde in een bepaald geval;
- het verkrijgen van inzicht in de betrokkenheid van personen bij bepaalde ernstige bedreigingen van de rechtsorde.

Politiediensten richten zich bij het verzamelen van gegevens vooral op het verkrijgen van een voor de goede uitoefening van de politietaak benodigde informatiepositie en het verzamelen van bewijs in het kader van een onderzoek naar een strafbaar feit.

Eenmaal verzamelde data kunnen binnen de politie (en de Koninklijke Marechaussee) relatief breed worden gedeeld, omdat de wet uitgaat van een systeem van 'free flow of information' (Staatscourant 2011, nr. 16405: 8). Politiegegevens mogen worden gebruikt binnen de politieorganisatie, mits zij bijdragen aan een goede uitvoering van de politietaak. Bovendien mogen de gegevens die onder art. 8-10 Wpg zijn verzameld, geautomatiseerd met elkaar worden vergeleken om naar verbanden te zoeken. Wanneer deze verbanden inderdaad blijken te bestaan, is verdere verwerking mogelijk (art. 11 Wpg). Deze grondslag voor de verwerking kan onder andere worden gebruikt voor het in kaart brengen van netwerken van personen die in georganiseerd verband misdrijven beramen of plegen die een ernstige inbreuk op de rechtsorde opleveren en/of betrokken zijn bij ernstige schendingen van de openbare orde. Juist op dit punt liggen er kansen voor Big Data-analyses waarmee effectiever opgespoord kan worden. Maar er ontstaan ook nieuwe risico's van inbreuken op de rechten van burgers als gevolg van deze analyses. De fase van uitwisseling, koppeling en analyse van data is echter minder streng gereguleerd in vergelijking met de regels omtrent de verzameling van gegevens (in het WvSv). Zo mogen gegevens betreffende onderzoeken gecombineerd worden wanneer de bevoegde functionaris (de leider van het onderzoek binnen de politie) dit goedkeurt. Wanneer bijzondere opsporingsbevoegdheden zijn gebruikt moet daarnaast de officier van justitie toestemming geven (art. 126dd WvSv). Van een onafhankelijke rechterlijke toetsing door de rechter-commissaris is echter geen sprake. Dit terwijl Big Data-analyses mogelijk wel nieuwe risico's opleveren voor de privacy van betrokkenen.

Naast het 'klassieke' opsporingsonderzoek kan de politie ook 'verkennend onderzoek' doen (art. 126gg e.v. WvSv). Hierbij mag aan de hand van gegevens uit open bronnen of op basis van vrijwillige medewerking van particulieren, een veel grotere en lossere verzameling van personen worden onderzocht. Doel van zo'n onderzoek is de voorbereiding van de opsporing van georganiseerde misdrijven, waaronder terrorisme. Hier vervaagt in zekere zin de grens tussen opsporing op basis van een redelijk vermoeden dat een misdaad beraamd wordt of gepleegd is, en meer algemene doelen om misdaad te voorkomen (Kooijmans en Mevis 2013: 11).

5.2.3 VEILIGHEIDSDIENSTEN

Voor de inlichtingen- en veiligheidsdiensten zijn vooral voorspellen en voorkomen belangrijk. De Wet op de inlichtingen- en veiligheidsdiensten (Wiv) regelt welke gegevens de AIVD en MIVD mogen verzamelen in het kader van hun taak de

nationale veiligheid te waarborgen. De AIVD richt zich daartoe onder meer op "het verrichten van onderzoek met betrekking tot organisaties en personen die door de doelen die zij nastreven, dan wel door hun activiteiten aanleiding geven tot het ernstige vermoeden dat zij een gevaar vormen voor het voortbestaan van de democratische rechtsorde, dan wel voor de veiligheid of voor andere gewichtige belangen van de staat" (art. 6 Wiv). De MIVD verricht onderzoek "omtrent het potentieel en de strijdkrachten van andere mogendheden, ten behoeve van een juiste opbouw en een doeltreffend gebruik van de krijgsmacht", en "naar factoren die van invloed zijn of kunnen zijn op de handhaving en bevordering van de internationale rechtsorde voor zover de krijgsmacht daarbij is betrokken of naar verwachting betrokken kan worden" (art. 7 Wiv).

De gegevensverwerking door de inlichtingendiensten is aan doelen gebonden, maar deze zijn zeer breed gedefinieerd. De inlichtingendiensten hebben zeer ruime bevoegdheden om gegevens te verzamelen, van het aftappen, ontvangen, opnemen en afluisteren van elke vorm van gesprek tot het inzetten van infiltranten en het opvragen van persoonsgegevens bij aanbieders van openbare telecommunicatienetwerken en openbare telecommunicatiediensten. Bovendien kunnen zij (bulk)gegevens van zusterdiensten ontvangen. De huidige discussie rondom de herziening van de Wiv richt zich op de wens van de diensten om ongericht gegevens te kunnen verzamelen via de kabel, waar het meeste internetverkeer zich afspeelt. In de huidige situatie mogen de inlichtingendiensten alleen informatie door de ether ongericht, dat wil zeggen in bulk, opvangen. Hierdoor kunnen de diensten grootschalige netwerkanalyses uitvoeren. Het doorzoeken van deze informatie vereist toestemming van de minister. Informatie via een kabel mag daarentegen alleen gericht op een specifieke persoon of organisatie worden verzameld, en ook daarvoor is toestemming van de minister nodig. Het kabinet-Rutte II wil dit onderscheid tussen ether en kabel opheffen en een nieuw stelsel van waarborgen invoeren. De mogelijkheid om eerst te verzamelen en daarna pas te analyseren wordt daarbij gehandhaafd en uitgebreid naar kabelgegevens (Jacobs 2016). De verwerking van informatie kan hierbij zeer uiteenlopende activiteiten omvatten, van statistische analyses van metadata tot het testen van profielen en zoeken naar nieuwe doelwitten. Ook hier geldt dat het werk in de boezem van de organisatie – de koppeling en de analyse van data – in belang toeneemt.

5.2.4 OVERIGE OVERHEIDSORGANEN BELAST MET VEILIGHEID EN OPENBARE ORDE

Een deel van de organisaties binnen het veiligheidsdomein valt tot slot onder de Wbp. Het gaat vooral om overheidsorganisaties en samenwerkingsverbanden die actief zijn op het terrein van fraudebestrijding en openbareordevraagstukken. De Wbp geeft uitwerking aan de Europese gegevensbeschermingsrichtlijn 95/46/EG. Momenteel lopen de laatste onderhandelingen over de vervanging van deze richtlijn door de Algemene Verordening Gegevensbescherming (AVG). In de Wbp

keren, in meer uitgewerkte vorm, de in het EVRM en het Handvest relevante bepalingen over gegevensverwerking terug. Deze worden aangevuld met bepalingen over bewaar- en vernietigingstermijnen en maatregelen die de verwerkende partijen in acht moeten nemen om persoonsgegevens te beveiligen tegen verlies, diefstal of onrechtmatig gebruik. De Wbp is expliciet niet van toepassing op de verwerking van persoonsgegevens door of ten behoeve van de inlichtingen- en veiligheidsdiensten, ten behoeve van de uitvoering van de politietaak, of ter uitvoering van de Wet justitiële en strafvorderlijke gegevens (art. 2 lid 2 onder b, c en e Wbp).

De Wbp verbindt aan de verwerking van persoonsgegevens in hoofdlijnen de volgende voorwaarden:

- Gegevens mogen alleen verwerkt worden voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden (art. 7 Wbp).
- Een verwerking is gerechtvaardigd als zij is gebaseerd op een van de limitatief opgesomde grondslagen uit artikel 8a-f Wbp. Deze grondslagen zijn onder andere dat personen daar ondubbelzinnig toestemming voor moeten hebben verleend, en dat de verwerkte gegevens noodzakelijk zijn voor de uitvoering van een wettelijke verplichting of voor de goede vervulling van publiekrechtelijke taak.
- Gegevens mogen niet verder worden verwerkt op een wijze die onverenigbaar is met de doeleinden waarvoor ze zijn verkregen.
- Bijzondere persoonsgegevens, die zeer privacygevoelig zijn, mogen – enkele uitzonderingen daargelaten – niet worden verwerkt (art. 16-23 Wbp).

5.3 TOEZICHT OP DE VERWERKING VAN PERSOONSGEGEVENS

Om een rechtmatige en zorgvuldige verwerking van persoonsgegevens te waarborgen, is er onafhankelijk toezicht op de verzameling en verdere verwerking van persoonsgegevens door de verschillende instanties die zijn belast met het handhaven van de (nationale) veiligheid.

5.3.1 POLITIE EN JUSTITIE

Het zwaartepunt van de waarborgen in relatie tot het handelen van politie en justitie bij de verwerking van persoonsgegevens ligt in de fase van het verzamelen. Door het gebruik van (bijzondere) opsporingsbevoegdheden kan de politie grote hoeveelheden gegevens verzamelen. De verantwoordelijkheid voor het opsporingsonderzoek waarin deze bevoegdheden worden gebruikt ligt bij het Openbaar Ministerie. De inzet van bevoegdheden die een meer dan geringe inmenging in de persoonlijke levenssfeer maken vindt enkel plaats na goedkeuring van de officier van justitie.

Daar waar sprake is van het gebruik van zware bevoegdheden die diep ingrijpen in de persoonlijke levenssfeer van burgers is er de onafhankelijke rechterlijke toetsing door de rechter-commissaris. De rechter-commissaris is belast met het toezicht op het opsporingsonderzoek en moet een machtiging verlenen voor de toepassing van bepaalde bijzondere opsporingsbevoegdheden (art. 170 lid 2 WvSv). Wanneer gegevens eenmaal rechtmatig verzameld zijn en in de politiesystemen zijn beland, dan is de Wpg van toepassing. De Autoriteit Persoonsgegevens is belast met het toezicht op de Wpg (en de Wet justitiële en strafvorderlijke gegevens). Ten slotte is er in concrete strafzaken de rechterlijke toetsing achteraf. Wanneer de grondrechten van de verdachte zijn geschonden, bijvoorbeeld door de onrechtmatige inzet van (bijzondere) opsporingsbevoegdheden, dan kan de strafmaat worden verlaagd; in uitzonderlijke gevallen kan het bewijs worden uitgesloten of kan het OM niet ontvankelijk worden verklaard.

De juridische kaders voor politie en justitie stellen dus de zwaarste eisen aan het verzamelen van persoonsgegevens door middel van de inzet van bijzondere opsporingsbevoegdheden. Wanneer de politieorganisatie de gegevens op rechtmatige wijze heeft verzameld, is er vervolgens minder zwaar toezicht op het hergebruik van deze gegevens, bijvoorbeeld in het kader van een ander strafrechtelijk onderzoek. Binnen de juridische kaders voor gegevensverwerking door politie en justitie bestaat dus gedeeltelijk een hiaat ten aanzien van de uitvoering van Big Data-analyses.

5.3.2 INLICHTINGEN- EN VEILIGHEIDSDIENSTEN

Het handelen van de inlichtingen- en veiligheidsdiensten is noch vooraf noch achteraf gebonden aan rechterlijke toetsing. De rechter komt wel in beeld als gegevens die zijn verzameld door de AIVD in strafzaken worden gebruikt. De verdediging kan bijvoorbeeld de rechtmatigheid van het verkregen bewijs betwisten.[3] Toezicht vindt plaats door middel van parlementaire controle en door de Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten.

5.3.3 OVERIGE OVERHEIDSORGANEN

Toezicht op de overheidsorganen die onder de Wbp vallen geschiedt door de Autoriteit Persoonsgegevens (AP). De AP doet onderzoek, adviseert, behandelt klachten over de naleving van de Wbp, bemiddelt bij geschillen over inzage en correctie, en legt boetes op aan organisaties die op onrechtmatige wijze persoonsgegevens verwerken. De AP behandelt geen individuele klachten van burgers over de verwerking van hun persoonsgegevens – dat doet de Nationale ombudsman – maar gebruikt deze klachten wel als signaal om handhavend op te treden. De ombudsman doet uit eigen beweging en op verzoek onderzoek naar de gedragingen van bestuursorganen, mede op basis van klachten. De uitspraken over deze klachten zijn niet bindend. Op wat grotere afstand speelt ook de Algemene Rekenkamer een rol, die toezicht houdt op de doelmatigheid van organisaties zoals

inspecties en nagaat of voor het bereiken van geplande beleidsprestaties de juiste hoeveelheid middelen wordt ingezet. Dit laatste is onder meer van belang voor de inzet van Big Data-analyses, waarover soms torenhoge verwachtingen leven.

5.4 BIG DATA EN DE KERNPRINCIPES VAN GEGEVENSBESCHERMING

De vormgeving van Big Data-processen zet een aantal van de kernprincipes van de juridische kaders voor gegevensverzameling binnen het veiligheidsdomein onder druk. Deze principes zijn 'doelbinding' en 'noodzakelijkheid' (opgebouwd uit effectiviteit, proportionaliteit en subsidiariteit). Tabel 5.1 laat zien hoe deze beginselen wettelijk zijn verankerd.

Tabel 5.1 Kernprincipes gegevensbescherming binnen het veiligheidsdomein I

	Doelbinding	Noodzakelijkheid
Wet politiegegevens	Art. 3: 3	Art. 3: 1 & 2
Wiv	Art. 12: 2	Art. 12: 2
Wbp	Art. 7 & 9	Art. 7, 8 & 11

Het principe van doelbinding koppelt de verzameling en verwerking van gegevens aan vooraf vastgestelde doeleinden. Dit staat op gespannen voet met het uitgangspunt van ongerichte verzameling en hergebruik van data binnen Big Data-processen. Het principe van noodzakelijkheid stelt dat gegevens alleen mogen worden verwerkt wanneer (1) de ingreep in relatie staat tot het te dienen belang (proportionaliteit) en (2) er niet een minder ingrijpende verwerking of ander middel voorhanden is. Noodzakelijkheid impliceert daarmee ook dataminimalisatie: de gegevensverzameling moet tot het hoogstnoodzakelijke worden beperkt. In het Big Data-tijdperk geldt daarentegen juist dat meer data resulteren in betere uitkomsten. Ten slotte impliceert de toets van noodzakelijkheid ook een effectiviteitstoets: een verwerking van persoonsgegevens die niet effectief is om het doel te bereiken, kan immers nooit noodzakelijk zijn. Voor deze afweging ontbreekt momenteel echter voldoende kennis van en inzicht in de effectiviteit van Big Data-analyses.

5.4.1 DOELBINDING VERSUS HERGEBRUIK VAN INFORMATIE

Doelbinding heeft betrekking op zowel het verzamelen als het verder verwerken van persoonsgegevens. Het verzamelen mag niet willekeurig zijn en dient te geschieden op basis van een wettelijk omschreven taak of bevoegdheid. Het verder verwerken van gegevens mag alleen in het kader van deze taak of bevoegdheid plaatsvinden (de in de voorgaande paragraaf genoemde uitzonderingen daargelaten). Doelbinding verhoudt zich slecht tot de logica van Big Data-processen. Uit het verbod op verdere verwerking voor andere doeleinden volgt dat secundair

gebruik – een van de kernprincipes van Big Data-processen – in principe niet is toegestaan. Maar de doelen van organisaties in het veiligheidsdomein zijn vaak ruim omschreven – openbare orde, veiligheid, nationale veiligheid – hetgeen de doelbinding een zekere rekkelijkheid geeft.

Spanningen met Big Data

In Big Data-processen wordt vaak een koppeling van verschillende databases gemaakt. Het oorspronkelijke doel waarvoor gegevens zijn verzameld gaat daarbij verloren. Naarmate meer data in samenwerkingsverbanden en informatienetwerken belanden en hergebruik van data de regel wordt, neemt de druk op het principe van doelbinding toe. Big Data-processen zijn in het veiligheidsdomein zeker nog geen standaardwerkwijze, maar de tendens om data te delen is binnen de overheid zeker aanwezig. Bovendien komt ook het combineren van publieke gegevens met private data steeds vaker voor. Overheidsorganisaties begeven zich daarmee steeds meer op terreinen die niet traditioneel tot hun takenpakket behoren en laten data die verzameld zijn op basis van afgebakende taken en doelstellingen over de grenzen van die organisatie en taken heenvloeien. Een recente brief van minister Van der Steur (Ministerie van Veiligheid en Justitie 2015) onderstreept deze tendens.

In de wereld van de veiligheids- en inlichtingendiensten is deze ontwikkeling waarschijnlijk het verst voortgeschreden. De gegevensverzameling door Amerikaanse en Britse diensten lijkt gebaseerd op het idee om zoveel mogelijk data te verzamelen. Voor Nederland geldt dit niet, maar de uitbreiding van de wettelijke bevoegdheden van de Nederlandse inlichtingen- en veiligheidsdiensten die nu in voorbereiding is, maakt het verzamelen van bulkgegevens via de kabel wel mogelijk. De algemene tendens is dat de dataverzameling in eerste instantie zeer groot en breed kan zijn – zie de voorbeelden in hoofdstuk 3 – en dat er pas in tweede instantie wordt geselecteerd en wordt beslist op welke datapunten er een analyse wordt gedaan (Jacobs 2016). Hoewel deze vervolgstap zich wel tot het vereiste van een legitiem doel kan verhouden, is de ongerichte verzameling van data een probleem. Dat betekent dat in het domein van veiligheid steeds vaker ook de gegevens van 'niet-verdachte personen' verzameld en verwerkt worden in onderzoeken, controle, toezicht en recherchewerk, hetgeen op gespannen voet staat met de onschuldpresumptie (Hildebrandt 2016). Mede naar aanleiding van de Snowden-onthullingen zijn er echter ook wetgevingsinitiatieven die de risico's van ongerichte bulkverzameling moeten verkleinen. Zo is in de Verenigde Staten in 2015 de USA Freedom Act aangenomen (H.R. 2048), die – deels – grenzen stelt aan de ongerichte verzameling van persoonsgegevens.

Ook de internationale samenwerking en data-uitwisseling met buitenlandse inlichtingen- en veiligheidsdiensten roept vragen op over de grenzen van doelbinding. De Nederlandse wet stelt begrenzingen aan de mogelijkheden van de

AIVD en de MIVD om gegevens over Nederlandse burgers te verzamelen. Het kan echter voorkomen dat de Amerikaanse inlichtingendiensten, die minder begrensd zijn ten aanzien van het vergaren van gegevens over Nederlandse onderdanen, deze gegevens verzamelen en vervolgens uitwisselen met de Nederlandse inlichtingendiensten. Zodoende kan de AIVD gebruikmaken van gegevens die zij volgens de Nederlandse wet niet zelf zou mogen verzamelen. In Nederland loopt er een rechtszaak over deze internationale samenwerking tussen inlichtingendiensten. De rechter heeft geoordeeld dat Nederlandse inlichtingendiensten er vooralsnog van uit mogen gaan dat de gegevens die zij verkrijgen via zusterorganisaties in het buitenland rechtmatig en legitiem zijn verkregen.[4] Dit vraagstuk beperkt zich overigens niet tot het delen van data bij inlichtingenwerk. Het doet zich tevens voor als een organisatie toegang krijgt tot gegevens van een andere organisatie, die hij zelf niet had mogen verzamelen.

5.4.2 NOODZAKELIJKHEID VERSUS GROOTSCHALIG EN ONGERICHT VERZAMELEN

De huidige juridische kaders zijn gebaseerd op het beginsel van noodzakelijkheid. Noodzakelijkheid impliceert dataminimalisatie. Dit uitgangspunt houdt in dat er in principe zo min mogelijk gegevens moeten worden verzameld en in elk geval niet meer dan nodig zijn om het specifiek geformuleerde doel te bereiken. Bovendien moeten deze gegevens in de meeste gevallen weer worden verwijderd of geanonimiseerd als het doel is bereikt. Het principe van dataminimalisatie is bedoeld om ervoor te zorgen dat de verwerkende partijen zich beperken tot hun taak en opdracht. Het feit dat er meer data te verkrijgen zijn, is op zichzelf geen afdoende reden om deze te verzamelen. De noodzakelijkheidstoets valt uiteen in een proportionaliteits-, subsidiariteits- en effectiviteitstoets. Alle drie de onderdelen staan op gespannen voet met de uitgangspunten van Big Data.

Proportionaliteit

Doelen als nationale veiligheid, openbare veiligheid en het voorkomen van strafbare feiten zijn tamelijk onomstreden. Datzelfde geldt voor de uitvoering van wettelijke en publiekrechtelijke taken op het terrein van toezicht, controle en fraudebestrijding. Tegelijkertijd is duidelijk dat niet alle middelen geoorloofd zijn om deze doelen na te streven. Over de inzet van die middelen is telkens een zorgvuldige afweging vereist. Die afweging veronderstelt inzicht in de mate waarin Big Data-analyses ingrijpen in de privacy van burgers. Een probleem is echter dat Big Data-processen zich vaak niet op specifieke individuen richten, maar op groepen mensen of potentieel iedereen. Daardoor is het lastig na te gaan welke schade dataverzameling veroorzaakt. Hier komt bij dat gegevens uit verschillende bronnen een breder beeld kunnen geven van het leven van een persoon dan wanneer hij of zij fysiek in de gaten wordt gehouden. Big Data-processen zijn in hun aard en reikwijdte onvergelijkbaar met meer traditionele onderzoeks- en opsporingsmiddelen.

Subsidiariteit

Subsidiariteit hangt nauw samen met het beginsel van dataminimalisatie. In Big Data-processen komt ook dit uitgangspunt onder druk te staan. In plaats van 'select before you collect' is 'collect before you select' hier het gangbare procedé. De interne logica van Big Data-processen gebiedt immers dat zoveel mogelijk gegevens worden verzameld, zoveel mogelijk databases worden gekoppeld en dat er altijd nieuwe doeleinden voor al gebruikte data te vinden zijn. Ook de eis dat gegevens moeten worden verwijderd of geanonimiseerd als ze niet langer noodzakelijk zijn voor het verwezenlijken van het doel waarvoor ze zijn verzameld, komt zo onder druk te staan. Het principe van dataminimalisatie is sterk verbonden met de doelbindingsvereisten. Als deze worden gewijzigd, of door de Big Data-omstandigheden worden ingehaald, zal het principe van dataminimalisatie mee (moeten) bewegen.

Effectiviteit

Ook is kennis vereist over de effectiviteit van Big Data-analyses en daarover is vooralsnog weinig bekend. Dat heeft deels met methodologische problemen te maken. Het effect van Big Data-analyses is in de praktijk vaak moeilijk van andere beleidsinspanningen te isoleren. Deels ook ontbreekt het simpelweg aan goed onderbouwd wetenschappelijk onderzoek. Veel Big Data-toepassingen zijn te kort in gebruik om goed geëvalueerd te kunnen worden. De beschikbare studies zijn daarnaast vaak gebrekkig uitgevoerd en de resultaten onvolledig (Van Brakel 2016b). Dit maakt het lastig om te beoordelen of de inzet van Big Data-analyses effectief en gerechtvaardigd is.

De discussie over de effectiviteit van Big Data-analyses speelt bij uitstek ook bij de uitbreiding van de bevoegdheid om gegevens te verwerken. De aanslagen van 9 september 2001 in New York en de aanslagen die daarop volgden – de moord op Pim Fortuyn (2002) en Theo van Gogh (2004), Madrid (2004), Londen (2005), Parijs (2015, bij herhaling) en Brussel (2016) – hebben de uitbreiding van bevoegdheden op dit terrein sterk versneld. Over de effectiviteit van deze uitbreiding bestaan echter de nodige twijfels. De Amerikaanse Privacy and Civil Liberties Oversight Board (2014) evalueerde na de onthullingen van Snowden de effectiviteit van de grootschalige en permanente verzameling van metadata van Amerikaanse telefoongesprekken door de NSA. De toezichthouder concludeerde dat het programma (gebaseerd op art. 215 van de Patriot Act) in geen enkel geval rechtstreeks had bijgedragen aan de ontdekking van tot dan toe onbekende terroristische complotten of aanvallen (idem 2014: 144-155).

Deze discussie speelt – hoewel beperkt – ook in Nederland. Zo betreurt de Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten (2015a) het gebrek aan discussie over de effectiviteit van interceptiebevoegdheden. Volgens het nieuwe wetsvoorstel ter vervanging van de Wiv 2002 vindt de regering het nodig

dat de inlichtingen- en veiligheidsdiensten ongericht bulkgegevens via de kabel kunnen onderscheppen. De toezichthouder is echter van mening dat eerst overtuigend moet worden aangetoond dat de effectiviteit van de huidige bevoegdheden ernstig tekortschiet, alvorens het gesprek over de noodzaak van nieuwe bevoegdheden aan de orde is.

Wat ten slotte in meer algemene zin de noodzakelijkheidstoets in het kader van Big Data bemoeilijkt, is het loslaten van het principe van doelbinding. Binnen het huidige juridische kader is de noodzakelijkheidstoets primair gekoppeld aan het verzameldoel. Voordat gegevens mogen worden verzameld wordt een doel gespecificeerd en dit doel vormt het uitgangspunt voor de noodzakelijkheidstoets. Deze methode komt bij Big Data echter steeds meer onder druk te staan. Zo kan volgens de US National Research Council (2015: 57) bijvoorbeeld geen enkel technisch middel de bulkverzameling van metagegevens volledig vervangen. Bulkgegevens bevatten namelijk historische data, die vaak pas betekenis krijgen in het licht van nieuwe gebeurtenissen. Hiertoe moeten deze data echter wel eerst worden binnengehaald en bewaard. Dit impliceert dat de noodzakelijkheid van Big Data-analyses eigenlijk alleen pas achteraf – dus na de verzameling – is vast te stellen.

5.4.3 BEVEILIGING EN JUISTHEID VAN (BIG) DATA

Tot slot bevat de wet- en regelgeving voor gegevensverwerking in het veiligheidsdomein ook bepalingen over de juistheid en beveiliging van data (zie tabel 5.2). Een deel van deze wetgeving is algemeen van aard, zoals de Wbp en daarop gebaseerde uitvoeringswetgeving, en een deel is sectorspecifiek zoals de Wiv en de Wpg. De wettelijke bepalingen over de beveiliging van data in de Wbp gelden als tamelijk algemeen (Kleve et al. 2013). Er is namelijk niet of nauwelijks gespecificeerd aan welk beveiligingsniveau organisaties moeten voldoen: het moet 'passend' zijn en rekening houden met 'de stand van de techniek', 'de kosten', 'de risico's van de verwerking' en 'de aard' van de betrokken persoonsgegevens. In de aankomende Algemene Verordening Gegevensbescherming zullen de bepalingen omtrent technische veiligheid waarschijnlijk worden uitgebreid en bestendigd (art. 23, 28, 30-32 AVG). Ook de zorg voor juiste en nauwkeurige gegevens is slechts globaal omschreven. Zo vermeldt de Wpg enkel dat de verantwoordelijke gebleken onjuiste of onvolledige politiegegevens moet verbeteren of vernietigen (art. 4 lid 1 Wpg).

Tabel 5.2 Kernprincipes van gegevensbescherming binnen het veiligheidsdomein II

	Beveiliging	Juistheid
Wet politiegegevens	Art. 4: 3	Art. 4: 1
Wiv	Art. 16b	(Art. 12: 4: Art. 16a)*
Wbp	Art. 13 (14)	Art. 11: 2

* Art. 12 lid 4 Wiv rept van 'een aanduiding omtrent de mate van betrouwbaarheid'.

Spanningen met Big Data

In Big Data-processen komt de (eind)verantwoordelijkheid voor de juistheid van gegevens onder druk te staan. Een van de kenmerken van Big Data is dat de accuraatheid van gegevens niet altijd gegarandeerd is (Lukoianova en Rubin 2014). In het merendeel van de gevallen is de verantwoordelijke de organisatie die de data verzamelt en/of in bezit heeft. In sommige gevallen strekt deze verantwoordelijkheid zich ook deels uit naar de organisatie die in de praktijk de gegevens analyseert. Een voorbeeld hiervan is het Inlichtingenbureau, dat in opdracht handelt van onder meer gemeenten en diverse overheidsorganisaties die aan het SyRI deelnemen. Wanneer organisaties data met elkaar delen, bewerken en opnieuw contextualiseren, is er geen logische verantwoordelijke meer aan te wijzen voor de kwaliteit en betrouwbaarheid van het proces en van het eindresultaat (WRR 2011). Dat was al een probleem, maar naarmate Big Data-processen zich verder ontwikkelen en 'gewoner worden', wordt het probleem pregnanter. Dezelfde problematiek geldt als verschillende partijen werken met en toegang hebben tot een database waarvan de gegevens vertrouwelijk en veilig moeten worden opgeslagen.

5.5 BIG DATA EN DE INVLOED OP FUNDAMENTELE RECHTEN NAAST PRIVACY

Naast het recht op privacy heeft de burger ook een aantal andere vrijheidsrechten die een rol spelen in Big Data-processen. De belangrijkste hiervan zijn de vrijheid van religie, de vrijheid van meningsuiting en de vrijheid van vergadering en betoging. Big Data-processen dienen om patronen te construeren. Wanneer gedrag, handelingen of content aan een of meer patronen voldoet, bijvoorbeeld dat van een potentiële terrorist, dan kan deze persoon vroegtijdig worden gevolgd of in zijn gedrag worden gehinderd. Hij of zij kan zelfs worden opgepakt wanneer er sterke aanwijzingen zijn voor het beramen van een misdaad. Het is echter problematisch om preventieve maatregelen uitsluitend op basis van statistische waarschijnlijkheden te nemen. Bovendien zijn deze maatregelen altijd over-inclusief en perken zij derhalve de vrijheid van burgers in. Niet elke persoon die aan een profiel voldoet zal immers in de praktijk overgaan tot het plegen van strafbare feiten. Binnen het veiligheidsdomein moeten hier moeilijke keuzes gemaakt worden: de dreiging van geweld en terreur gaat tenslotte ook ten koste van de vrijheid en de veiligheid van de burger. Dergelijke moeilijke keuzes in Big Data-processen vragen om zorgvuldige procedurele waarborgen.

5.5.1 PERSOONLIJKE AUTONOMIE

Big Data-processen kunnen het effect hebben dat mensen hun gedrag aanpassen. Wanneer burgers het gevoel hebben dat hun gedrag wordt geobserveerd, dan kan dit leiden tot aanpassing van individueel gedrag aan de norm die door de machthebber is gesteld en tot het achterwege laten van – vermeend – niet-normconform gedrag (zoals bijvoorbeeld kritische uitlatingen over de machthebber). Deze vrees

heet ook wel het *chilling effect* (zie hoofdstuk 4). Dit effect wordt expliciet erkend door het Europese Hof voor de Rechten van de Mens in relatie tot de vrijheid van meningsuiting (art. 10 EVRM).[5] Maar *chilling effects* kunnen ook een bedreiging vormen voor bijvoorbeeld het recht op vrijheid van gedachte, geweten en godsdienst. *Chilling effects* worden belangrijker naarmate Big Data meer voet aan de grond krijgt en in steeds meer onderdelen van het dagelijks leven een rol gaat spelen. Onzekerheid over toekomstige normen en gedragsconsequenties kunnen deze effecten sterker maken. De gevolgen van *chilling effects* zijn echter lastig te duiden en te specificeren. Zo signaleerden Marthews en Tucker (2015) dat na de Snowden-onthullingen mensen bepaalde zoektermen op internet vermeden. Een causaal verband is daarmee echter nog niet aangetoond. Ook zijn de gevolgen van *chilling effects* vaak pas op de langere termijn waarneembaar, omdat mensen hun gedrag stapsgewijs aanpassen. Bovendien gaat het behalve om individuele belangen veeleer om een algemene en structurele tendens die zich laat voelen op een collectief niveau.

5.5.2 RECHT OP BESCHERMING TEGEN DISCRIMINATIE EN STIGMATISERING

Big Data-processen kunnen het gevaar van discriminatie in zich herbergen. Dit risico is vooral aanwezig bij voorspellende uitspraken op basis van profielen en geaggregeerde groepskenmerken. Zo kan de gegevensstroom "moslim+vakantie naar Jemen+bezoek aan website X" een reden zijn om iemand scherper in de gaten te houden. Deze persoon voldoet dan wellicht aan het profiel van een potentiële terrorist. Dergelijke profielen kunnen echter te sterk de nadruk leggen op – deels onvervreemdbare – eigenschappen als ras, geloof of andere gevoelige kenmerken, waardoor het risico van discriminatie en ongelijke behandeling op de loer ligt. Daarbij komt de vrees dat dergelijke profielen en de acties die daarop zijn gebaseerd, kunnen leiden tot stigmatisatie en een bevestiging van stereotypes, zowel in toezicht en opsporing als in het maatschappelijke debat (Custers et al. 2014).

Daarnaast is er het risico dat door profiling de zwakkeren in de samenleving in 'slechte' categorieën worden ingedeeld en daardoor minder kansen krijgen, terwijl sterkere personen in 'goede' categorieën worden ingedeeld. Bias in bestaande en nieuwe databestanden kan dit proces verder versterken (Boyd en Crawford 2012). Stel bijvoorbeeld dat in wijk X veel wordt gesurveilleerd wegens overlast van bepaalde etnische groepen. Als gevolg hiervan gaat de database informatie over voornamelijk deze bevolkingsgroepen bevatten. Dit kan als gevolg hebben dat het initiële beeld wordt bevestigd. Vice versa geldt overigens hetzelfde, namelijk dat bepaalde bevolkingsgroepen niet of nauwelijks in bepaalde databases worden opgenomen en derhalve uit het zicht van de overheid blijven, bijvoorbeeld illegale immigranten en daklozen. Hierdoor kunnen zij in mindere mate profiteren van de voordelen van Big Data-processen.

5.5.3 HET RECHT OP EEN EERLIJK PROCES

Big Data-processen raken tot slot aan het recht op een eerlijk proces, met name als het ongericht verzamelen van gegevens een grotere rol gaat spelen in het veiligheidsdomein. Het Nederlandse recht, het EVRM en het Handvest van de grondrechten van de Europese Unie kennen het recht op een eerlijk proces, dat kort gezegd luidt dat eenieder tegen wie een vervolging is ingesteld voor onschuldig dient te worden gehouden totdat zijn schuld in rechte is komen vast te staan.

Spanningen met Big Data

Door Big Data-processen kan dit recht onder druk komen te staan (Crawford en Schultz 2014), vooral wanneer deze processen repressieve preventie ten doel hebben (Kerr en Earle 2013). De politie mag ter uitvoering van haar algemene politietaak 'ogen en oren open houden' (art. 3 Politiewet). Hierbij is er nog geen sprake van een meer dan geringe inmenging in de persoonlijke levenssfeer. In hoeverre geautomatiseerde surveillance (*open source intelligence*) en Big Data-analyses hierbinnen passen is nog voorwerp van discussie (zie in dit kader Oerlemans en Koops 2012).

Voor de toepassing van dieper ingrijpende opsporingsmethoden moet er een verdenking van een misdrijf zijn, of moet er een redelijk vermoeden bestaan dat strafbare feiten worden gepleegd in georganiseerd verband, dan wel dienen er aanwijzingen te zijn dat een terroristisch misdrijf wordt beraamd of gepleegd (titel IVa-VE SV). De vraag is of en zo ja, hoe ongerichte bulkverzameling van gegevens en de daarbij behorende Big Data-analyses moeten worden gezien als dergelijke bijzondere opsporingsbevoegdheden. Hoe dan ook staat het verzamelen van grote hoeveelheden gegevens over zeer grote groepen mensen zonder concrete verdenking op gespannen voet met de onschuldpresumptie. Door de nadruk op profiling en *preventive* en *predictive policing* krijgen profielen en andere methoden van analyse een grotere rol in het bepalen van handelingen, maatregelen en sancties van de politie. Sommige van deze vooronderzoeken en andere activiteiten vormen uiteindelijk toch de opmaat tot een daadwerkelijk strafrechtelijk onderzoek gericht op een verdachte, of spelen een rol in de vervolgstappen van dat onderzoek. De vraag is dan evenwel wanneer er een redelijk vermoeden van schuld is gerezen en wanneer er dus rechtmatig (bijzondere) opsporingsbevoegdheden ingezet mochten worden.

Transparantie en equality of arms

Daarbij komt dat de politie vaak veel betere technologische middelen heeft om te profileren en aan dataverwerking te doen. De burger weet vaak niet hoe bepaalde algoritmes, computersystemen en profielen precies werken en kan zich er dus ook moeilijk tegen verdedigen. Ditzelfde geldt voor de verdediging in strafzaken die zich geconfronteerd ziet met bewijs gebaseerd op Big Data-analyses. Hierdoor ontstaat er *inequality of arms*, zoals ook lange tijd het geval was bij DNA-onderzoek en

andere dure forensische onderzoeken, waarvan het voor de verdediging vaak moeilijk was een onafhankelijk tegenonderzoek te laten bekostigen. Bij een veroordeling van een verdachte kan niet worden gewerkt met aannames of assumpties die uitsluitend volgen uit geautomatiseerde besluitvorming of algemene groepsprofielen. Deze gelden niet als wettig en overtuigend bewijsmiddel. Big Data zullen derhalve vooral een rol spelen als ondersteuning van de bewijslast. Big Data-analyses kunnen een zaak bovendien in een bepaald licht zetten ('framing') en de verdediging moet een dergelijk beeld kunnen weerleggen. De vraag is of zij – of hun advocaten – zich hiertegen voldoende kunnen wapenen. Als de autoriteiten een grote technologische voorsprong hebben, staat de zogenoemde *equality of arms* tussen de partijen in de rechtszaal onder druk (Fairfield en Luna 2014).

5.6 BIG DATA EN DE GRENZEN VAN HET HUIDIGE JURIDISCHE KADER

Big Data legt problemen met het huidige juridische kader bloot die fundamenteel van aard zijn. De werkbaarheid van uitgangspunten als doelbinding en noodzakelijkheid vormt hierbij het voornaamste vraagstuk. Wanneer deze principes onder druk komen te staan, heeft dat consequenties voor de wijze waarop deze rechten zijn te effectueren. Maar naast dit centrale probleem waar Big Data het huidige juridische kader voor plaatst, zijn er nog enkele andere gerelateerde vraagstukken. Deze worden hierna besproken.

5.6.1 DE WERKBAARHEID VAN HET BEGRIP PERSOONSGEGEVEN

Een persoonsgegeven is volgens de Wbp een gegeven dat iemand direct of indirect identificeert of kan identificeren (art. 1a Wbp), een definitie die in de diverse sectorwetgeving is gevolgd. Als gegevens niet tot een persoon zijn te herleiden, dan valt de verwerking daarvan in principe buiten de geboden bescherming. Het onderscheid tussen gewone gegevens en persoonsgegevens verliest in het Big Data-tijdperk echter steeds meer aan zeggingskracht.

In het Big Data-tijdperk doorlopen gegevens steeds meer een circulair proces: zij worden gekoppeld, geaggregeerd en geanonimiseerd en vervolgens weer ontdaan van anonimisatie, verrijkt met andere gegevens en tot persoonsprofielen gemaakt, enzovoort. Mede hierdoor zijn gegevens tegenwoordig vrijwel altijd weer te herleiden tot een individu (Ohm 2010; Koot 2012; Sweeney 2002). Het is vooraf meestal niet te zeggen hoe gegevens moeten worden gekwalificeerd (als persoonsgegeven of niet), welke rol deze gegevens krijgen, welk gebruik daarvan zal worden gemaakt en in hoeverre ze iets over individuen zullen zeggen. Dit heeft er al toe geleid dat de juridische definitie van 'persoonsgegeven' in de loop der jaren steeds breder is geformuleerd (Van der Sloot 2014a); hierdoor kunnen meer gegevens onder het huidige gegevensbeschermingsrecht worden gebracht.[6] In zijn algemeenheid geldt echter dat des te wijder een recht is, des te zwakker. Het is daarmee

de vraag of het in de toekomst bij het gebruik van Big Data voor veiligheidsdoeleinden überhaupt nog wel zinvol is om te spreken over een onderscheid tussen 'gewone gegevens' en 'persoonsgegevens'. Immers, spontane identificatie van personen door de combinatie van losse, op zichzelf anonieme data, zal steeds sneller plaatsvinden (Bennett en Bayley 2016). Ook is de toepassing van Big Data in het veiligheidsdomein in de meeste gevallen juist gericht op het – in een vroegtijdig stadium – vinden (en waar mogelijk identificeren) van personen die een risico vormen voor de veiligheid. Daar waar de mogelijkheid bestaat om een persoon uit een groep te pikken (*to single out*), is het gegevensbeschermingsrecht reeds van toepassing.[7]

Dit probleem strekt zich ook uit tot de kwalificatie betreffende de 'gevoeligheid' van de gegevens. Metadata zijn in tegenstelling tot de inhoud van de communicatie bijvoorbeeld niet speciaal beschermd (Koops 2013b). Toch kunnen metadata in sommige gevallen een zeer gedetailleerd beeld opleveren van iemands persoon en persoonlijke leven en kunnen zij dus raken aan de bescherming van de privésfeer (Felten 2013). Het huidige juridische kader houdt hier maar in zeer beperkte mate rekening mee.

Een aanverwante kwestie is de koppeling van ongevoelige persoonsgegevens aan andere ongevoelige persoonsgegevens of meer algemene statistieken, waarbij het resultaat gevoelig wordt. Een simpel voorbeeld is de koppeling van het algemene kenmerk 'in wijk X heeft 80 procent van de mensen ziekte Y' met het ongevoelige gegeven 'persoon Z woont in wijk X'. In combinatie levert dit een gevoelig gegeven over persoon Z op, omdat de eigenschappen van de groep (hoge prevalentie van ziekte Y) in de praktijk veelal toegedicht zullen worden aan een individueel lid van de groep (persoon Z).

Ook het onderscheid tussen gewone persoonsgegevens en bijzondere persoonsgegevens wordt in het kader van Big Data steeds minder relevant. Volgens de Wbp is het bijvoorbeeld in de meeste gevallen niet toegestaan om gegevens betreffende etniciteit te verwerken. Dit betekent dat deze gegevens uit een dataset moeten worden verwijderd. Echter, in veel gevallen zijn er achterliggende data in de dataset die (in samenhang) alsnog wat over etniciteit kunnen zeggen (bijvoorbeeld de wijk waar iemand woont, favoriete televisieprogramma's, muzieksmaak, koopgedrag, hobby's, enzovoort). Met behulp van Big Data kunnen dit soort correlaties zichtbaar worden gemaakt en kan alsnog de etniciteit met een zekere waarschijnlijkheid worden afgeleid.

5.6.2 VERMENGING VAN JURIDISCHE KADERS

Zoals geschetst in de voorgaande hoofdstukken is er een duidelijke trend zichtbaar waarbij instanties samenwerken om veiligheidsvraagstukken te adresseren. Binnen dit soort samenwerkingsverbanden worden veel gegevens uitgewisseld en

worden de gegevenssets van individuele instanties in samenhang geanalyseerd. Het wordt echter steeds onduidelijker welk regime van toepassing is op deze analyses en de daaruit verkregen informatieproducten en kennis. De huidige juridische kaders bieden regels en waarborgen die primair gekoppeld zijn aan de instantie waarop het juridisch kader betrekking heeft. In het Big Data-tijdperk valt dit onderscheid echter steeds moeilijker zuiver te houden.

5.6.3 REGULERING VAN BIG DATA IN RELATIE TOT COLLECTIEVE RECHTEN EN BELANGEN

De huidige juridische kaders zijn gericht op de bescherming van het individu en van individuele belangen. Zeker bij mensenrechten is het uitgangspunt dat deze van toepassing zijn op de bescherming van natuurlijke personen en hun belangen. Voor sommige rechten geldt echter dat ze naast persoonlijke belangen ook maatschappelijke belangen beschermen. De vrijheid van meningsuiting bijvoorbeeld, beschermt het individu in zijn wens tot expressie en ontplooiing, maar wordt tevens als voorwaarde gezien voor de onafhankelijkheid van de journalistiek. Dit laatste geldt op zijn beurt weer als een voorwaarde voor een vrije, democratische samenleving (Nieuwenhuis 2015). Andere rechten worden vaak uitsluitend in termen van persoonlijke belangen besproken, zoals het recht op privacy en gegevensbescherming. Zo wordt privacy doorgaans gekoppeld aan de bescherming van de menselijke waardigheid, de individuele autonomie en de persoonlijke vrijheid (Rössler 2005). Dit heeft in het maatschappelijk debat nog wel eens tot gevolg dat privacy het als individueel belang verliest van het collectieve belang veiligheid.

Big Data-processen richten zich vaak – in eerste instantie – niet op specifieke individuen, maar op groepen mensen of potentieel iedereen. De verzamelde en verwerkte data zijn niet altijd direct te herleiden tot een individu. Als dat al zo is, dan is nog de vraag of en hoe deze Big Data-processen schade toebrengen aan de specifieke positie en de belangen van een individu. Het is bijvoorbeeld lastig te duiden welke schade de dataverzameling door de NSA heeft toegebracht aan de concrete belangen van de doorsnee-Amerikaanse of -Europese burger. Evenzo lijkt de massale plaatsing van beveiligingscamera's in steden niet zozeer direct een impact te hebben op specifieke individuen, maar simpelweg op iedereen die in die steden woont en hebben camera's gevolgen voor de samenleving als geheel. Kort gezegd gaan veel Big Data-processen om algemene en grootschalige projecten, die een impact hebben op grote groepen of de samenleving als geheel, terwijl het recht voornamelijk inzet op de bescherming van individuen en individuele belangen. Dit is een niveauverschil dat in toenemende mate problematisch wordt.

5.7 TOEZICHT, TRANSPARANTIE EN ONAFHANKELIJKE TOETSING

In een Big Data-wereld is de rol van het onafhankelijk toezicht cruciaal. Deze rol wordt door het groeiend gebruik van Big Data en de toenemende complexiteit van gebruikte modellen en algoritmes echter wel ingewikkelder en veeleisender. Het opnieuw doordenken van de bevoegdheid en capaciteit van handhavende organisaties is daarmee van groot belang (Eskens et al. 2015). Toezicht veronderstelt daarnaast voldoende transparantie, om inzicht te kunnen krijgen in Big Data-processen. Veel van deze processen zijn echter aan het oog van de buitenwereld onttrokken, zeker in het veiligheidsdomein. Voor effectief toezicht is een grotere mate van transparantie desondanks noodzakelijk. Die transparantie geldt voor de verzameling van data, maar ook voor de analyse en het gebruik van die data. Tot slot is het belangrijk dat individuen in staat zijn om wetgeving en beleid te (laten) toetsen, door klachten in te dienen of naar de rechter te stappen. Een groot probleem hierbij is de eerder geconstateerde nadruk op individuele schade en het gebruik van persoonsgegevens. Op dit vlak loopt de wet- en regelgeving achter op de praktijk.

5.7.1 HERVORMING VAN HET HANDHAVINGSREGIME

Toezicht op en handhaving van de huidige juridische kaders verschilt per instantie. In aanvulling daarop bestaan functies als de ombudsman en instanties als het College voor de Rechten van de Mens.

Het toezicht op het verzamelen van gegevens door de politie gebeurt ex-ante door de officier van justitie en de rechter-commissaris en ex-post door de zittingsrechter. Het toezicht op de verwerking van gegevens conform de eisen van de Wpg wordt gedaan door de Autoriteit Persoonsgegevens (AP). Het is duidelijk dat in de fase van uitwisseling, koppeling en analyse van data (het domein van de Wpg) de wettelijke waarborgen en het toezicht daarop aanzienlijk minder sterk zijn dan de waarborgen in de verzamelfase (het WvSv). Wanneer de uitwisselings-, koppelings- en analysefase wint aan belang – en dat ligt in de lijn der verwachting – kan dit nieuwe risico's voor de bescherming van fundamentele rechten opleveren.

Voor het toezicht op de inlichtingen- en veiligheidsdiensten is er een apart toetsend orgaan, namelijk de Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten (CTIVD). De CTIVD ziet zich in toenemende mate geconfronteerd met Big Data-toepassingen in het werk van de inlichtingen- en veiligheidsdiensten, en wil daarom investeren in meer kennis van dataverwerkingssystemen en geautomatiseerd toezicht (CTIVD 2015a: 11). De CTIVD ziet vooral toe op de rechtmatigheid en in mindere mate op de proportionaliteit van het handelen van de diensten. Zij constateert zelf dat er meer aandacht zou mogen zijn voor doelbinding en de effectiviteit van met name grootschalige data-analyses (CTIVD 2015a: 28).

Het algemene gegevensbeschermingsrecht wordt gehandhaafd door de AP die – formeel en informeel – aanspreekpunt vormt voor een groot aantal van de in hoofdstuk 3 beschreven praktijken met Big Data, vooral van bestuursorganen. De vraag is of de AP, gezien zijn bevoegdheden, capaciteit, middelen en mankracht, voldoende is toegerust op een sterke groei van Big Data-toepassingen binnen en buiten de overheid. De aanstaande Algemene Verordening Gegevensbescherming biedt een gedeeltelijke oplossing voor dit probleem. Zij zet namelijk in op de hervorming van het handhavingsregime van het gegevensbeschermingsrecht. Een belangrijk element vormt de toekenning van een ruimere boetebevoegdheid aan handhavende organisaties (art. 79-79b AVG). Dit dwingt organisaties ertoe privacyregels hoog op de agenda te zetten. Ook zal meer worden ingezet op het aanstellen van privacyfunctionarissen binnen organisaties (art. 35-37 AVG) en op het vaststellen van gedragscodes en certificaties (art. 38-39a AVG). Daarnaast wordt er een betere samenwerking nagestreefd tussen de handhavende organisaties van de verschillende nationale lidstaten (art. 54a-72 AVG), zoals nu reeds gebeurt in de Artikel 29-werkgroep (het onafhankelijke advies- en overlegorgaan van Europese privacytoezichthouders). Dit wordt gestimuleerd door het feit dat de regels uit een verordening, in tegenstelling tot regels uit een richtlijn, directe werking hebben en dus niet meer hoeven te worden geïmplementeerd door nationale overheden; ook de handhaving en kaderstelling wordt deels op Europees niveau geplaatst (art. 64-72 en 86-87 AVG). Hierbij moet wel worden aangetekend dat deze eisen enkel gelden voor verantwoordelijken die onder het toepassingsbereik van de verordening vallen (de overheidsinstanties niet zijnde politie, justitie en inlichtingen- en veiligheidsdiensten) en er op punten binnen de verordening ruimte bestaat voor lidstaten om voor overheidstaken uitzonderingen te maken.

Voor de uitwisseling van politiële en justitiële gegevens is een nieuwe richtlijn in de maak, die het eerdergenoemde Kaderbesluit zal vervangen.[8] Evenals de verordening bevat deze richtlijn de verplichting een functionaris gegevensbescherming in te stellen (afdeling 3) alsook een meldplicht gegevensverlies (art. 28 en 29). En ook hier moet de toezichthouder meer bevoegdheden krijgen (art. 39 e.v.). Het is aan de lidstaten zelf om te bepalen op welke wijze er sancties worden opgelegd (art. 55), mits deze 'doeltreffend, evenredig en afschrikkend' zijn.

Toezicht op de inlichtingendiensten

Op andere terreinen lijkt evenwel minder animo voor het versterken van toezichthoudende instanties. Zo is er in de zomer van 2015 een conceptwetsvoorstel gepubliceerd om de Wiv te wijzigen. Een belangrijk element in dit wetsvoorstel is de verruiming van de bevoegdheden tot ongerichte interceptie van niet-kabelgebonden communicatie door de inlichtingen- en veiligheidsdiensten. Hoewel de commissie-Dessens deze verruiming van de interceptiebevoegdheden uitdrukkelijk koppelde aan een versterking van het toezicht door de CTIVD, heeft het kabinet weinig toegevoegd aan het regime voor handhaving en controle (wetsvoorstel

34027). Dit besluit kwam het kabinet op veel kritiek te staan, met name tijdens de internetconsultatie. Volgens Loof et al. (2015) voldoet het wetsvoorstel niet aan de voorwaarden voor het toezicht op de inlichtingen- en veiligheidsdiensten, zoals die zich hebben uitgekristalliseerd in jurisprudentie en in op internationaal niveau geformuleerde *soft law*-voorwaarden. Omdat ex-postrechtmatigheidstoezicht en bindende oordelen in klachtenprocedures in het wetsvoorstel ontbreken, bestaat het risico van een toekomstig internationaal of Europees oordeel dat het niet voldoet aan de mensenrechtelijke eisen.

5.7.2 MEER TRANSPARANTIE

Voor effectief toezicht is bovendien transparantie een belangrijke vereiste. Transparantie is een belangrijk uitgangspunt bij gegevensverwerking. Het idee is dat burgers weten wie welke persoonsgegevens verwerkt, een uitgangspunt dat Bennett (1992) de theorie van *information privacy* noemt. Dit maakt het mogelijk om de gegevensverwerkende partij ter verantwoording te roepen en/of onjuiste data te signaleren en corrigeren. Een belangrijk instrument hiertoe is de melding van gegevensverwerking door de verantwoordelijke partijen aan de AP (hoofdstuk 4, paragraaf 1 Wbp), dat van deze meldingen een openbaar register bijhoudt. Binnen het veiligheidsdomein ligt dit wegens geheimhoudingsbepalingen een stuk gecompliceerder. Het EHRM erkent bijvoorbeeld dat inlichtingendiensten niet effectief kunnen opereren wanneer zij volledige openheid van zaken geven (Van der Sloot 2014b). Anderzijds geldt er een plicht om personen ervan te verwittigen als zij zijn afgetapt door de politie, moeten gebieden met camerabewaking duidelijk zijn aangegeven en geldt ook in het gegevensbeschermingsrecht een transparantievereiste. Zo bepaalt artikel 126bb WVS dat de officier van justitie aan betrokkene schriftelijk mededeling doet van de uitoefening van de bevoegdheden zodra het belang van het onderzoek dat toelaat. De mededeling blijft evenwel achterwege indien uitreiking van de mededeling redelijkerwijs niet mogelijk is of als er andere bijzondere omstandigheden gelden.

Overigens keert de huidige meldplicht niet terug in de aankomende Verordening gegevensverwerking. Deze wordt vervangen door een documentatieplicht (art. 28 AVG). Daarnaast worden de algemene informatieplichten aangescherpt (art. 12-14a AVG). Ook moet meer verantwoording worden afgelegd over de logica van geautomatiseerde besluitvorming en profiling. Deze eisen beperken zich echter tot verantwoordelijken onder de verordening. Wanneer dataverwerkende organisaties niet in algemene zin gestimuleerd worden om, ten minste op elementen, meer openheid van zaken te geven, heeft dit hoe dan ook zijn weerslag op de (geringe) mate van transparantie van processen van dataverwerking binnen het veiligheidsdomein.

Spanningen met Big Data

Het principe van transparantie komt in verschillende opzichten onder druk te staan door Big Data. Data-subjecten weten vaak simpelweg niet dat hun gegevens zijn verzameld door middel van cookies, mobiele telefoons of geheime taps en zullen dus niet zo snel hun informatierecht inroepen. Zelfs als zij dat wel zouden weten, dan nog blijft het punt dat in de toekomst gegevensverwerking waarschijnlijk zo wijdverbreid is dat het bijna ondoenlijk wordt om van alle partijen en alle informatiestromen op de hoogte te blijven. Hier komt bij dat transparantie over het verzamelen van gegevens in het Big Data-tijdperk steeds minder zegt over hoe, door wie en voor welke doeleinden deze gegevens uiteindelijk in de praktijk worden gebruikt. Dit geldt evenzeer voor de organisaties die data verwerken. Voor hen is het vaak niet duidelijk op wie de gegevens betrekking hebben, waar de gegevens vandaan komen en hoe zij de datasubjecten zouden kunnen bereiken, zeker bij de koppeling van databases en het hergebruik van informatie. Zouden alle verantwoordelijken inderdaad alle datasubjecten blijven informeren over 'de ontvangers of de categorieën ontvangers van de gegevens', dan zal niet alleen de burger worden overspoeld met dergelijke berichten, maar wordt ook sisyfusarbeid van de organisaties verwacht.

De groeiende complexiteit en automatisering van analyseprocessen maakt daarnaast dat het steeds moeilijker te begrijpen is hoe organisaties tot bepaalde beslissingen komen. Veel van de gebruikte algoritmen zijn bovendien bedrijfsgeheim en niet toegankelijk voor nadere inspectie. Dat geldt behalve voor het bedrijfsleven ook voor de overheid, die voor haar Big Data-analyses vaak gebruikmaakt van diensten en producten van private partijen. Pasquale (2015: 1912-206) typeert deze ontwikkeling in termen van een 'Black Box Society'. We hoeven volgens hem niet exact te begrijpen hoe een algoritme werkt. Net als bij de motor van een auto moeten we echter goed genoeg kunnen beoordelen of hij ons veilig en comfortabel naar onze bestemming kan brengen. In het Big Data-tijdperk krijgt transparantie aldus een bredere betekenis. Het belang daarvan strekt zich over de fase van verzameling heen ook naar de fase van analyse en gebruik.

Transparante besluitvorming

Een te sterke geheimhoudingscultuur bij overheidsorganisaties kan leiden tot spanningen tussen overheid en samenleving. Er zal in Nederland (en daarbuiten) dan ook moeten worden gezocht naar een zinnige manier om in het Big Data-tijdperk de transparantie van de macht te waarborgen en een nieuwe balans te zoeken tussen wat wel en niet bekend kan worden gemaakt. In het Big Data-tijdperk zijn burgers in toenemende mate transparant voor overheidsorganisaties en bedrijfsleven. De transparantie van overheid en bedrijfsleven is geregeld binnen het kader van de regels voor databescherming. Deze regels richten zich echter vooral op de fase van het verzamelen van gegevens. Een alternatief is om *decision transparancy* te organiseren, dat wil zeggen meer openheid over beslissingen die

raken aan het leven van individuen (Koops 2013a). Dit moet grote dataverwerkende partijen dwingen tot meer zorgvuldige en eerlijke beslissingen, ongeacht hoe de verzameling en analyse van data heeft plaatsgevonden. Niet alleen het proces van gegevensverwerking, maar vooral de uitkomsten daarvan, in de vorm van beslissingen over een individu, raken aan de rechten van burgers. Transparantie moet in het Big Data-tijdperk dus vooral op besluitvorming en besluitvormingsprocessen zijn gericht, en individuen in staat stellen om te begrijpen welke data en welke wijze zijn gebruikt om tot besluiten te komen. Wanneer burgers weet hebben van welke datasets overheidsorganisaties voor beslissingen gebruiken, kunnen zij deze beslissingen ook beter toetsen en/of aanvechten.

5.7.3 ONAFHANKELIJKE TOETSING VAN BELEID EN WET- EN REGELGEVING

Het is van belang om de positie van burgers in Big Data-processen te versterken, mede om te voorkomen dat de nuttige toepassing van Big Data door de politie en andere instanties op wantrouwen stuit. Dit gebeurt deels door toezichthouders beter voor hun controlebevoegdheden toe te rusten en de transparantie over de gegevensverwerking door overheidsorganisaties te vergroten. Behartiging van algemene en maatschappelijke belangen door NGO's en burgerrechtenorganisaties kan daarin, net als op andere gebieden zoals het fysieke leefmilieu, een belangrijke rol spelen (Regan 1995; Allen 2011). Het EHRM ontvangt bijvoorbeeld klachten van zulke organisaties over massasurveillance door de overheid (Van der Sloot 2016b). Zo richtte het EHRM zich in de zaak Roman Zakharov versus Russia op de vraag of er voldoende wettelijke waarborgen zijn tegen misbruik van macht door overheidsorganen.[9] Dit zijn zogenoemde *in abstracto*-claims, waardoor niet alleen burgers, maar ook rechtspersonen voor het publiek belang kunnen opkomen. Een voordeel hiervan is dat rechtspersonen vaak beter geëquipeerd zijn om langdurige rechtszaken te voeren en zich op bredere, meer algemene belangen kunnen richten dan de individuele burgers die klachten indienen.

De combinatie van enerzijds lichtere eisen ten aanzien van de bewijsvoering voor individuele schade en anderzijds acties die het algemeen belang dienen, komt tegemoet aan het probleem dat individuele schade moeilijk aantoonbaar is en ook het gebruik van niet-persoonsgegevens een inmenging in de privacy kan vormen. De mogelijkheid voor rechtspersonen om in het algemeen belang voor privacy op te komen (art. 3: 305a BW; WODC 2014), leek in Nederland sterk verruimd toen het Gerechtshof in Den Haag de Stichting Privacy First ontvankelijk verklaarde in haar klacht over de Nederlandse Paspoortwet, die van burgers verlangt dat zij hun vingerafdruk afstaan.[10] Maar de Hoge Raad oordeelde anders en beperkte de ontvankelijkheid tot vorderingen als de onderhavige waarin een rechtspersoon opkomt voor belangen van personen die terzake geen rechtsingang hebben bij de bestuursrechter, of voor zover zij opkomt voor een eigen belang waarvoor zij geen rechtsingang heeft bij de bestuursrechter.[11] Anders dan in landen die een vorm van onafhankelijke rechterlijke toetsing van de legitimiteit van beleid en wetgeving aan

grondrechten kennen, zoals veel andere Europese staten, ontbreekt in het Nederlandse staatsbestel in zulke situaties een onafhankelijke rechterlijke mogelijkheid tot toetsing. Daardoor heeft het zwaartepunt van deze toetsing zich verplaatst naar instellingen als het EHRM en Hof van Justitie van de Europese Unie.

Niet-juridische middelen

Burgers kunnen vanwege de kosten, tijdsdruk en complexiteit van rechtszaken ook gebruikmaken van niet-juridische mogelijkheden om klachten in te dienen over Big Data-toepassingen. De Autoriteit Persoonsgegevens neemt bijvoorbeeld 'tips' aan, die vervolgens aanleiding kunnen zijn voor onderzoek.[12] De Nationale ombudsman behandelt klachten van burgers, waarbij minder strenge regels gelden omtrent de juridische status. Dit stelt burgers in staat om ook meer algemene klachten in te dienen, wat bevorderlijk kan zijn bij klachten over de bredere effecten van Big Data-toepassingen. De ombudsman kan over de uitkomst van klachtonderzoek echter uitsluitend aanbevelingen en geen bindende uitspraken doen. Klachten over de inlichtingenoperaties moeten worden ingediend bij de minister, die na onderzoek en advies door de CTIVD vervolgens een oordeel uitbrengt. In het nieuwe wetsvoorstel van de regering krijgt de CTIVD naar verwachting een rol als zelfstandige externe klachtenbehandelaar, hetgeen in lijn is met de Europese en internationale jurisprudentie hierover (Loof et al. 2015: 28-37). Het voordeel van deze vormen van klachtbehandeling is dat ze vaak minder strikte regels en lagere kosten kennen, en sneller tot uitspraken (kunnen) komen (FRA 2015). Tegelijkertijd is het aantal organisaties dat klachten in behandeling neemt zeer beperkt en ontbreekt de bevoegdheid om bindende oordelen te vellen.

De toetsing van Big Data-toepassingen is in de eerste plaats een verantwoordelijkheid van de wetgevende macht en van het parlement, ondersteund door adviezen van de Autoriteit Persoonsgegevens, de Raad van State en de Algemene Rekenkamer. Hierbij klinkt de stem van de burgers indirect, en bemiddeld door politieke instanties, door in het toetsingsproces. Omdat veel van de grote dataverwerkingsprojecten het individu overstijgen in aard en omvang is het belangrijk om de positie van NGO's en burgerrechtenorganisaties te versterken en te verstevigen in de toetsing van Big Data-toepassingen. Hierdoor kunnen burgers ook meer direct of via belangenorganisaties stem geven aan hun belang bij vrijheid en veiligheid.

5.8 CONCLUSIE

Veiligheid is binnen de bestaande juridische kaders voor gegevensverwerking een belangrijke beperkingsgrond van het recht op privacy. De overheidsorganisaties die op dit terrein actief zijn, hebben een ruimere bevoegdheid om gegevens te verzamelen. Ook is, afhankelijk van de sector en de doelstelling van de desbetreffende organisatie, een kleiner of groter deel van hun werkzaamheden aan geheimhouding onderworpen. Sommige van deze organisaties, zoals de politie en de

inlichtingendiensten, vallen niet onder de Wet bescherming persoonsgegevens, maar hebben hierop toegesneden eigen kaders voor gegevensverwerking en -bescherming.

Big Data-toepassingen kunnen aan de algemene vrijheid en veiligheid bijdragen mits er voldoende waarborgen zijn ingebouwd tegen verkeerd gebruik en verkeerde interpretaties van gegevens. Ook zullen deze gegevens en het gebruik daarvan goed moeten zijn beveiligd tegen manipulatie door criminele organisaties. Big Data-toepassingen stellen echter ook centrale uitgangspunten van de juridische kaders, in het bijzonder doelbinding en noodzakelijkheid, op de proef. Immers, deze toepassingen gaan uit van grootschalige gegevensverzameling en de waarde ervan ligt vooral in secundair gebruik. Naarmate meer data in samenwerkingsverbanden en informatienetwerken belanden, ontstaat bovendien een vermenging van juridische kaders. Daardoor wordt steeds onduidelijker welk regime van toepassing is op Big Data-analyses en de daaruit verkregen informatieproducten en kennis.

De juridische kaders voor gegevensverwerking – binnen en buiten het veiligheidsdomein – zijn voornamelijk gericht op het moment dat gegevens worden verzameld en opgeslagen. Op dat moment moet onder andere het doel worden bepaald voor het verwerken van gegevens en dient te worden bekeken of de gegevensverzameling legitiem is. De fase van uitwisseling, koppeling en analyse van data is in vergelijking met de regels omtrent de verzameling van gegevens minder streng gereguleerd en het toezicht daarop is minder zwaar. Dit geldt behalve voor de juridische regimes waaronder politie en inlichtingen- en veiligheidsdiensten vallen, ook voor de Wbp, waaronder de overheidsorganisaties vallen die actief zijn op het terrein van openbare orde en fraudebestrijding. Wanneer de uitwisselings-, koppelings- en analysefase wint aan belang – en dat ligt in de lijn der verwachting – kan dit de effectiviteit van de opsporing vergroten, maar ook nieuwe risico's voor de bescherming van fundamentele rechten opleveren. Het nu volgende hoofdstuk werkt op systematische wijze een aantal voorstellen uit om deze beide aspecten van Big Data in goede banen te kunnen leiden. De belangrijkste aanbeveling daarbij is om de regulering van de gegevensverwerking uit te breiden van de fase van het verzamelen naar de fase van de analyse en de fase van het gebruik.

NOTEN

1 Bij deze laatste categorie moet bijvoorbeeld worden gedacht aan de burgemeester die op grond van artikel 172 van de Gemeentewet de verantwoordelijkheid heeft voor de bescherming van de openbare orde en veiligheid.

2 Deze eis vloeit wat de opsporing betreft ook voort uit het strafvorderlijk legaliteitsbeginsel.

3 Zie onder andere ECLI:NL:GHSGR:2003:AF7798, r.o 2.4; ECLI:NL:HR:2006:AV4122 en ECLI:NL:HR:2011:BP7544

4 Rechtbank Den Haag, 23 juli 2014, ECLI:NL:RBDHA: 2014:8966. In de zaak dient inmiddels hoger beroep. Verder heeft het Europees Hof voor de Rechten van de Mens op verzoek van de eisers de zaak gevoegd met een vergelijkbare zaak die tegen het Verenigd Koninkrijk aanhangig is gemaakt (Application no. 58170/13, Big Brother Watch and others v. the United Kingdom, 4 september 2013).

5 Zie bijvoorbeeld Goodwin v. United Kingdom, Application no. 17488/90.

6 Zie in dit kader bijvoorbeeld ook de opinie van de Artikel 29-Werkgroep over het concept persoonsgegeven (advies 4/2007 over het begrip persoonsgegeven goedgekeurd op 20 juni 2007, 01248/07/NL WP 136).

7 Zie in dit kader ook de nieuwe definitie van persoonsgegevens in de Algemene Verordening Gegevensbescherming en overweging 23 daarbij: "To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by any other person to identify the individual directly or indirectly".

8 Voorstel voor een richtlijn van het Europees Parlement en de Raad betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens door bevoegde autoriteiten met het oog op de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, en betreffende het vrije verkeer van die gegevens (Interinstitutioneel dossier: 2012/0010 (COD)).

9 Roman Zakharov v. Russia App, nr. 47143/06 ECtHR, 4 december 2015.

10 Gerechtshof Den Haag, Privacy First e.a. v. Nederlandse Staat, 200.087.09-01, 2014.

11 Hoge Raad, Privacy First e.a. v. Nederlandse Staat, ECLI:NL:HR:2015:1296, 22-05-2015.

12 Met de inwerkingtreding van de Algemene Verordening Gegevensbescherming gaan toezichthouders als de AP ook klachten van burgers ontvangen en in behandeling nemen.

6 CONCLUSIES EN AANBEVELINGEN

6.1 INLEIDING

Wat kan de betekenis van Big Data voor het veiligheidsbeleid zijn? Dat is de vraag die de WRR heeft verkend en geanalyseerd in antwoord op de adviesaanvraag 'Big Data, privacy en veiligheid' van de regering. Hiertoe zijn in de inleiding de volgende onderzoeksvragen geformuleerd:

1. Wat is Big Data?
2. Hoe en in welke mate manifesteert Big Data zich in het (Nederlandse) veiligheidsdomein?
3. Wat levert een sterkte-zwakteanalyse van Big Data en veiligheid op?
4. Waar liggen de spanningen tussen de huidige (juridische) kaders voor gegevensverwerking in het veiligheidsdomein en de ontwikkeling van Big Data?
5. Wat is nodig om het beleid en het (juridische) kader zo te organiseren dat Big Data verantwoord ingezet kan worden, fundamentele rechten van burgers beschermd zijn en een goede verhouding tussen transparantie en effectiviteit van het veiligheidsbeleid gevonden wordt?

De nu volgende paragrafen geven antwoord op deze vragen, daarbij deels teruggrijpend op bevindingen in de voorgaande hoofdstukken. De voornaamste conclusie van dit rapport is dat de huidige wet- en regelgeving in een Big Data-tijdperk niet langer volstaat. Deze richt zich overwegend op de fase van gegevensverzameling, terwijl de belangrijkste kansen en risico's van Big Data juist liggen in de twee fasen die daarop volgen: de gegevensanalyse en het gebruik daarvan. Daarom schetst dit slothoofdstuk een nieuw regulerend kader voor het gebruik van Big Data in het veiligheidsdomein, waarbij het zwaartepunt van de regulering komt te liggen bij de analyse en het gebruik van gegevens. Deze verzwaring is noodzakelijk om te voorkomen dat het gebruik van Big Data een risico in plaats van een hulpbron voor een vrije samenleving gaat vormen.

6.2 BIG DATA EN VEILIGHEID

Er is sprake van een revolutie in de bestaande technologische en statistische technieken om grote hoeveelheden data snel en efficiënt te verwerken en te analyseren. Deze Big Data-revolutie doet zich nog weinig voor in de bredere samenleving en het veiligheidsdomein, maar in het commerciële domein zijn de ontwikkelingen al verder. Dit zal de komende jaren echter gaan veranderen. Wat verstaan we onder Big Data (vraag 1) en hoe manifesteren Big Data-toepassingen zich in het veiligheidsdomein (vraag 2)?

6.2.1 WAT IS BIG DATA?

Wat Big Data is, blijkt niet zo eenvoudig te definiëren. Dit komt door een gebrek aan consensus over wat de doorslaggevende kenmerken ervan zijn. De meeste definities van Big Data verwijzen naar de alomtegenwoordige drie V's (Laney 2001). Daarbij staat de eerste V voor *Volume* (het gebruik van grote hoeveelheden data), de tweede voor *Variety* (het gebruik van verschillende databronnen die in verschillende structuren of zelfs ongestructureerd zijn opgeslagen) en de derde staat voor *Velocity*, oftewel de snelheid van de dataverwerking (data worden vaak realtime geanalyseerd). Een aantal auteurs voegt meer V's aan dit drietal toe, zoals Veracity (IBM 2015; Klous 2016), Variability (Hopkins en Evelson 2011; Tech America Foundation 2012), Value (Dijcks 2012; Dumbill 2013) en Virtual (Zikopoulos en Eaton 2011; Akerkar et al. 2015).

De WRR gebruikt enkele belangrijke kenmerken van Big Data als leidraad voor dit rapport. Deze zijn gegroepeerd rondom een drietal hoofdaspecten: data(verzameling), analyse(technieken) en gebruik (tabel 6.1).

Tabel 6.1 Kenmerken van Big Data

Data	• Omvang van de data: het gaat om grote hoeveelheden gegevens. • Structuur van de data: het kan gaan om gestructureerde of ongestructureerde gegevens of een combinatie van beide. • Variëteit van de data: het gaat om de combinatie van verschillende databronnen.
Analyse	• Methode van analyse: de analyse is *data-gedreven*, er wordt dus gezocht naar patronen in de data zonder vooraf opgestelde hypothesen. • Oriëntatie van de analyse: hoewel Big Data-analyses ook inzicht kunnen geven in het verleden (retrospectieve analyses), zijn het met name de analyses van het heden (*realtimeanalyses / nowcasting*) en de toekomst (*predictive analyses / forecasting*) die de aandacht trekken.
Gebruik	• Ontschotting van domeinen: data uit het ene domein worden gebruikt voor beslissingen in het andere domein. • *Actionable knowledge**: conclusies op geaggregeerd niveau kunnen worden toegepast voor beslissingen op groeps- of individueel niveau (persoon of object).

* De term is van Degli Esposti (2014).

De WRR beschouwt Big Data als het samenspel van deze kenmerken en niet zozeer als een vastomlijnd en definieerbaar object. Er zijn weinig praktijkvoorbeelden waarbij al deze kenmerken tegelijkertijd aanwezig zijn. In de meeste gevallen gaat het om een combinatie van een aantal van deze kenmerken en zelden om de volle breedte daarvan. Het zijn voornamelijk de mogelijkheden die in de nabije toekomst binnen handbereik liggen die bepaalde datasystemen tot Big Data-systemen maken.

Big Data is tot op zekere hoogte het resultaat van de evolutie van steeds meer geavanceerde technologische en statistische technieken om data te verwerken en te analyseren om beter inzicht te krijgen in de samenleving. In een aantal opzichten is Big Data tegelijk ook een revolutie, met name op de volgende punten:

– *Exponentiële groei van data.* Hoewel de afgelopen twee eeuwen steeds meer data beschikbaar kwamen, is die groei nauwelijks te vergelijken met de huidige explosie van data. Hierbij komen tevens de verbeterde mogelijkheden om ongestructureerde en ongelijksoortige data (zoals video's, foto's en boeken) in één analyse te verenigen. Data kunnen daarbij het product zijn van gerichte verzameling ('directed'), van geautomatiseerde processen ('automated'), of van vrijwillige verstrekking ('volunteered') (Kitchin 2014b: 87-98). Vooral de hoeveelheid data uit de laatste twee categorieën is met de komst van slimme apparaten, sociale media en digitale transacties exponentieel toegenomen. Er is een 'dataficatie' van alledaagse handelingen waarneembaar, waarbij de dataverzameling vrijwel ongemerkt plaatsvindt, buiten de controle en toestemming en vaak zelfs het besef van het individu om (PCAST 2014; Zuiderveen Borgesius 2015; Schneier 2015).
– *Een scherper beeld.* De toegenomen hoeveelheid data zorgt voor een kwalitatieve sprong in de analyse. Dit wordt wel 'the unreasonable effectiveness of data' genoemd: matig presterende algoritmen hebben met hele grote hoeveelheden data betere uitkomsten dan betere algoritmen met kleinere hoeveelheden data. Bij steekproeven kunnen vaak alleen over de gehele populatie of grote subgroepen daarvan uitspraken worden gedaan, omdat bij verdere specificatie de groep snel te klein wordt om nog statistisch significante uitspraken te doen. Daarentegen kan Big Data een veel gedetailleerder beeld leveren. De kracht van Big Data ligt in de grote nauwkeurigheid van het beeld op geaggregeerd niveau. Het vertalen van die inzichten naar kleinere groepen of zelfs individuen – iets dat in het veiligheidsbeleid vaak gewenst is – gaat wel met haken en ogen gepaard.
– *Data-gedreven analyse.* Een belangrijk kenmerk van Big Data is de data-gedreven analyse, een heel andere benadering dan de traditionele statistische methode. Het doel van de analyse is niet het toetsen van hypotheses maar het vinden van interessante verbanden en patronen, waarvan de gedachte is dat ze relevant zijn voor commerciële doelen of doelen van de overheid, zoals veiligheid. Dergelijke analyses kunnen interessante en onverwachte correlaties en inzichten opleveren. Maar de causaliteit is – zeker in eerste aanleg – niet duidelijk en het risico bestaat dat correlaties tot causaliteiten 'verheven' worden.
– *Wegvallen van grenzen tussen domeinen.* Het laatste 'nieuwe' aspect aan de huidige ontwikkelingen is het wegvallen van de grenzen tussen domeinen, binnen overheden en tussen publieke, private en 'sociale' bronnen. De kracht

van Big Data ligt in het wegvallen van deze domeingrenzen, want dat levert een rijker beeld op van de werkelijkheid. Maar juist op dit punt rijzen ook vragen, die verderop ter sprake komen.

Wanneer al deze kenmerken van Big Data in volle omvang aanwezig zijn, is sprake van een revolutie. Momenteel geldt echter – zeker binnen de overheid – dat Big Data nog slechts een beperkte rol speelt. Hoewel de revolutie dus nog even op zich laat wachten, vinden intussen wel degelijk belangrijke veranderingen plaats. Binnen het domein van de dataverzameling tellen de verschillende kwantitatieve sprongen – de verzameling en opslag van veel, gevarieerde en ongestructureerde data – op tot een kwalitatieve sprong. Binnen het domein van de analyse komen onder de vleugels van langer bestaande technieken en algoritmen, nieuwe methoden van analyse op, zoals zelflerende algoritmen en *machine learning*. Deze methoden gelden als potentiële *game changers*, zeker als deze in de toekomst de norm voor veel toepassingen worden. De impact op het gebruik van Big Data hangt af van antwoorden op vragen als: wordt Big Data gebruikt om het verleden te begrijpen of om de toekomst te voorspellen? En wat is de impact van beslissingen op basis van Big Data-analyses op het dagelijks leven van burgers en consumenten? Binnen het domein van veiligheid is deze laatste vraag uiteraard van groot belang.

6.2.2 BIG DATA IN HET VEILIGHEIDSDOMEIN

Hoe en in welke mate Big Data-toepassingen zich in het (Nederlandse) veiligheidsdomein manifesteren, is door geheimhouding en het experimentele karakter van sommige toepassingen niet gemakkelijk in kaart te brengen. De casestudies in dit rapport geven de indruk dat er – met uitzondering van de inlichtingen- en veiligheidsdiensten, de Belastingdienst en fraudebestrijding – nog weinig toepassingen zijn die aan alle kenmerken van Big Data voldoen. Dat komt vooral doordat de omschakeling naar nieuwe werkwijzen tijd en geld kost. Tussen traditionele werkwijzen en nieuwe technologie in worden intussen combinaties van traditioneel, door menselijke inzichten geleid rechercheren en benutting van Big Data beproefd. Ondanks de grootschalige preparatie en koppeling van databases lijken veel organisaties echter nog niet de keuze te hebben gemaakt om omvangrijke, data-gedreven analyses uit te voeren.

Een duidelijk voorbeeld van data-gedreven analyses is de infobox Crimineel en Onverklaard Vermogen (iCOV). Deze organisatie brengt data van verschillende overheidsorganisaties bijeen om het aanpakken van crimineel vermogen makkelijker te maken. Nu is het zo dat de aangesloten organisaties een concreet verzoek kunnen doen om een verdacht persoon of bedrijf na te trekken. Voor de iCOV dient zich daarnaast de mogelijkheid aan om in de eigen omvangrijke databanken voortaan op basis van datamining en profilering potentiële fraudeurs op te sporen. Op deze wijze hoopt men veel meer fraudeurs op het spoor te komen. Wel twijfelt

men over wat er wel en niet is toegestaan binnen de huidige juridische kaders met betrekking tot gegevensverwerking. Een ander voorbeeld is de Inspectie SZW (2014), die al een stap verder is in het implementeren van nieuwe analysetechnieken. Onder het kopje *Gerichter inspecteren met datamining* stelt de inspectie in haar jaarverslag 2014 dat zij innovatieve technieken heeft verkend die het mogelijk maken om met datamining risicovolle bedrijven te selecteren. Op basis van de zeer positieve ervaringen is besloten om deze technieken ook in 2015 toe te passen om inspecties nog gerichter te kunnen uitvoeren.

Wanneer we deze ontwikkelingen doortrekken naar de nabije toekomst, zijn er ingrijpende effecten te verwachten in de verzameling, analyse en gebruik van data binnen het veiligheidsdomein. Een aantal van die effecten is nu al waarneembaar.

Data: verdergaande koppeling, grensvervaging en nieuwe databronnen
Op het punt van de verzameling van data zijn er drie ontwikkelingen gaande: de verdergaande koppeling van gegevensbestanden, de vervaging van grenzen tussen data uit publieke en private bronnen, en een sterke toename van nieuwe databronnen.

– De koppeling van gegevens neemt een hoge vlucht. Het zijn immers vaak niet de gegevens zelf die waardevol zijn, maar de koppeling daarvan, en dan vooral de koppeling van grote hoeveelheden gegevens. In samenhang hiermee zal ook het organiseren van de gegevensverzameling en -uitwisseling in het veiligheidsdomein de komende jaren sterk van karakter veranderen. Een toenemend aantal organisaties zal zich willen aansluiten bij bestaande en nog op te richten samenwerkingsverbanden die gegevens uitwisselen en daarop analyses laten uitvoeren. Daarbij gaat het zowel om gegevensuitwisseling met private partijen en het vorderen van gegevens als om het aankopen van data op de private markt (Werkgroep Verkenning kaderwet gegevensuitwisseling 2014; Kitchin 2014b).
– Private dataverzamelingen spelen steeds vaker een rol in de analyses binnen het veiligheidsdomein, als aanvulling op de data uit overheidsbronnen. Die private gegevens zijn te vorderen met inachtneming van de procedurele eisen van het Wetboek van Strafvordering, maar kunnen ook met medewerking van de betrokken organisaties worden geraadpleegd, zonder dat ze daarna nog worden opgeslagen. Aldus verandert de aard en herkomst van gegevens. Relatief harde data (financiële gegevens, en allerhande registraties) kunnen in de analyses worden gekoppeld aan zachtere, meer sociale gegevens. Ook komen er steeds meer data over personen beschikbaar als bijproduct van digitale processen en sensoren. Het onderscheid tussen persoonsgegevens en niet-persoonsgegevens wordt in een Big Data-wereld steeds meer betekenisloos nu op basis van verschillende datapunten relatief eenvoudig een persoon te 'construeren' is.

– In toenemende mate worden binnen het veiligheidsdomein (Big) Data over niet-verdachte mensen verzameld en geanalyseerd als gevolg van het gebruik van steeds grotere en meer diverse gegevensbestanden.

Analyse: realtime en geautomatiseerd

Organisaties in het veiligheidsdomein zullen meer gebruik gaan maken van informatie-gedreven methodes en taakinvullingen (Galdon Clavell 2016; WRR 2011). Deze verschuiving wordt veroorzaakt door groeiende zorgen over de veiligheid, nieuwe technologische mogelijkheden, bezuinigingen en de wens tot efficiënter werken. Het is te verwachten dat data-analyses in belang toenemen, steeds vaker realtime zijn en – naarmate de dataomvang groeit – in toenemende mate een geautomatiseerd karakter krijgen.

– Instrumenten en technieken die eerder in het kader van nationale veiligheid werden ingezet, zullen steeds vaker ook bij misdaadbestrijding en toezicht worden gebruikt.
– Big Data-analyses worden vaak gebruikt in toezicht (surveillance) en daaraan voorafgaand onderzoek voor gerichtere controles, terwijl de juridische inkadering daarvan onderontwikkeld is in vergelijking met het strafrecht, dat een redelijk vermoeden van schuld als voorwaarde stelt voor het verwerken van gegevens.
– Data-analyses worden steeds complexer en spelen een belangrijkere rol bij het nemen van beslissingen in het veiligheidsbeleid. Dit betekent dat de inzet van algoritmen en *machine learning* groter zal worden. Zeker naarmate er meer 'zachte', sociale data aan de mix van gegevens toegevoegd worden, stelt dat hogere eisen aan de kwaliteit van deze algoritmen.
– Correctieprocedures – in de software zelf of extern door middel van transparantie, audits of toezicht – zullen een groter gewicht in de schaal gaan leggen. Hebben overheidsorganisaties voldoende kennis in huis om de kwaliteit van de analysetechnieken te bewaken en hebben toezichthouders daar voldoende zicht op? Een complicatie in het veiligheidsdomein zijn de geheimhoudingsvereisten, die moeten voorkomen dat verdachten zich kunnen wapenen tegen de gehanteerde opsporingsmethoden en/of deze kunnen manipuleren. Het legitieme belang van het voorkomen van 'gaming the system' staat op gespannen voet – maar niet op onoverkomelijke wijze – met het organiseren van effectief toezicht en de bescherming van fundamentele rechten.

Gebruik: voorspellend, maar met welke doel?

Veiligheidsinstanties passen Big Data-analyses momenteel toe als aanvulling op de bestaande methoden van onderzoek, opsporing en surveillance. Zij kunnen daardoor hun interventiemiddelen efficiënter inzetten. Ook kunnen zij de dienstverlening aan burgers erdoor verbeteren. Naar verwachting zullen Big Data-analyses in de toekomst bij besluitvormingsprocessen een groeiend gewicht in de

schaal gaan leggen. Hoewel de toekomstige inzet van Big Data uiteindelijk afhangt van politieke besluitvorming, is ook het maatschappelijk vertrouwen van burgers van invloed op de gekozen doelen.

– Big Data-analyses – en daarmee statistische waarschijnlijkheden – zullen de komende jaren een grotere rol in de besluitvorming spelen en zullen vaker sturend zijn voor interventies. Deze ontwikkeling moet afgewogen worden tegen de foutmarges die we bereid zijn te accepteren. Wat is de ruimte voor en noodzaak van toetsing en menselijke afwegingen?
– Voorspellende analyses kunnen verschillende vormen aannemen (Kerr en Earle 2013). Ze zijn te gebruiken om mensen te helpen een juiste keuze te maken (*consequential*), onze voorkeuren in kaart te brengen (*preferential*), of om keuzemogelijkheden in te perken (*pre-emptive)*. Big Data-analyses worden binnen het veiligheidsdomein hoofdzakelijk ingezet voor surveillance en opsporing (*pre-emptive*) en nog weinig om burgers te helpen om beter toegerust te zijn voor de omgang met mogelijke risico's en dreigingen (*conseqential*).

6.3 STERKTE-ZWAKTEANALYSE VAN BIG DATA EN VEILIGHEID

Het gebruik van Big Data in het veiligheidsdomein kent sterke en zwakke kanten (vraag 3). Voor specifieke vormen van criminaliteit (zoals fraude) heeft Big Data positieve resultaten laten zien, al ontbreekt in veel gevallen een betrouwbare onderbouwing en evaluatie van de effectiviteit van de gebruikte analysemethoden. Tegelijkertijd is Big Data geen wondermiddel en kent Big Data diverse beperkingen die het gebruik ervan aan banden zouden moeten leggen. Daarnaast zijn er praktische en organisatorische hindernissen te overwinnen alvorens Big Data vruchten zal kunnen afwerpen. Tot slot brengt het gebruik van Big Data-oplossingen ook risico's met zich mee, die de legitimiteit en het draagvlak van Big Data-oplossingen kunnen maken en breken.

Kansen

De beschikbaarheid van veel meer data en de verfijning van analysetechnieken bieden overduidelijk kansen voor verbetering van het veiligheidsbeleid, mits het gebruik hiervan niet zelf een oorzaak van onveiligheid wordt:

– Big Data kan bijdragen aan hogere operationele efficiëntie. Er is vooral winst te behalen in organisaties die zich bezighouden met de verzameling en analyse van data en informatie. Analyses die voorheen soms dagen, weken of maanden duurden, kunnen door middel van Big Data-analyses in enkele minuten of uren worden uitgevoerd of feiten aan het licht brengen die anders een onvindbare speld in een hooiberg zouden blijven.

- Big Data maakt scherpere risicoanalyses mogelijk. Dit komt enerzijds door de grotere omvang en diversiteit van de gebruikte databestanden. Anderzijds zijn Big Data-analysemethoden gericht op het 'ontdekken' van onverwachte verbanden, die in risicoprofielen kunnen worden verwerkt. Dit resulteert onder meer in gerichtere inspecties en in een efficiëntere inzet van schaarse middelen, zoals ambulances, mobiele eenheid of politiehelikopters waar en wanneer ze de grootste kans hebben om effect te sorteren.
- Big Data-analyses kunnen helpen bij het reconstrueren van gebeurtenissen in het verleden (bijvoorbeeld direct na aanslagen). Ook kunnen Big Data-analyses nuttig zijn bij het realtime volgen van ontwikkelingen in het heden. Dit kan bijvoorbeeld van grote waarde zijn na een ramp of voor *crowd control* bij evenementen. Het is dan belangrijk om snel een beeld te krijgen van de situatie ter plaatse, zodat hulp geboden kan worden of kan worden ingegrepen bij dreigende situaties.
- Met Big Data-analyses kunnen zelfs voorspellingen worden gedaan. In het veiligheidsdomein heersen hoge verwachtingen rond het voorspellen van de tijd en plaats waar een verhoogd risico op misdaad is en van het identificeren van toekomstige daders en slachtoffers. Deze kennis maakt het mogelijk om preventief op te treden en individuen en organisaties te waarschuwen voor potentiële risico's. Hiernaast kunnen voorspellingen de pakkans van criminelen vergroten, door hun handelen inzichtelijker te maken.

Beperkingen

Er zijn van Big Data geen wonderen te verwachten. Big Data-oplossingen zijn niet voor alle veiligheidsproblemen even bruikbaar en/of geschikt en kennen bovendien inherente tekortkomingen.

- De *juiste* data zijn – ook in een digitale wereld – niet altijd beschikbaar. Soms zijn deze er domweg niet, soms zijn er problemen met bewaartermijnen en soms blijken verschillende dataplatformen niet interoperabel met elkaar. Ook de kwaliteit van data is geen gegeven, data kunnen verouderd, gecorrumpeerd of zelfs gemanipuleerd zijn. Vervalsingen of misleidingen kunnen doorwerken in Big Data-analyses en zo de zorg voor de veiligheid ondermijnen.
- De *bias* die elke dataset in meerdere of mindere mate kenmerkt, kan tot uitkomsten leiden die stelselmatig bepaalde groepen bevoor- of benadelen. Behalve door gebiaste datasets kan de oorzaak van dit effect ook in de gebruikte algoritmes liggen.
- Datamining – de voor Big Data kenmerkende analysevorm – is niet voor alle vormen van misdaadbestrijding even geschikt. Datamining is voor het voorkomen van terroristische aanslagen waarschijnlijk een ineffectieve methode. Patroonherkenning werkt het beste bij overtredingen die een vast en

terugkerend patroon laten zien. Omdat elke terroristische aanslag uniek is, is het nagenoeg onmogelijk om een goed profiel te maken. In combinatie met een gering aantal aanslagen levert dit te hoge foutpercentages op.
– Een basale, maar niet te onderschatten voorwaarde voor Big Data-analyses is voorts dat de resultaten kloppen. Maar zelfs geavanceerde algoritmes kennen hun beperkingen. Bij de analyse van grote hoeveelheden data zijn de resultaten dan ook niet automatisch correct. Correlaties zijn zelden voldoende informatie om conclusies op te baseren. Big Data-analyses vereisen vrijwel standaard nader onderzoek.
– Big Data-analyses zijn per definitie op historische data en datapatronen gebaseerd, die naar hun aard slechts een gefragmenteerd en probabilistisch beeld van de toekomst bieden. De sociale werkelijkheid kent echter weinig patronen en wetmatigheden. Bovendien is het gebruik van profielen bij opsporing lastig, omdat criminelen hun gedrag kunnen aanpassen om opsporing te ontlopen. Een risico is ook dat mensen die hun leven hebben gebeterd, tot bepaalde risicogroepen blijven behoren, omdat ze nu eenmaal zo geregistreerd staan.

Dit soort beperkingen verdient binnen het veiligheidsdomein serieuze aandacht. Als correlatie voor causaliteit wordt aangezien, worden gemakkelijk stevige conclusies getrokken, zeker als de disciplinerende werking van bewijsvoeringsprocedures voor een rechtszaak ontbreekt. Vooral in het veiligheidsdomein kan dat zeer problematisch zijn. Een verkeerde inzet van de politie is nu eenmaal van een kwalitatief andere orde dan een verkeerde aanbeveling voor een boek door Amazon.com. Daarom kunnen de uitkomsten van Big Data-analyses in het strafrecht nooit meer dan een – potentieel belangrijk – hulpmiddel bij de opsporing zijn, maar niet de strikte eisen van bewijsvoering opzijzetten.

Ook geldt dat elke analyse gebaseerd op statistische waarschijnlijkheden zowel vals positieve als vals negatieve resultaten produceert. In het domein van de veiligheid geeft dit een duivels dilemma. Vals positieve resultaten criminaliseren onschuldige mensen, en vals negatieve resultaten laten veiligheidsrisico's onopgemerkt voortbestaan. Het verlagen van het aantal vals negatieven betekent in de regel dat het aantal vals positieven toeneemt, en ook het omgekeerde is het geval. Kortom: er zijn telkens zorgvuldige afwegingen nodig, zowel *binnen* Big Data-analyses als in de keuze *tussen* het gebruik van Big Data-oplossingen en andere middelen.

Randvoorwaarden

Om Big Data effectief en legitiem in te kunnen zetten, is het bovendien nodig om aan een aantal praktische en organisatorische randvoorwaarden te voldoen. Dat is in het veiligheidsdomein niet anders dan in andere domeinen. Ten eerste zijn flinke investeringen nodig om het potentieel van Big Data te kunnen realiseren.

Denk hierbij aan dataverzameling, expertise, mankracht, beveiliging en gebruik. Daarnaast zal de beveiliging van databases up-to-date moeten zijn, mede omdat criminele netwerken zich steeds vaker op grootschalige datadiefstal richten.

- Big Data-analyses vergen specialistische technologische kennis, die schaars is. In het veiligheidsdomein is dat probleem nog wat groter: men is op zoek naar mensen met technische kennis, met een publieke taakopvatting (en bijbehorende salariseisen) die een brug kunnen slaan tussen veiligheidsbeleid en *data science*.
- De ontvankelijkheid van 'de rest van de organisatie' is een punt van aandacht. Zonder goede inbedding van analyses in de dagelijkse praktijk – van de agent op straat bijvoorbeeld – blijven zelfs de meest veelbelovende uitkomsten zonder gevolgen.
- De beveiliging van data is een onderwerp van permanente inspanning en zorg. Naarmate meer data gedeeld, geanalyseerd en centraal opgeslagen worden, neemt het belang van een adequate beveiliging daarvan navenant toe. Grote databanken – ook die van de overheid – zijn een 'single point of failure', hetgeen grote kwetsbaarheid met zich meebrengt voor grote groepen mensen. Op die manier is Big Data ook een potentiële bron van *onveiligheid.*

Risico's

Het gebruik van Big Data in het veiligheidsdomein brengt tot slot ook risico's met zich mee:

- Big Data-analyses kunnen leiden tot een toename van sociale stratificatie door de *bias* te reproduceren en te versterken die in elke dataset zit. Zonder correctie vertaalt zich dit op termijn in discriminatie en oneerlijke behandeling van bepaalde groepen in de maatschappij (Zarsky 2016: 126-127).
- In het uiterste geval resulteren Big Data-methoden in datadeterminisme. Daarbij worden individuen beoordeeld op basis van probabilistische kennis (correlaties en inferenties) over wat ze misschien zullen doen, in plaats van wat ze daadwerkelijk hebben gedaan.
- Big Data staat op gespannen voet met individuele privacy. Daarnaast kan Big Data ook privacy als een collectieve invulling van vrijheid aantasten. Het gaat dan niet om de individuele schade van personen – volgens het principe van *individual harm* dat in het huidige recht voorop staat – maar om schade aan het fundamentele recht zelf, door de veelheid van individuele privacyschendingen.
- Big Data-oplossingen zijn gevoelig voor *function creep,* oftewel gebruik van data anders dan het doel waarvoor die data zijn verzameld. *Function creep* is in het domein van het veiligheidsbeleid een punt van zorg vanwege (a) verschillen in bevoegdheid omtrent gegevensverzameling en (b) de ingrijpende gevolgen die aan Big Data-analyses verbonden kunnen worden.

– Door de grootschalige verzameling, opslag en analyse van data door overheden, waaronder inlichtingen- en veiligheidsdiensten, en het verlies van anonimiteit op het internet kunnen mensen het gevoel krijgen dat hun privacy en ongehinderde meningsuiting in gevaar zijn, hetgeen kan leiden tot een aantasting van burgerlijke vrijheden (*chilling effects*). Dat geldt in het bijzonder voor mensen en organisaties die van groot belang zijn voor het functioneren van de democratie, zoals journalisten, schrijvers, klokkenluiders, NGO's en advocaten.
– Big Data kan tot een transparantieparadox leiden: burgers worden steeds transparanter voor de overheid, terwijl de profielen, algoritmen en methoden die overheidsorganisaties gebruiken nauwelijks transparant of navolgbaar zijn, dan wel op een of andere manier zijn gemanipuleerd. Daarmee verschuift ook het machtsevenwicht tussen staat en burger. Het geheime karakter van activiteiten in het veiligheidsdomein versterkt deze transparantieparadox. Nu via Big Data-methoden steeds grotere groepen burgers in beeld komen – naast verdachte ook niet-verdachte burgers – gaat dat steeds meer wringen (Hildebrandt 2016).

6.4 EEN REGULATIEF KADER VOOR BIG DATA

Het is noch mogelijk noch wenselijk om de ontwikkeling van Big Data tegen te houden. Hoewel Big Data in de volle breedte van de 'definitie' nog zeldzaam is, zijn de aanzetten al duidelijk zichtbaar en zal de ontwikkeling de komende jaren waarschijnlijk een hoge vlucht nemen, ook binnen het veiligheidsdomein. Daarom is het zaak om het gebruik van Big Data in goede banen te leiden. Big Data-analyses kunnen in potentie een waardevolle bijdrage leveren aan de maatschappelijke veiligheid en vrijheid. Ze moeten echter betrouwbare, robuuste kennis opleveren, die bruikbaar is voor de besluitvorming van overheidsorganisaties. Ook moeten ze rusten op een stevige wettelijke basis, die helder is over de risico's die Big Data met zich meebrengt en maatregelen bevat voor de omgang daarmee. De twee nu volgende paragrafen geven antwoord op de vierde en vijfde vraag van de onderzoeksopzet, namelijk (a) waar de spanningen liggen tussen de huidige (juridische) kaders voor gegevensverwerking in het veiligheidsdomein en de ontwikkeling van Big Data; en (b) wat er nodig is om het beleid en het (juridische) kader zo te organiseren dat Big Data verantwoord ingezet kan worden, fundamentele rechten van burgers beschermd zijn en er een goede verhouding is tussen transparantie en effectiviteit van het veiligheidsbeleid.

6.4.1 EEN WELOVERWOGEN INZET VAN BIG DATA

Allereerst is het van belang om bij de inzet van Big Data aandacht te besteden aan de hierboven genoemde kwaliteiten van Big Data-analyses. Ook moet voldaan zijn aan de voorwaarden voor een goede uitvoering en juist gebruik van Big Data-analyses, zodat deze ook daadwerkelijk een effectieve bijdrage aan de maatschappelijke veiligheid en vrijheid kunnen leveren. Hierbij is het raadzaam om Big Data

in eerste instantie daar in te zetten waar de verwachting is dat de grootste successen zijn te boeken. Veiligheidsvraagstukken met een regelmatig en terugkerend karakter, en die zich lenen voor patroonherkenning, lijken hiervoor in aanleg het meest geschikt. Daarnaast is het belangrijk de inzet van Big Data toe te spitsen op situaties waar kwalitatief hoogwaardige data beschikbaar zijn. Het gebruik van 'zachtere' gegevens vereist grotere voorzichtigheid en extra toetsing, anders wordt de kans op vals positieve en vals negatieve resultaten vergroot, wat door discriminatie extra negatief kan uitwerken voor zwakkere groepen.

Gebruik Big Data in de eerste plaats voor de analyse van veiligheidsvraagstukken die zich goed voor patroonherkenning lenen en waarover data met een hoog onderscheidend vermogen beschikbaar zijn.

Ongeacht de effectiviteit van Big Data-analyses zal hierbij telkens in overweging moeten worden genomen of de ernst en omvang van de feiten waarnaar onderzoek wordt gedaan het benutten van een gegevensverzameling rechtvaardigt. Dit geldt zowel de gegevensverzameling om profielen te construeren als de data om individuen of groepen mensen te volgen. Bovendien moet altijd de afweging gemaakt worden of een Big Data-oplossing het beste instrument voor het oplossen van een bepaald probleem is. De hype rondom data-analyse moet er niet toe leiden dat andere, in de praktijk beproefde en bewezen, instrumenten en methoden uit beeld verdwijnen.

Ten tweede is het belangrijk om de inzet van Big Data niet te beperken tot doeleinden van opsporing en surveillance, maar de uitkomsten daarvan tevens te gebruiken om burgers, private organisaties en bedrijven in te lichten over waarschijnlijke risico's en dreigingen, zodat zij zich daarop beter kunnen voorbereiden en hun gedragingen kunnen aanpassen. Dit effect – spiegelbeeldig aan de eerdergenoemde *chilling effects* – kan de maatschappelijke aanvaardbaarheid van Big Data sterk vergroten.

Zorg bij de inzet van Big Data voor een evenwichtige spreiding van doeleinden. Gebruik Big Data behalve voor repressieve doeleinden ook voor dienstverlening gericht op preventie, die rechtstreeks ten goede komt aan burgers, private organisaties en bedrijven.

6.4.2 UITGANGSPUNT VOOR REGULERING: VAN VERZAMELEN NAAR ANALYSE EN GEBRUIK

Het reguleringskader voor Big Data in het veiligheidsdomein dat de WRR in dit advies formuleert, moet aan twee randvoorwaarden voldoen. Dat kader moet enerzijds niet te stringent zijn, zodat er ruimte is voor innovatie en experiment, en moet anderzijds voldoende waarborgen scheppen voor de bescherming van indivi-

duele fundamentele rechten en maatschappelijke belangen. In zijn ideaalvorm is Big Data gebaseerd op het principe van ongerichte gegevensverzameling, alsmede op koppeling en hergebruik van al verzamelde gegevens voor andere doeleinden. Het lijkt daarom voor de hand te liggen om bij Big Data de regulering van het verzamelen, in het bijzonder de handhaving van doelbinding, te verzwaren ten opzichte van het huidige juridische kader. Dit zou echter een groot deel van de belofte van Big Data in de kiem smoren. Het principe van doelbinding moet worden gehandhaafd; artikel 8, tweede lid, van het Handvest van de grondrechten van de EU eist dit ook. Alleen op basis van specifieke bevoegdheden en met het oog op zwaarwegende doeleinden mag dit principe worden doorkruist ten behoeve van andere wettelijke omschreven doeleinden zoals bestrijding van ernstige misdrijven. In die situaties mogen data worden geraadpleegd die met een ander doel zijn verzameld. Deze mogelijke doorbreking van de doelbinding is reeds in de huidige wetgeving verankerd (zie paragraaf 5.2), maar het waarborgkarakter daarvan wordt uitgehold door de steeds bredere relevantie van bijvoorbeeld commerciële data in het kader van Big Data-analyses ten behoeve van de veiligheid. De WRR pleit daarom voor een benadering waarbij meer accent komt te liggen op de regulering van data-analyse en -gebruik. De beleidsaanbevelingen zijn echter wel een samenhangend pakket dat uit balans raakt als het als een *pick and choose*-model wordt behandeld. *Regulering van analyse en gebruik is een conditio sine qua non voor het niet zwaarder reguleren van het verzamelen van gegevens.*

Structureel gebruik van Big Data-methoden binnen het veiligheidsdomein vereist dat burgers erop kunnen vertrouwen dat overheidsorganisaties zorgvuldig met hun gegevens omspringen. De huidige juridische kaders zijn vooral gericht op het inkaderen van het verzamelen en delen van gegevens. Dit geldt voor de gegevensbeschermingsregels zoals neergelegd in het Handvest, de Richtlijn bescherming persoonsgegevens (RBP), de Wet bescherming persoonsgegevens (Wbp), Wet politiegegevens (Wpg) en de nieuwe EU-verordening. En het geldt ook voor het strafrechtelijke kader, dat een specifiek strafvorderlijk doel vereist om gegevens over personen te verzamelen en te gebruiken bij opsporing en vervolging. Daarnaast is het doelbindingsprincipe van toepassing, dat bepaalt dat gegevens niet mogen worden gebruikt voor een ander doel dan waarvoor ze zijn verzameld. Al deze bepalingen zullen ook in de toekomst in acht moeten worden genomen. Toereikend zijn deze normen echter niet. Zij moeten naar het oordeel van de raad worden aangevuld met nieuwe normen voor de analyse en het gebruik van gegevens in Big Data-processen. Alleen door extra eisen te stellen aan het toepassen van Big Data-analyses kan het vertrouwen worden gecreëerd en bevestigd dat de overheid niet sluipenderwijs penetreert in de persoonlijke vrijheid van de burgers.

6.4.3 HET VERZAMELEN VAN GEGEVENS

Zowel de bestaande als de in ontwikkeling zijnde regulering voor het verzamelen van gegevens heeft naar het oordeel van de WRR ook in het Big Data-tijdperk onverminderd een belangrijke functie. Data moeten door degene die ze verwerkt rechtmatig zijn verkregen, wat in het Big Data-tijdperk extra inspanning kan vergen. Dit geldt voor het zoeken in bronnen op het internet, voor het delen van informatie met andere overheidsdiensten, voor het verkrijgen van gegevens van buitenlandse diensten en voor gegevens van of via derde partijen. De ontvangende dienst heeft de plicht na te gaan of de verstrekkende partij deze gegevens rechtmatig heeft verkregen en of hij de ontvangen gegevens zelf had mogen verzamelen volgens zijn wettelijke bevoegdheden. Wanneer de overheid serieus met Big Data-analyses aan de slag gaat, zal ook een adequate beveiliging van de verzamelde gegevens en bijbehorende beheersystemen meer aandacht vragen, temeer daar de hoeveelheid gegevens en de opslagcapaciteit naar verwachting sterk zullen blijven groeien. Daarbij vereist de komst van kwantumcomputers, die een groot deel van de reguliere encryptiesleutels onveilig maken, de ontwikkeling van nieuwe cryptografische algoritmes (Gursüs en Preneel 2016). De bepalingen hierover in de Wbp zijn echter nogal open geformuleerd, en vele toezichthouders ervaren de governance op de informatiehuishouding van de rijksoverheid als onvoldoende (WRR 2011). Deze achterstand kan de maatschappelijke aanvaardbaarheid van Big Data-projecten sterk verminderen.

Zoals gezegd, legt Big Data ook druk op belangrijke rechtsprincipes als doelbinding en dataminimalisatie, die sterk verbonden zijn met het moment van het verzamelen van gegevens. Doelbinding bepaalt dat gegevens alleen mogen worden verzameld en opgeslagen wanneer het doel van de gegevensverwerking helder is. Met de beschikbaarheid van grote hoeveelheden data is het steeds gebruikelijker om eerst gegevens te verzamelen en pas later een schifting te maken in gegevens die bruikbaar en onbruikbaar zijn. Bovendien kan de combinatie van meerdere databases met niet-identificerende gegevens door koppeling en analyse leiden tot profielen en nieuwe persoonsgegevens (op basis van een combinatie van informatiepunten). Het principe van dataminimalisatie houdt in dat er zo min mogelijk gegevens worden verzameld en in elk geval niet meer dan nodig zijn om het specifiek geformuleerde doel te bereiken. Bovendien moeten deze gegevens weer worden verwijderd als het doel is bereikt. De logica van Big Data dat 'meer beter is' staat met deze principes op gespannen voet, net als de idee dat de waarde van Big Data-analyse zit in het secundair gebruik van data en het vinden van onverwachte – en dus niet aan een doel gebonden – toepassingen. Het moment van de gegevensverzameling als aangrijpingspunt voor regulering staat daarmee onder druk.

De WRR zet daarom bij de uitwerking van het kader in op een versterking van de regulering van de fases van de analyse en het gebruik van Big Data-processen. De raad is van mening dat er meer winst te behalen valt in de latere fasen van Big Data-processen dan in een extra intensivering van de regulering van het verzamelen van data.

6.4.4 DE ANALYSE VAN GEGEVENS

Bij de regulering van de fase van de analyse van gegevens constateert de WRR een hiaat in de regelgeving. In Big Data-processen zijn de keuzes die in de analysefase worden gemaakt (algoritmen, categorisering, wegingsfactoren etc.) van eminent belang. Juist op dit vlak kunnen zich diverse risico's voordoen, die te maken hebben met *bias* in de gegevensbestanden, discriminatie en het risico op overinclusiviteit. Op de achtergrond hiervan dreigen *chilling effects* op de vrije meningsuiting en voor beroepsgroepen als journalisten en advocaten die een vitale rol spelen in de democratische rechtsstaat. De analysefase is in het huidige regime nagenoeg ongereguleerd gebleven: 'algorithmic accountability' ontbreekt.[1] De WRR bepleit daarom een aanvullende normering in de flexibele, maar betekenisvolle vorm van een expliciete zorgplicht.

Gegeven de ruimere bevoegdheden van overheidsorganisaties op het terrein van misdaad- en fraudebestrijding en gegeven de toename in de omvang van gegevensverwerkingsprocessen moeten die organisaties een wettelijk te omschrijven *zorgplicht* in acht nemen.

Het is onwerkbaar om vooraf exact voor te schrijven waaraan de analysefase moet voldoen: dat is per geval verschillend. Wel geldt een aantal algemene vereisten voor de kwaliteit van de data en de deugdelijkheid van de gehanteerde analysemethoden:

- Overheidsdiensten moeten ervoor zorgen dat hun gegevens up to date zijn en dat hun datasets geen *bias* bevatten, een plicht die zich tevens uitstrekt tot gegevens die zij van derden verkrijgen.
- De algoritmes en methoden die bij data-analyses worden gebruikt moeten deugdelijk zijn en aan de wetenschappelijke criteria voor goed (statistisch) onderzoek voldoen.
- Ze moeten daarom voor toezicht toegankelijk zijn, wat problematisch kan zijn als het 'hart' van analysesystemen uit commerciële algoritmen bestaat, die de datadienstverleners als bedrijfsgeheim presenteren. Hierbij moeten de onderzoeksresultaten, de profielen en correlaties ook op hun merites worden gecontroleerd: de gegevensverwerkende partijen moeten duidelijk kunnen maken hoe zij tot bepaalde uitkomsten komen.

De diverse onderdelen van de zorgplicht – in feite een reeks kwaliteitsmaatstaven – zijn onderwerp van gesprek tijdens het monitoren van het analyseproces en ex-posttoezicht door de toezichthouder. De verantwoordelijkheid voor de kwaliteit van de data en de deugdelijkheid van de gehanteerde analysemethoden blijft hierbij te allen tijde liggen bij de gegevensverwerkende partij.

Gezien het grote belang van de fase van gegevensanalyse – in feite het hart van Big Data-processen – zou op dit punt bovendien ook het externe toezicht versterkt moeten worden.

Big Data-projecten en -toepassingen in het veiligheidsdomein moeten onderwerp zijn van een externe review door de toezichthouder, die in het bijzonder toeziet op de gemaakte keuzes inzake data en methode van analyse. Deze review toetst ook of er aan de zorgplicht is voldaan.

De organisatie van de review kan aansluiten bij de interne audits die ook nu al wegens artikel 33 Wpg worden gedaan en worden toegestuurd aan het College bescherming persoonsgegevens, dat 'in het maatschappelijk verkeer'[2] wordt aangeduid als Autoriteit persoonsgegevens (AP). De audits moeten een uitgebreide analyse geven van de deugdelijkheid van de gegevens, van de onderzoeksmethoden en van de gevonden resultaten. Dit moet op een jaarlijkse basis geschieden voor grote dataverwerkingsprojecten binnen de overheid, in het bijzonder binnen het veiligheidsdomein gezien de potentiële gevolgen voor (individuele) burgers. Het rapport dat aan de AP wordt toegezonden moet de autoriteit in staat stellen om een goed beeld te krijgen van het gebruik van bronnen en de gehanteerde methode. Eenzelfde plicht geldt voor de inlichtingendiensten; de instantie waaraan in dat geval wordt gerapporteerd is de Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten (CTIVD). Voor zowel de AP als de CTIVD geldt dat deze rol vereist dat de technische en statistische capaciteit en expertise binnen de organisatie versterkt dient te worden. De AP en de CTIVD rapporteren vervolgens aan de Tweede Kamer. In het geval van de inlichtingendiensten geldt daarbij: waar mogelijk aan de vaste commissies voor Binnenlandse Zaken en Defensie en uitsluitend waar noodzakelijk aan de Commissie voor de Inlichtingen- en Veiligheidsdiensten (CIVD). Daarin geven zij een nader advies over mogelijke regulering of het stellen van begrenzingen. Bij gebleken gebreken in de rapportage kunnen de AP en de CTIVD de audits aanscherpen, opvoeren en in uiterste gevallen overlaten aan een onafhankelijke derde.

Ook is het belangrijk om bij Big Data-projecten vooraf een evaluatiemoment in te plannen. Dit vanwege de potentiële impact van data-gedreven toepassingen in het veiligheidsdomein. Een andere reden is het feit dat de overheid grote ICT-projecten vaak laat doorlopen, ook als beëindiging een reële en wellicht betere optie is.

Grote dataverwerkingsprojecten binnen de overheid, vooral door de politie, inlichtingen- en veiligheidsdiensten, inspecties, de Belastingdienst en samenwerkingsorganen op het terrein van misdaad- en fraudebestrijding, moeten een horizon van 3 tot 5 jaar krijgen.

De termijn ligt tussen 3 en maximaal 5 jaar. Dit geeft projecten voldoende tijd om zich te ontwikkelen en te bewijzen, maar is ook kort genoeg om nog in een vroegtijdig stadium te kunnen ingrijpen. Bij de evaluatie van de projecten zijn er drie specifieke punten van controle:

– Ten eerste wordt nagegaan of de noodzaak voor het project nog steeds bestaat. Omstandigheden kunnen zich immers wijzigen.
– Ten tweede wordt onderzocht of het gegevensverwerkingsproces effectief was: heeft Big Data zijn belofte ingelost? Beide punten veronderstellen dat er aan het begin van het project realistische en controleerbare doelen zijn geformuleerd. De evaluatie moet verduidelijken of deze doelen wel, niet of gedeeltelijk zijn verwezenlijkt en in welke mate dat toe te schrijven is aan de Big Data-analyse. Als blijkt dat deze doelen niet of nauwelijks zijn gehaald zal het project moeten worden stopgezet. Bij beperkte positieve resultaten is een plan nodig voor herziening van het project.
– Ten derde moet de evaluatie een kosten-batenanalyse van het project bevatten. Daarin moet expliciet de proportionaliteits- en subsidiariteitstoets ten aanzien van de effecten voor persoonlijke vrijheid en veiligheid zijn meegenomen. De fundamentele rechten moeten daarbij expliciet aan de orde komen: wat heeft de inbreuk op deze rechten van burgers concreet opgeleverd? Voor deze evaluatie kan aansluiting gezocht worden bij de zogenoemde Surveillance Impact Assessments (SIA). Deze zijn opgebouwd uit vier samenhangende elementen: de impact op de individuele privacy; de impact op individuele relaties, posities en vrijheden; de impact op groepen en categorieën; en de bredere impact op de samenleving en het politieke systeem (Bennett en Bayley 2016). De evaluatie is dus aanzienlijk meeromvattend dan de conventionele Privacy Impact Assessment (PIA), zoals de aankomende Verordening die vereist.

Ook van deze drie punten zal een rapport moeten worden opgesteld. Dat rapport wordt gestuurd naar de AP respectievelijk het CTIVD, die vervolgens kunnen rapporteren aan de volksvertegenwoordiging. Het feit dat de verantwoording in beide gevallen getrapt verloopt – eerst een *trusted* externe partij in de vorm van een toezichthouder en dan openbare rapportage aan de Tweede Kamer – betekent dat 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.

6.4.5 HET GEBRUIK VAN GEGEVENS

Bij het gebruik van de gegevensanalyses in Big Data-processen is de problematiek rond *profiling* van belang. Hierbij moeten drie overwegingen in acht worden genomen. Ten eerste ligt de kracht van Big Data-analyses voornamelijk in algemene conclusies en structurele patronen; bij de toepassing van deze algemene beelden op concrete situaties en specifieke individuen bestaat altijd een mismatch, omdat een profiel altijd zowel over- als onder-inclusief is. Er zullen dan ook benchmarks moeten worden ontwikkeld voor de toelaatbare foutmarge in het toepassingsbereik. Het ligt voor de hand om deze te koppelen aan zowel het belang voor de desbetreffende dienst of organisatie als de impact op (individuele) burgers.
Ten tweede worden profielen steeds meer leidend in het maken van keuzes en beslissingen. Er is een tendens om de profielen en patronen relatief kritiekloos te volgen en computeranalyses als quasi-objectief te beschouwen.

De WRR beveelt aan om bij profiling nadere regels over toelaatbare foutmarges te stellen, het verbod op geautomatiseerde besluitvorming strikter te handhaven en alert te zijn op semi-automatische besluitvorming.

De huidige regels inzake geautomatiseerde individuele besluiten in de RBP (art. 15) en in de aanstaande Verordening worden algemeen als zwak beschouwd (Bygrave 2001; Hildebrandt 2009: Savin 2013). Nederland zou een voortrekkersrol moeten nemen en ervoor moeten zorgdragen dat geautomatiseerde besluitvorming in ieder geval verboden is en blijft. 'Computer says no' kan niet het eind zijn van het gesprek tussen overheid en burger. Ook moet er alerter worden gereageerd op semi-automatische besluitvorming, waarbij formeel een mens het besluit neemt, maar deze de facto niet afwijkt van het digitale advies. Ten derde moet er daarom voor gewaakt worden dat data-analyses en profielen niet 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.

Het principe dat de verantwoordelijkheid voor de juistheid van Big Data-processen te allen tijde bij de gegevensverwerkende partij blijft liggen, moet juridisch worden verankerd. Deze dient aan te tonen waar een beslissing op gebaseerd is en welke factoren en wegingen daarin zijn meegenomen.

6.5 TOEZICHT, TRANSPARANTIE EN RECHTERLIJKE TOETSING

Het gebruik van Big Data veronderstelt een substantiële verzwaring van het toezicht. Effectief en vertrouwenwekkend toezicht vereist op zijn beurt een grotere mate van transparantie van dataverwerkingsprocessen. Transparantie is daarbij geen doel op zich, maar is dienstbaar aan *accountability*. Ook moeten burgers en organisaties mogelijkheden hebben om de juistheid en evenredigheid van beslissingen op basis van data-analyses door overheidsinstanties, ter discussie te stellen en eventueel te laten toetsen.

Toezicht

Het huidige toezicht op gegevensverwerking laat veel te wensen over, zeker in het licht van de huidige, snelle ontwikkelingen. Zowel de AP als de CTIVD is onvoldoende toegerust voor de uitdagingen van het Big Data-tijdperk in termen van bevoegdheden, expertise en financiële middelen. Vele partijen, waaronder de CTIVD (2015c) zelf, zijn van mening dat de voorgenomen uitbreiding van bevoegdheden van de MIVD en AIVD vraagt om een significante uitbreiding van de capaciteit en expertise van de toezichthouder op alle niveaus. Voor de parlementaire CIVD, die nauwelijks eigen ondersteuning heeft, geldt dat wellicht nog sterker. Hoewel de bevoegdheden en middelen van de AP door de nieuwe Verordening werden verstevigd, is het voornamelijk aan de nationale wetgever om de bijbehorende financiële middelen, bevoegdheden en capaciteiten toe te kennen. Om onder deze omstandigheden het toezicht in goede banen te leiden, zouden ook andere toezichthouders daarin een belangrijke rol kunnen spelen. De Algemene Rekenkamer kan toezien op het punt van de rechtmatigheid en doelmatigheid van Big Data-toepassingen in het veiligheidsdomein. En de Nationale ombudsman kan burgers ondersteunen in hun relatie met de digitale overheid.

Omdat de analysefase in de toekomst steeds belangrijker wordt, moet ook het toezicht daarop worden uitgebreid. Dit geldt in het bijzonder voor de inlichtingendiensten. Er is nu een wetsvoorstel in behandeling dat hun bevoegdheden ten aanzien van dataverwerking met name op het punt van het internetverkeer sterk wil uitbreiden. Vele partijen die bijdroegen aan de internetconsultatie over dit wetsvoorstel vroegen aandacht voor de versterking van het externe toezicht op inlichtingen- en veiligheidsdiensten, dat als een zwak punt wordt gezien. Op dit punt heeft de regering de aanbevelingen van de – door de regering zelf ingestelde – commissie-Dessens (Commissie evaluatie Wiv 2002) niet opgevolgd.

De toegenomen mogelijkheden om data te verzamelen en analyseren, dienen gepaard te gaan met een versteviging van het onafhankelijke toezicht. Voor het toezicht op de inlichtingendiensten zou, gelet op de taak van de CTIVD inzake de bescherming van fundamentele rechten, de introductie van doorzettingsmacht/de mogelijkheid bindende rechtmatigheidsoordelen te vellen gepast zijn.

Transparantie

Ook zal er meer transparantie moeten komen in de dataverwerkingsprocessen van de overheid. Op dit punt is nog een wereld te winnen, want de gegevensverwerking is in veel gevallen een black box. Hier komt bij dat datasubjecten vaak simpelweg niet weten dat hun gegevens zijn verzameld en dus niet zo snel hun informatierecht zullen inroepen. De WRR begrijpt dat hier, gegeven de gevoeligheden, geen volledige transparantie over kan bestaan, omdat er het gevaar bestaat van 'gaming the system'. Desalniettemin kan er worden gewerkt met een getrapte vorm van transparantie, zoals hierboven al is voorgesteld ten aanzien van de rapportages van de politie en de inlichtingendiensten aan de AP en de CTIVD, die vervolgens weer rapporteren aan de volksvertegenwoordiging.

Ook is het wenselijk om burgers meer inzicht te geven in de frequentie van de gegevensverzameling, de doelen waarvoor dat gebeurt en – waar mogelijk – welk effect complexe gegevensanalyses scoren. Dit is van belang, of het nu om inlichtingenoperaties, criminaliteitsbestrijding, belastingkwesties of uitkeringsfraude gaat. In het veiligheidsdomein hebben sommige organisaties het wettelijke recht om (delen van) hun taak voor datasubjecten en het algemene publiek geheim te houden. De groeiende hoeveelheden data die de overheid onder de bestaande geheimhoudingsbepalingen kan binnenhalen lopen echter uit de pas met de vereiste transparantie van de gegevensverwerking. Betere informatie voorziet bovendien in een democratische behoefte; een weloverwogen discussie over het gebruik van Big Data-oplossingen vereist meer inzicht in het gegevensgebruik van overheidsorganisaties op het terrein van het veiligheidsbeleid.

De transparantie van de gegevensverwerking moet worden vergroot, en er moet een beter evenwicht komen tussen het vereiste van geheimhouding en het belang van openbaarheid over Big Data-toepassingen die aan fundamentele vrijheden raken.

Op ten minste twee niveaus is meer transparantie nodig. Een groeiend aantal organisaties binnen het veiligheidsdomein is met Big Data-toepassingen aan de slag, vooral op het terrein van fraudebestrijding. Hiervan ontbreekt een goed overzicht. Veel relevante informatie over gegevensverwerking binnen samenwerkingsverbanden staat bijvoorbeeld in convenanten en besluiten, die weliswaar openbaar maar niet erg toegankelijk zijn. Ook is een betere zichtbaarheid nodig per individueel voornemen om met Big Data-toepassingen te werken. Dat kan bijvoorbeeld door organisaties te verplichten een beleidsplan op te stellen. Daarin vermelden zij welke Big Data-toepassingen zij gebruiken om uitwerking te geven aan hun wettelijke en/of publiekrechtelijke taken, en wat de kosten en de beoogde resultaten daarvan zijn. Ook op het vlak van de verantwoording is meer mogelijk dan nu gebeurt, bijvoorbeeld in de jaarverslaglegging. Op het terrein van de inlichtingendiensten vindt een discussie plaats over de geheimhouding van met name tap-

statistieken. Diverse Europese landen betrachten een aanzienlijk grotere mate van openbaarheid over de technieken en operaties van het inlichtingenwerk, zonder dat dit het functioneren van de inlichtingendiensten noemenswaardig belemmert. In België gebeurt dat expliciet met als doel een geïnformeerde discussie te kunnen voeren over de werking, de bevoegdheden en de controle op de inlichtingen- en veiligheidsdiensten.[3]

Toetsing in het algemeen belang

Zoals eerder aangegeven zal de machtsongelijkheid tussen de burger en de overheid in verband met dataverwerkingscapaciteiten en -technieken in het Big Data-tijdperk waarschijnlijk alleen maar groter worden. Het is dus van belang om de positie van de burger te versterken. Deels gebeurt dit door meer bevoegdheden en controlemogelijkheden aan de toezichthoudende organisaties toe te kennen en de transparantie over de gegevensverwerking te vergroten; en deels gebeurt dat door de bewijslast voor de juistheid van Big Data-processen bij de gegevensverwerkende partijen te verankeren. Het is volgens de WRR echter ook van belang om de stem van de burger *zelf* in de toetsing van Big Data-toepassingen te versterken.

Dit is in de eerste plaats een verantwoordelijkheid van de wetgevende macht en van het parlement in zijn controlerende functie, ondersteund door adviezen van de Autoriteit Persoonsgegevens, de Rekenkamer en de Raad van State. Burgers kunnen echter ook direct of via belangenorganisaties stem geven aan hun belang bij vrijheid en veiligheid. In de huidige situatie is het klachtrecht sterk verbonden aan individuele schade en zijn de mogelijkheden voor collectieve procedures bij de rechter gebonden aan de criteria van artikel 3: 305a van het Burgerlijk Wetboek. Dit geeft de burger – en organisaties waarin burgers zich verenigen – te weinig mogelijkheden om besluitvorming op basis van Big Data-processen te bevragen zolang zij geen gezamenlijke persoonlijke benadeling kunnen aanvoeren. Het ontbreken in ons constitutioneel bestel van een onafhankelijke rechterlijke toetsing van wetgeving wanneer (nog) niet van persoonlijke schade is gebleken, en de onbepaaldheid van artikel 10 van de Grondwet hebben ertoe geleid dat het zwaartepunt van de rechterlijke toetsing zich verplaatst naar het Europese Hof voor de Rechten van de Mens en het Hof van Justitie van de Europese Unie. Burgers die zich zorgen maken over de maatschappelijke effecten van Big Data-toepassingen, moeten zich dus in feite tot Luxemburg wenden om Nederlandse wetgeving en beleid omtrent Big Data-toepassingen te toetsen. Dit is gezien de collectieve impact van Big Data-processen in het Nederlandse veiligheidsdomein een onbevredigende situatie.

De WRR is van mening dat de mogelijkheden voor rechterlijke toetsing van wetgeving en beleid omtrent Big Data-toepassingen verbeterd moeten worden.

Omdat veel van de grote dataverwerkingsprojecten het individu overstijgen in aard en omvang is het belangrijk om niet slechts in te zetten op individuele rechten, maar om de positie van NGO's en burgerrechtenorganisaties ook in juridische procedures te versterken en te verstevigen. Dit betekent niet dat de deur naar de rechter voor elke collectieve klacht open dient te gaan, maar dat selectief zaken worden toegelaten die recht doen aan collectieve zorgen en bijdragen aan de opbouw van jurisprudentie op dit belangrijke en relatief onontgonnen terrein (Zwenne en Schmidt 2016).

6.5.1 INTERNATIONALE EN EUROPESE CONTEXT

De WRR is er bij dit alles van doordrongen dat Big Data-analyses ook, in steeds toenemende mate, plaatsvinden buiten het bereik van de nationale staat. Wanneer activiteiten van personen en organisaties zich – zoals aan de orde van de dag is – over staatsgrenzen heen ontwikkelen terwijl negatieve gevolgen van Big Data-analyses hen vanuit het buitenland treffen (bijvoorbeeld belemmeringen bij reizen), dreigt een juridisch vacuüm te ontstaan.

De regering moet de voorbereidingen van adequate wetgeving zoveel mogelijk in EU-verband entameren. De Europese Unie zal internationale normering effectiever kunnen bevorderen dan afzonderlijke lidstaten. Daarnaast zal de Nederlandse regering kunnen bevorderen dat het onderwerp Big Data bij de Raad van Europa hogere prioriteit krijgt.

6.6 SLOT

Big Data herbergt grote beloften in zich voor toezicht, opsporing en preventie binnen het veiligheidsdomein. Big Data verkleint echter ook de afstand van de overheid tot haar burgers. Big Data-processen kunnen burgers dicht op de huid komen te zitten, ook wanneer zij onschuldig zijn en op hen geen enkele verdenking rust. De inzet van Big Data moet daarom gepaard gaan met extra maatregelen om de fundamentele rechten te beschermen. Alleen onder die voorwaarde kan Big Data een belangrijke bijdrage leveren aan de veiligheid, en daarmee aan de vrijheid van de burgers in en van Nederland.

NOTEN

1 De term is van Nicholas Diakopoulos (2013).

2 Artikel 51 lid 4 Wpg.

3 www.comiteri.be/images/pdf/Jaarverslagen/Activiteitenverslag_2014.pdf. Zie p. 70-77 voor de statistieken over inlichtingenoperaties.

LITERATUURLIJST

Adey, P. (2004) 'Secured and sorted mobilities: Examples from the airport', *Surveillance and Society* 1, 4: 500-514.

ADMA (2013) *Best practice guideline: Big Data; A guide to maximising customer engagement opportunities through the development of responsible Big Data strategies*, Sydney: ADMA.

Adviescommissie Informatiestromen Veiligheid (AIV) (2007) *Data voor daadkracht. Gegevensbestanden voor veiligheid: observaties en analyse*, Den Haag: Deltahage.

Akerkar, R., G. Lovoll, S. Grumbach, A. Faravelon, R. Finn en K. Wadhwa (2015) 'Understanding and mapping Big Data', *deliverable 1.1 BYTE project*, beschikbaar op: http://byte-project.eu/wp-content/uploads/2016/03/BYTE-D1.1-FINAL-post-Y1-review.compressed.pdf [7 februari 2015].

Algemene Inlichtingen- en Veiligheidsdienst (2015) *Jaarverslag 2014*, Den Haag.

Allen, A.L. (2011) *Unpopular privacy: What must we hide?*, New York/Oxford: Oxford University Press.

Ambrose, J. en M. Leta (2014) 'Lessons from the avalanche of numbers: Big Data in historical context', *I/S: A Journal of Law and Policy for the Information Society 2014-2015*, august 1, beschikbaar op SSRN: http://ssrn.com/abstract=2486981 [1 augustus 2014].

Anderson, C. (2008) 'The end of theory: The data deluge makes the scientific method obsolete', *Wired Magazine* 16.07, beschikbaar op: www.uvm.edu/~cmplxsys/wordpress/wp-content/uploads/reading-group/pdfs/2008/anderson2008.pdf [30 maart 2016].

Andrejevic, M. en K. Gates (2014) 'Big Data surveillance editorial', *Surveillance and Society* 12, 2: 185-196.

Article 29 Data Protection Working Party (2014) *Opinion 5/2014 on anonymisation techniques*, beschikbaar op: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [20 mei 2015].

Artikel 29-Werkgroep (2007) *Advies 4/2007 over het begrip Persoonsgegeven*, goedgekeurd op 20 juni 2007, 01248/07/NL WP 136.

Bakhshi, H., A. Bravo-Biosca en J. Mateos-Garcia (2014) *Inside the datavores: Estimating the effect of data and online analytics on firm performance*, Londen: NESTA, beschikbaar op: www.nesta.org.uk/sites/default/files/inside_the_datavores_technical_report.pdf.

Bakker, B. (2014) 'Hoe groot wordt big data?', *Accountant*, 3: 12-17, beschikbaar op: www.accountant.nl/magazines/accountant-maart-2014/hoe-groot-wordt-big-data/.

Ball, K. en F. Webster (red.) (2003) *Intensification of surveillance. Crime, terrorism and warfare in the information age*, Londen: Pluto Press.

Balzacq, T. (2008) 'The policy tools of securitization. Exchange, EU foreign and interior policies', *Journal of Common Market Studies* 46, 1: 75-100.

Balzacq, T. (2015) 'What kind of theory – if any – is secititization?', *International relations* 29, 1: 96.

Banko, M. en E. Brill (2001) 'Scaling to very very large corpora for natural language disambiguation', *ACL '01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*: 26-33.

Barnes, T.J. (2013) 'Big data, little history', *Dialogues in Human Geography* 3, 3: 297-302.

Barnes, T.J. en M.W. Wilson (2014) 'Big Data, social physics, and spatial analysis: The early years', *Big Data and Society*, April-June 2014: 1-14.

Beck, C. en C. McCue (2009) 'Predictive policing: What can we learn from Wal-Mart and Amazon about fighting crime in a recession?', *The Police Chief* LXXVI, 11, beschikbaar op: www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=1942&issue_id=112009.

Bellanova, R. en P. de Hert (2013) 'Practices and modes of transatlantic data processing: From sorting countries to sorting individuals', blz. 514-535 in *The Routledge Handbook of European Criminology*, Londen: Routledge.

Bennett, C.J. (1992) *Regulating privacy: Data protection and public policy in Europe and the United States,* Ithaca: Cornell University Press.

Bennett, C.J. en R.M. Bayley (2016) 'Privacy protection in the era of 'Big Data': Regulatory challenges and social assessments', blz. 205-227 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Bennett Moses, L. en J. Chan (2014) 'Using Big Data for legal and law enforcement decisions: Testing the new tools', *University of New South Wales Law Journal* 37, 2: 643-678.

Bongers, F., C.-J. Jager en R. te Velde (2015) *Big Data in onderwijs en wetenschap: Inventarisatie en essays*, Utrecht: Dialogic.

Boston, K. (2015) *Introducing partner audiences*, beschikbaar op: https://blog.twitter.com/2015/introducing-partner-audiences.

Boyd, D. en K. Crawford (2012) 'Critical questions for Big Data. Provocations for a cultural, technological and scholarly phenomenon', *Information, Communication and Society* 15, 5: 662-679.

Brakel, R. van (2016a) 'The rise of preemptive surveillance of children in England: Unintended social and ethical consequences', in T. Rooney en E. Taylor (red.) *Surveillance and Childhood*, Londen: Ashgate (te verschijnen).

Brakel, R. van (2016b) 'Pre-emptive Big Data surveillance and its (dis)empowering consequences: The case of predictive policing', blz. 117-141 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Broeders, D. (2007) The new digital borders of Europe EU databases and the surveillance of irregular migrants', *International sociology* 22, 1: 71-92.

Broeders, D. (2009) *Breaking down anonymity: Digital surveillance of irregular migrants in Germany and the Netherlands*, Amsterdam: Amsterdam University Press.

Broeders, D. (2011) 'A European 'border' surveillance system under construction', blz. 40-67 in H. Dijstelbloem en A. Meijer (red.) *Migration and the new technological borders of Europe*, Houndsmills, Basingstoke and Hampshire: Palgrave.

Broeders, D. en J. Hampshire (2013) 'Dreaming of seamless borders: ICTs and the pre-emptive governance of mobility in Europe', *Journal of Ethnic and Migration Studies* 39, 8: 1201-1218.

Broeders, D. en H. Dijstelbloem (2016) 'The datafication of mobility and migration management: The mediating state and its consequences', blz. 242-260 in I. van der Ploeg en J. Pridmore (red.), *Digitizing identities*, Londen: Routledge.

Brownstein, J., C. Freifeld en L. Madoff (2009) 'Digital disease detection — Harnessing the Web for public health surveillance', *New England Journal of Medicine* 360: 2153-2157.

Brynjolfsson, E., L.M. Hitt en H.H. Kim (2011) 'Strength in numbers: How does data-driven decisionmaking affect firm performance?', beschikbaar op: http://ssrn.com/abstract=1819486 of http://dx.doi.org/10.2139/ssrn.1819486.

Bygrave, L.A. (2001) 'Automated profiling: Minding the machine: Article 15 of the EC Data Protection Directive and Automated Profiling', *Computer Law and Security Review* 17, 1: 17-24.

Cao, L. (2012) 'Actionable knowledge discovery and delivery', *WIREs Data Mining and Knowledge Discovery* 2012, 2: 149-163.

Castells, M. (1996) *The rise of the network society, the information age: Economy, society and culture, Vol. I*, Oxford: Blackwell.

Cavoukian, A. en D. Castro (2014) *Setting the record straight: De-identification does work*, Information and Privacy Commissioner Ontario, Canada.

CBP (2014) *Advies conceptbesluit SyRI*, Den Haag, beschikbaar op: https://cbpweb.nl/sites/default/files/atoms/files/z2013-00969.pdf.

Citron, D.K en F.A. Pasquale (2014) 'The scored society, due process for automated predictions', *Washington Law Review* 89: 1-33.

Clarke, R., M. Morell, G. Stone, C. Sunstein en P. Swire (2013) *Liberty and security in a changing world. Report and Recommendations of The President's Review Group on Intelligence and Communications Technologies* 12 december 2013, beschikbaar op: www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf.

Commissie evaluatie Wiv 2002 (2013) *Evaluatie Wet op de inlichtingen- en veiligheidsdiensten 2002, Naar een nieuwe balans tussen bevoegdheden en waarborgen*, beschikbaar op: www.rijksoverheid.nl/documenten/rapporten/2013/12/02/rapport-evaluatie-wet-op-de-inlichtingen-en-veiligheidsdiensten-2002 [8 februari 2016].

Commissie van Toezicht betreffende de Inlichtingen- en Veiligheidsdiensten (2015a) *Jaarverslag 2014 - 2015*, Den Haag: CTIVD.

Commissie van Toezicht betreffende de Inlichtingen- en Veiligheidsdiensten (2015b) *Toezichtsrapport over de samenwerking van MIVD met buitenlandse inlichtingen- en/of veiligheidsdiensten*, Den Haag: CTIVD.

Commissie van Toezicht betreffende de Inlichtingen- en Veiligheidsdiensten (2015c) *Reactie CTIVD op het concept-wetsvoorstel Wet op de inlichtingen- en veiligheidsdiensten 20XX (consultatieversie juni 2015)*, 26 augustus 2015, Den Haag: CTIVD.

Cox, M. en D. Elssworth (1997) 'Application-controlled demand paging for out-of-core visualization', beschikbaar op: www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf.

Craig, T. en M.E. Ludoff (2011) *Privacy and big data. The players, regulators, and stakeholders*, Sebastopol: O'Reilly Media.

Crawford, K., K. Miltner en M.L. Gray (2014) 'Critiquing big data: Politics, ethics, epistemiology, special section introduction', *International Journal of Communication*, 8: 1663-1672.

Crawford, K. en J. Schultz (2014) 'Big data and due process: Toward a framework to redress predictive privacy harms' *BCL Rev.* 55, 93.

Curtin, D. (2011) *Top secret Europe*, Inaugural lecture, Amsterdam: University of Amsterdam.

Custers, B., T. Calders, B. Schermer en T. Zarsky (red.) (2013) *Discrimination and privacy in the information society - Data mining and profiling in large databases*, Heidelberg: Springer.

Daalhuijsen, T., S. Steenman en A. Meijer (2015) 'Big Data: een revolutie in gemeentelijk beleid?', *Bestuurswetenschappen* 69, 3: 6-25.

Davenport, T.H., P. Barth en R. Bean (2012) 'How 'Big Data' is different', *MIT Sloan Management Review*, fall 2012.

Davenport, T.H. en D.J. Patil (2012) 'Data scientist: The sexiest job of the 21st century', *Harvard Business Review, October 2012 issue*, beschikbaar op: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/.

Degli Esposti, S. (2014) 'When Big Data meets dataveillance: The hidden side of analytics', *Surveillance and Society* 12, 2: 209-225, beschikbaar op: www.surveillance-and-society.org.

Deibert, R. (2013) *Black code. Inside the battle for cyber space*, Toronto: Signal.

DeNardis, L. (2014) *The global war for internet governance*, New Haven: Yale University Press.

Desrosières, A. (2001) 'History of statistics', *International Encyclopedia of the Social and Behavioral Sciences*, 15080-15085.

Devlin, B., S. Rogers en J. Myers (2012) *Big Data comes of age*, An Enterprise Management Associates (EMA) and 9sight Consulting Research Report, beschikbaar op: 9sight.com/pdfs/Big_Data_Comes_of_Age.pdf.

Diakopoulos, N. (2013) *Algorithmic accountability reporting: On the investigation of black boxes*, Tow Center for Digital Journalism, beschikbaar op: http://towcenter.org/wp-content/uploads/2014/02/78524_Tow-Center-Report-web-1.pdf, [8 februari 2016].

Diebold, F.X. (2000) *Big data dynamic factor models for macroeconomic measurement and forecasting*, Discussion read to the 8th World Congress of the Econometric Society, Seattle, August, beschikbaar op: www.ssc.upenn.edu/~fdiebold/papers/paper40/temp-wc.PDF.

Dijcks, J.P. (2012) *Oracle: Big Data for the enterprise*, Oracle White Paper, beschikbaar op: www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf.

Dijstelbloem, H. en A. Meijer (red.) (2009) *De migratiemachine. De rol van technologie in het migratiebeleid*, Amsterdam: Van Gennep.

Dumbill, E. (2013) 'Making sense of big data editorial', *Big Data* 1, 1: 1-2, beschikbaar op: http://online.liebertpub.com/doi/abs/10.1089/big.2012.1503.

Economist Intelligence Unit (2012) 'The deciding factor: Big data and decision making', *Economist Intelligence Unit* commissioned by Capgemini, 4 juni, beschikbaar op: www.uk.capgemini.com/resource-file-access/resource/pdf/The_Deciding_Factor__Big_Data___Decision_Making.pdf.

Edwards, L. (2016) 'Privacy, security and dataprotection in smart cities: A critical EU law perspective', *European Data Protection Law Review*, beschikbaar op: http://ssrn.com/abstract=2711290.

Ekbia, H., M. Mattioli, I. Kouper, G. Arave, A. Ghazinejad, T. Bowman, V. Ratandeep Suri, A. Tsou, S. Weingart en C.R. Sugimoto (2015) 'Big Data, bigger dilemmas: A critical review', *Journal of the Association for Information Science and Technology* 66, 8: 1523-1545.

Eskens, A., O. van Daalen en N. van Dijk (2015) *Ten standards for oversight and transparancy of national intelligence services*, Amsterdam: IViR, beschikbaar op: www.ivir.nl/publicaties/download/1591.

European Commission (2014) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *Towards a thriving Data-Driven Economy*, COM (2014) 442, beschikbaar op: http://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy.

European Data Protection Supervisor (2014) *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, Preliminary Opinion, March 2014, beschikbaar op: https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf.

Europol (2015) *Exploring tomorrow's organised crime*, beschikbaar op: www.europol.europa.eu/sites/default/files/Europol_OrgCrimeReport_web-final.pdf.

Fairfield, J.A.T. en E. Luna (2014) 'Digital innocence', 99 *Cornell L. Rev.* 981, beschikbaar op: http://scholarship.law.cornell.edu/clr/vol99/iss5/1.

Fayyad, U., G. Piatetsky-Shapiro en P. Smyth (1996) 'From data mining to knowledge discovery in databases', *AI Magazine* 17, 3: 37-54.

Felten, E.W. (2013) 'Written Testimony of Edward W. Felten, Professor of Computer Science and Public Affairs, Princeton University', United States Senate, Committee on the Judiciary Hearing on Continued Oversight of the Foreign Intelligence Surveillance Act October 2, 2013, beschikbaar op: www.cs.princeton.edu/~felten/testimony-2013-10-02.pdf.

Floridi, L. (2012) 'Big Data and their epistemological challenge', *Philosophy and Technology* 25, 4: 435-437.

FRA (2015) 'Surveillance by intelligence services: Fundamental rights safeguards, and remedies in the EU. Mapping Member States' legal frameworks', beschikbaar op: http://fra.europa.eu/sites/default/files/fra_uploads/fra-2016-surveillance-intelligence-services_en.pdf.

Fukuda-Parr, S. en C. Messineo (2011) 'Human security: A critical review of the literature', *CRPD Working Paper* No. 11, https://soc.kuleuven.be/web/files/12/80/wp11.pdf.

Galdon-Clavell, G. (2016) 'Policing, big data and the commodification of security', blz. 89-115 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Gandy, O.H. jr. (1993) *The panoptic sort: A political economy of personal information*, Boulder: Westview Press.

Gandy, O.H. jr. (2009) *Coming to terms with chance: Engaging rational discrimination and cumulative disadvantage*, Farnham and Burlington: Ashgate.

Gillespie, T. (2014) 'The relevance of algorithms', in T. Gillespie, P. Boczkowski en K. Foot (red.) *Media technologies: Essays on communication, materiality, and society*, Cambridge, MA: MIT Press.

Glennon, M. (2014) 'National security and double government', *Harvard National Security Journal* 5, 1: 1-114.

González Fuster, G. en A. Scherrer (2015) 'Big Data and smart devices and their impact on privacy. Document requested by the committee on civil liberties, justice and home affairs', Brussel, beschikbaar op: www.europarl.europa.eu/RegData/etudes/STUD/2015/536455/IPOL_STU(2015)536455_EN.pdf.

Goodman, M. (2015) *Future crimes: A journey to the dark side of technology – and how to survive it*, Londen: Transworld Publishers.

Graaf, B.A. de (2011) 'Nationale veiligheid in Nederland: de geschiedenis van een ordeningsprincipe', blz. 199-225 in E.R. Muller (red.) *Veiligheid: Veiligheid en veiligheidsbeleid in Nederland*, Deventer: Kluwer.

Graham, M. (2010) 'Neogeography and the palimpsests of place: Web 2.0 and the construction of a virtual earth', *Tijdschrift voor Economische en Sociale Geografie* 101, 4: 422-436.

Greenfield, A. (2006) *Everyware: The dawning age of ubiquitous computing*, Boston: New Riders.

Griffioen, H. (2011) *Privacy en vormen van 'intelligente' mobiliteit: De impact van ict-applicaties voor de weg en het spoor*, Den Haag: WRR, beschikbaar op: www.wrr.nl/publicaties/publicatie/article/privacy-en-vormen-van-intelligente-mobiliteit/.

Gürses, S. en B. Preneel (2016) 'Cryptology and privacy in the context of Big Data', blz. 49-86 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Hacking, I. (1990) *The taming of chance*, Cambridge: Cambridge University Press.

Hald, A. (2003) *A history of probability and statistics and their applications before 1750*, Hoboken: Wiley.

Halevy, A., P. Norvig en F. Pereira (2009) 'The unreasonable effectiveness of data', *Intelligent Systems, IEEE 24*, 2: 8-12, beschikbaar op: http://static.googleusercontent.com/media/research.google.com/nl//pubs/archive/35179.pdf.

Hand, D., H. Mannila en P. Smyth (2001) *Principles of data mining*, The MIT Press.

Harcourt, B. (2007) *Against prediction: Profiling, policing and punishing in an actuarial age*, The University of Chicago.

Harford, T. (2014) Big data: are we making a big mistake? *Financial Times*, beschikbaar op: https://next.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0#axzz3F5qexEZu.

Heide, L. (2009) *Punched-card systems and the early information explosion, 1880-1945*, Baltimore: Johns Hopkins University Press.

Hert, P. de (2005) 'Balancing security and liberty within the European human rights framework: A critical reading of the Court's case law in the light of surveillance and criminal law enforcement strategies after 9/11', *Utrecht Law Review* 1, 1: 68-96.

Hert, P. de en H. Lammerant (2016) 'Predictive profiling and its legal limits: Effectiveness gone forever?', blz. 145-173 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Hildebrandt, M. (2008) 'Defining profiling: A new type of knowledge?', blz. 17-45 in M. Hildebrandt en S. Gutwirth (red.) *Profiling the European citizen*, Dordrecht: Springer.

Hildebrandt, M. (2009) 'Who is profiling who? Invisible visibility', blz. 239-252 in S. Gutwirth, Y. Poullet, P. de Hert, J. Nouwt en C. de Terwangne (red.) *Reinventing data protection?*, Dordrecht: Springer Science.

Hildebrandt, M. (2015) *Smart technologies and the (ends) of Law. Novel entanglements of law and technology*, Cheltenham: Edward Elgar Publishing.

Hildebrandt, M. (2016) 'Data gestuurde intelligentie in het strafrecht', blz. 137-240 in E.M.L. Moerel, J.E.J. Prins, M. Hildebrandt, T.F.E Tjong Tjin Tai, G-J. Zwenne en A.H.J. Schmidt, *Homo Digitalis*, Handelingen Nederlandse Juristen-Vereniging 146e jaargang/2016-I, Wolters Kluwer, Beschikbaar op: njv.nl/wp-content/uploads/2011/04/Preadviezen-NJV-2016.pdf

Hildebrandt, M. en K. de Vries (red.) (2013) *Privacy, due process and the computational turn*, Abingdon: Routledge.

Hillard, R. (2012) 'It's time for a new definition of Big Data', *MIKE 2.0*, beschikbaar op: http://mike2.openmethodology.org/blogs/information-development/2012/03/18/its-time-for-a-new-definition-of-big-data/.

Hirsch Ballin, M. (2014) 'Big Data in de strijd tegen terrorisme', *Christen Democratische Verkenningen*, Herfst: 97-107.

Hopkins, B. en B. Evelson (2011) Expand your digital horizon with big data', *Forrester*, beschikbaar op: www.asterdata.com/newsletter-images/30-04-2012/resources/Forrester_Expand_Your_Digital_Horiz.pdf.

Horn, E. (2011) 'Logics of political secrecy', *Theory, Culture & Society* 28, 7-8: 103-122.

Human Rights Watch (2014) *Annual report*.

IBM (2015) *What is Big Data?*, beschikbaar op: www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

ICTU (2014) 'Basisregistraties als fundament voor modernisering toezicht bij Inspectie SZW', beschikbaar op: www.digitaleoverheid.nl/images/stories/stelsel_van_basisregistraties/best-practice-iszw.pdf [21 februari 2016].

IDC (2012) 'Market analysis. Worldwide big data technology and services 2012-2016 forecast', beschikbaar op: http://laser.inf.ethz.ch/2013/material/breitman/additional%20reading/Worldwide%20Big%20Data%20Technology%20and%20Services%202012-2016%20Forecast.pdf.

Information Commissioner's Office (IMC) (2014) *Big data and data protection*, version 1.0, 28 July 2014.

Inspectie Sociale Zaken en Werkgelegenheid (2014) *Jaarverslag*, Den Haag, beschikbaar op: www.inspectieszw.nl/Images/Jaarverslag-Inspectie-szw-2014_tcm335-365553.pdf [8 februari 2016].

Intrado (2012) *Intrado arms agencies with new weapon to help keep responders and communities safe*, 15 augustus, beschikbaar op: www.intrado.com/news/press/2012/intrado-arms-agencies-new-weapon-help-keep-responders-and-communities-safe.

Ioannidis, J.P.A. (2005) 'Why most published research findings are false', *PLOS Medicine* 2, 8, beschikbaar op: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124.

Jacobs, A. (2009) 'The pathologies of Big Data', *Communications of the ACM* 52, 8: 36-44.

Jacobs, B. (2016) 'Select while you collect. Over de voorgestelde interceptiebevoegdheden voor inlichtingen- en veiligheidsdiensten', *Nederlands Juristenblad* 29-1-2016, 4: 256-261.

Jagadish, H.V., J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J.M. Patel, R. Ramakrishnan en C. Shahabi (2014) 'Big data and its technical challenges', *Communications of the ACM* 57, 7: 86-94.

Johnson, S.D., K.J. Bowers, D.J. Birks en K. Pease (2009) 'Predictive mapping of crime by ProMap: Accuracy, units of analysis and the environmental backcloth', blz. 171-198 in D. Weisburd, W. Bernasco en G.J.N. Bruinsma (red.) *Putting crime in its place*, Dordrecht: Springer.

Johnston, L. en C. Shearing (2003) *Governing security; Explorations in policing and justice*, Londen: Routledge.

Kaldor, M. (2007) *Human security: Reflections on globalization and intervention*, Cambridge: Polity Press.

Kaldor, M. (2014) 'Filling the security gap. Human security, human rights and human development', in M. Martin en T. Owen (red.) *Routledge handbook of human security*, Londen: Routledge.

Kentonline (2013) PredPol software which targets crime down to small zones has slashed north Kent crime by 6%, www.kentonline.co.uk/kent/news/crime-innorth-kent-slashed-4672.

Kerr, I. en J. Earle (2013) 'Prediction, preemption, presumption: How Big Data threatens big picture privacy', *Stanford Law review* 66: 65, beschikbaar op: www.stanfordlaw review.org/sites/default/files/online/topics/66_StanlrevOnline_65_Kerr Earle.pdf [8 februari 2016].

Kerr, O.J. (2011) 'An equilibrium adjustment theory of the Fourth Amendment', *Harvard Law Review* 125: 476-543.

Keymolen, E. en D. Broeders (2013) 'Innocence lost? Care and control in Dutch digital youth care', *The British Journal of Social Work* 43, 1: 41-63.

King, G. (2016) 'Preface: Big Data is not about the data!', blz. 1-4 in R.M. Alvarez (red.) *Computational social science: Discovery and prediction*, Cambridge: Cambridge University Press.

Kitchin, R. (2013) 'Big Data and human geography: Opportunities, challenges and risks', *Dialogues in Human Geography* 3, 3: 262-267.

Kitchin, R. (2014a) 'The real time city? Big data and smart urbanism', *Geojournal* 79: 1-14.

Kitchin, R. (2014b) *The data revolution: Big Data, open data, data infrastructures and their consequences*, Londen: Sage.

Kitchin, R (2014c) Thinking critically about and researching algorithms, *The Programmable City Working Paper* 5, beschikbaar op: http://dx.doi.org/10.2139/ssrn.2515786.

Kitchin, R. en T.P Lauriault (2014) 'Small data, data infrastructures and big data', *The Programmable City Working Paper 1*, beschikbaar op: http://dx.doi.org/10.2139/ ssrn.2376148.

Kleve, P., C. van Noortwijk en P.C. van Schelven (2013) 'De rol van informatietechnologie', blz. 589-629 in E.R. Muller (red.) *Veiligheid en veiligheidsbeleid in Nederland*, Deventer: Kluwer.

Klous, S. (2016) 'Sustainable harvesting of the Big Data potential', blz. 27-47 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Koning, M.E. (2013) 'Publiek-private samenwerking in cyberspace – de gegevensvergaring', *Computerrecht* 1: 33-41.

Kool, L., J. Timmer en R. van Est (2015) *De datagedreven samenleving*, Den Haag: Rathenau Instituut.

Koops, B.J. (2013a) 'On decision transparency, or how to enhance data protection after the computational turn', blz. 196-220 in M. Hildebrandt en K. de Vries (red.) *Privacy, due process and the computational turn*, Abingdon: Routledge.

Koops, B.J. (2013b) 'Juridische kwalificatie van verkeersgegevens in het licht van artikel 13 Grondwet', beschikbaar op: https://pure.uvt.nl/portal/files/1506360/Koops_Juridische_kwalificatie_van_ verkeersgegevens_en_art_13_Gw_2013def.pdf.

Kooijmans, T. en P.A.M. Mevis (2013) *ICT in the context of criminal procedure: The Netherlands*, TLS/EUR/AIDP.

Koot, M.R. (2012) *Measuring and predicting anonymity*, Amsterdam: Universiteit van Amsterdam.

Kraska, T. (2013) 'Finding the needle in the Big Data systems haystack', *IEEE Internet Computing* 17, 1: 84-86.

Laney, D. (2001) '3D data management: Controlling data volume, velocity and variety', Gartner, beschikbaar op: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Laney, D. (2012) 'The importance of 'Big Data': A definition', beschikbaar op: www.gartner.com/doc/2057415/importance-big-data-definition.

Lazarus, L. en B.J. Goold (2007) 'Security and human rights: The search for a language of reconciliation', blz. 1-24 in B.J. Goold en L. Lazarus (red.) *Security and human rights*, Oxford: Hart Publishing.

Lazer, D., R. Kennedy, G. King en A. Vespignani (2014) 'The parable of Google Flu: Traps in big data analysis', *Science* 14, 343: 1203-1205, beschikbaar op: http://j.mp/1ii4ETo.

Leek, J. (2015) *The elements of data analytic style; A guide for people who want to analyze data*, Leanpub, beschikbaar op: http://leanpub.com/datastyle.

Lerman, J. (2013) 'Big Data and its exclusions', *Stanford Law Review Online*, beschikbaar op: http://dx.doi.org/10.2139/ssrn.2293765.

Levitt, S.D en S.J. Dubner (2005) *Freakonomics. A rogue economist explores the hidden side of everything*, New York: Harper Collins.

Loof, J.P., J. Uzman, T. Barkhuysen, A.C. Buyse, J.H. Gerards en R.A. Lawson (2015) 'Het mensenrechtenkader voor het Nederlandse stelsel van toezicht op de inlichtingen- en veiligheidsdiensten', beschikbaar op: www.ctivd.nl.

Lukoianova, T. en V. Rubin (2014) 'Veracity roadmap: Is Big Data objective, truthful and credible?', *Advances in Classification Research Online* 24, 1: 4-15.

Lyon, D. (2003) *Surveillance as social sorting: Privacy, risk, and digital discrimination*, Londen: Routledge.

Lyon, D. (2014) 'Surveillance, Snowden, and Big Data: Capacities, consequences, critique', *Big Data & Society*, July-September: 1-13.

Lyon, D. (2015) *Surveillance after Snowden*, Cambridge: Polity Press.

Madden, S. (2012) 'From databases to Big Data', *Internet Computing, IEEE* 16, 3: 4-6.

Markoff, J. (2012) 'Scientists see promise in deep-learning programs', *New York Times* 23 november 2012, beschikbaar op: www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html.

Marthews, A. en C. Tucker (2015) 'Government surveillance and internet search behavior', beschikbaar op: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564.

Mayer-Schönberger, V. en K. Cukier (2013) *Big Data. A revolution that will transform how we live, work and think*, Londen: John Murray Publishers.

McAfee, A. en E. Brynjolfsson (2012) 'Big Data: The management revolution', *Harvard Business Review*, october.

McKinsey Global Institute (2011) *Big Data: The next frontier for innovation, competition and productivity*, beschikbaar op: www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.

Meier, P. (2015) *Digital humanitarians. How Big Data is changing the face of humanitarian response*, Boca Raton, FL: CRC Press.

Microsoft (2012) *The big bang: How the Big Data explosion is changing the world*, beschikbaar op: https://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/.

Militaire Inlichtingen- en Veiligheidsdienst (2015) *Jaarverslag 2014*, Den Haag: MIVD.

Ministerie van Economische Zaken (2014) 'Google zorgt voor nieuwe banen in Groningen', Nieuwsbericht 23-09-2014, beschikbaar op: www.rijksoverheid.nl/actueel/nieuws/2014/09/23/google-zorgt-voor-nieuwe-banen-in-groningen

Ministerie van Sociale Zaken en Werkgelegenheid (2015a) *Rapportage over SyRI-projecten*, 29 juni.

Ministerie van Sociale Zaken en Werkgelegenheid (2015b) *Financiële resultaten uit 21 interventieteamprojecten waarvoor met behulp van de voorloper van SyRI bestanden zijn gekoppeld en vervolgens een risico-analyse is gemaakt*, Den Haag, beschikbaar op: www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2015/06/29/bijlage-2-financiele-resultaten-uit-21-interventieteamprojecten-waarvoor-met-behulp-van-de-voorloper-van-syri-bestanden-zijn-ge/bijlage-2-financiele-resultaten-uit-21-interventieteamprojecten-waarvoor-met-behulp-van-de-voorloper-van-syri-bestanden-zijn-gekoppeld-en-vervolgens-een-risicoanalyse-is-gemaakt.pdf.

Minister van Veiligheid en Justitie (2014) *Big Data, privacy en veiligheid. Brief aan de voorzitter van de WRR*, 26 mei 2014.

Ministerie van Veiligheid en Justitie (2015) 'Brief aan de Tweede Kamer over gekoppelde databestanden', 24 maart 2015.

Mislove, A., S. Lehmann, Y. Ahn, J.P. Onnela en J.N. Rosenquist (2011) 'Understanding the demographics of Twitter users', beschikbaar op: http://dougleschan.com/the-recruitment-guru/wp-content/uploads/2014/01/Understanding-the-Demographics-of-Twitter-Users-Jukka-Pekka-....pdf.

Moessner, K.M., L. Blystad en E.Z. Tragos (2015) 'Internet of Things beyond the hype: Research, innovation and development', blz. 15-118 in O. Vermesan en P. Friess (red.) *Building the hyperconnected society IoT research and innovation value chains,*

ecosystems and markets, Aalborg: River Publishers, beschikbaar op: www.internet-of-things-research.eu/pdf/Internet%20of%20Things%20beyond%20the%20Hype%20-%20Chapter%203%20-%20SRIA%20-%20IERC%202015_Cluster_%20eBook_978-87-93237-98-8_P_Web.pdf.

Montjoye, Y-A. de, C.A. Hidalgo, M. Verleysen en V.D. Blondel (2013) 'Unique in the crowd: The privacy bounds of human mobility', *Scientific Reports* 3, beschikbaar op: http://dx.doi.org/10.1038/srep01376.

Narayanan, A. en V. Shmatikov (2008) 'Robust de-anonymization of large sparse datasets', 2008 IEEE symposium on security and privacy 111.

Nationale Ombudsman (2010) *Toegang verboden. Onderzoek naar de opname van vreemdelingen in het Schengen Informatie Systeem en de informatievoorziening hierover*, rapport 2010/115, Den Haag.

Neoclaus, M. (2007) 'Security, liberty and the myth of balance: Towards a critique of security politics', *Contemporary Political Theory* 6: 131-149.

Nieuwenhuis, A.J. (2015) *Over de grens van de vrijheid van meningsuiting: theorie, rechtsvergelijking, discriminatie, pornografie*, Nijmegen: Ars Aequi Libri.

Oerlemans, J.J. en B.J. Koops (2012) 'Surveilleren en opsporen in een internetomgeving', *Justitiële Verkenningen* 38, 5: 35-49.

OESO (2014) *Data-driven innovation: Big Data for growth and well-being*, Parijs: OECD Publishing.

Ohm, P. (2010) 'Broken promises of privacy: Responding to the surprising failure of anonymization', *UCLA Law Review* 57: 1701-1777.

Olsthoorn, P. (2016) *Big Data voor fraudebestrijding*, WRR Working Paper 21, Den Haag: WRR.

Omand, D., J. Bartlett en C. Miller (2012) *Intelligence*, Londen: DEMOS, beschikbaar op: www.demos.co.uk/files/_Intelligence_-_web.pdf?1335197327.

Omtzigt P. (2015) *Mass surveillance*, beschikbaar op: http://website-pace.net/documents/19838/1085720/20150126-MassSurveillance-EN.pdf/df5aae25-6cfe-450a-92a6-e903af10b7a2.

Oracle (2014) 'Oracle buys datalogix; Creates the world's most valuable data cloud to maximize the power of digital marketing', Press release 22 December 2014, beschikbaar op: www.oracle.com/us/corporate/press/2395487.

Ottes, L. (2016) *Big Data in de zorg*, WRR Working Paper 20, Den Haag: WRR.

Pamlin, D. en S. Armstrong (2015) *Global challenges: 12 Risks that threat human civilisation*, Stockholm: Global Challenges Foundation.

Paris, R. (2001) 'Human security; Paradigm shift or hot air?', *International Security* 26, 2: 87-102.

Pasquale, F. A. en D.K. Citron (2014) 'Promoting innovation while preventing discrimination: Policy goals for the scored society', *Washington Law Review* 89: 1413-24, beschikbaar op: http://ssrn.com/abstract=2552864.

Pasquale, P. (2015) *The Black Box society: The secret algorithms that control money and information*, Cambridge: Harvard University Press.

PCAST (2014) *Big Data and pricavy: A technological perspective*, Report to the President, beschikbaar op: www.whitehouse.gov/sites/default/files/microsites/ostp/pcast/pcast_big_data_and_privacy_-_may_2014.pdf [8 februari 2016].

Perez, S. (2014) 'Target's data breach gets worse: 70 million customers had info stolen, including names, emails and phones', *Techcrunch* 10 januari 2014, beschikbaar op: http://techcrunch.com/2014/01/10/targets-data-breach-gets-worse-70-million-customers-had-info-stolen-including-names-emails-and-phones.

Perry, W.L., B. McInnis, C.C. Price, S. Smith en J.S. Hollywood (2013) *Predictive policing: The role of crime forecasting in law enforcement operations*, Santa Monica: Rand.

Politie (2013) *BAVP 2013 - 2017. Aanvalsprogramma informatievoorziening politie*, beschikbaar op: www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2015/08/31/tk-aanvalsprogramma-informatievoorziening-politie-voortgangsrapportage/tk-aanvalsprogramma-informatievoorziening-politie-voortgangsrapportage.pdf.

Porter, T.M. (1986) *The rise of statistical thinking 1820-1900*, Princeton: Princeton University Press.

PredPol (2014) *Predictive policing*, beschikbaar op: www.predpol.com [20 november 2014].

Privacy and Civil Liberties Oversight Board (2014) *Report on the telephone records programme conducted under Section 215 of the USA PATRIOT Act and on the Operations of the Foreign Intelligence Surveillance Court*, beschikbaar op www.pclob.gov/library.html, [17 maart 2016].

Puschmann, C. en J. Burgess (2013) 'The politics of Twitter data', blz. 43-54 in K. Weller, A. Bruns, J. Burgess, M. Mahrt en C. Puschmann (red), *Twitter and society*, New York: Peter Lang.

PWC (2013) *PwC's 5th annual digital IQ survey: Digital conversations and the C-suite*, beschikbaar op: www.pwc.com/us/en/advisory/2013-digital-iq-survey/assets/2013-global-digital-iq-survey-report.pdf.

Quetelet, A. (1835) *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale*, Parijs: Bachelier.

Raab, C.D. (2014) 'Privacy as a security value', blz. 39-58 in D.W. Schartum, L. Bygrave en A.G.B. Bekken (red.) *Jon Bing: En Hyllest / A Tribute*, Gyldendal, www.research.ed.ac.uk/portal/en/publications/privacy-as-a-security-value(837c1adc-7037-46ec-8b7e-8120b3196843).html.

Raad van State (2014) *Advies W12.14.0102/III*, Den Haag, beschikbaar op: www.raadvanstate.nl/adviezen/advies.html?id=11339.

Ratcliffe, J. (2010) 'Crime mapping: Spatial and temporal challenges', blz. 5-24 in A.R. Piquero en D. Weisburd (red.) *Handbook of quantitative criminology*, New York, NY: Springer.

Regalado, A. (2014) 'Business adapts to a new kind of computer', *MIT Technology Review*, may 20, beschikbaar op: www.technologyreview.com/news/527356/business-adapts-to-a-new-style-of-computer/.

Regan, P.M. (1995) *Legislating privacy: Technology, social values, and public policy*, ChapelHill: University of North Carolina Press.

Richards, N.M. en H.J. King (2013) 'Three paradoxes of Big Data', *Stanford Law Review Online* 41, beschikbaar op: http://ssrn.com/abstract=2325537.
Richards, N.M. en J. King (2014) 'Big Data ethics', *Wake Forest Law Review*, 49: 393-432.
Rienks, R. (2014) *Predictive policing. Kansen voor een veiligere toekomst*, Apeldoorn: Politieacademie.
Rössler, B. (2005) *The value of privacy*, Cambridge: Polity Press.
Rubenstein, I. (2013) 'Big Data: The end of privacy or a new beginning?', *International Data Privacy Law* 3, 2: 74-87.
Rusitschka, S. en A. Ramirez (2014) *Big Data technologies and infrastructures*, Deliverable 1.4 BYTE project, final version 1.1, 2 september 2014.
Sanders, C.B., C. Weston en N. Schott (2015) 'Police innovations, 'secret squirrels' and accountability: Empirically studying intelligence-led policing in Canada', *British Journal of Criminology* 55: 711-729.
Sandijk, J. van (2013) 'CBS onderzoekt de mogelijkheden van Big Data', CBS, 30-33, beschikbaar op: www.pietdaas.nl/beta/pubs/pubs/Big_data_zomerrelatiemagazine.pdf.
SAP (2014) *Big Data. Real time. Real results*, SAP.
Savin, A. (2013) 'Profiling and automated decision making in the present and new EU data protection frameworks', blz. 249-270 in P. Arnt Nielsen, P. Koerver Schmidt en K. Dyppel Weber (red.) *Erhvervsretlige emner*, Kopenhagen: Juridisk Institut CBS.
Schacklette (2015) 'Improved analytics reduce false positives in credit card activity', *Tech Republic* 19 januari 2015, beschikbaar op: www.techrepublic.com/article/improved-analytics-reduce-false-positives-in-credit-card-activity/.
Schendel, S. van (2016) *Het gebruik van Big Data door de MIVD en AIVD*, WRR-Working Paper 18, Den Haag: WRR.
Schneier, B. (2015) *Data and Goliath. The hidden battles to collect your data and control your world*, New York: W.W. Norton & Company.
Scott, J.C. (1998) *Seeing like a state: How certain schemes to improve the human condition have failed*, New Haven: Yale University Press.
Searle, J. (1984) *Minds, brains, and science*, Cambridge: Harvard University Press.
Shaw, J. (2014) 'Why "Big Data" is a big deal', *Harvard Magazine*, March-April, http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal.
Skorup, B. (2014) *Cops scan social media to help assess your 'threat rating'*, Reuters, December 12, beschikbaar op: http://blogs.reuters.com/great-debate/2014/12/12/police-data-mining-looks-through-social-media-assigns-you-a-threat-level/ [geraadpleegd op van 15 augustus 2015].
Sloot, B. van der (2014a) 'Do data protection rules protect the individual and should they? An assessment of the proposed General Data Protection Regulation', *International Data Privacy Law* 4.
Sloot, B. van der (2014b), Privacy in the post-NSA era: Time for a fundamental revision?, *JIPITEC*, 2014, 1.
Sloot, B. van der, en S. van Schendel (2016) *International and comparative legal study on Big Data*, WRR-Working Paper 20, Den Haag: WRR.

Sloot, B. van der (2016a) 'The individual in the Big Data era: Moving towards an agent-based privacy paradigm', blz. 177- 203 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Sloot, B. van der (2016b) 'Is the human rights framework still fit for the Big Data era? A discussion of the ECtHR's case law on privacy violations arising from surveillance activities', in Serge Gutwirth et al. (red.) *Data protection on the move, law, governance and technology series 24*, Dordrecht: Springer.

Sloot, B. van der, D.W.J. Broeders en E.K. Schrijvers (2016) 'Introduction: Exploring the boundaries of Big Data', blz. 11-23 in B. van der Sloot, D.W.J. Broeders en E.K. Schrijvers (red.) *Exploring the boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Solove, D.J. (2007) ''I've got nothing to hide' and other misunderstandings of privacy', *San Diego Law Review* 44: 745-772.

Solove, D.J. (2011) *Nothing to hide: The false tradeoff between privacy and security*, New Haven: Yale.

Stehr, N. en M. Adolf (2015) *Ist wissen Macht? Erkenntnisse über Wissen*, Weilerswist: Velbrück Wissenschaft.

Sweeney, L. (2002) 'K-anonymity: A model for protecting privacy', *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 10, 5.

Talbot, D. (2014) 'Cell-phone data might help predict ebola's spread', *MIT Technology Review*, beschikbaar op: www.technologyreview.com/s/530296/cell-phone-data-might-help-predict-ebolas-spread/ [17 september 2014].

Taleb, N.N. (2013) 'Beware the big errors of "Big Data"', beschikbaar op: www.wired.com/.../big-data-means-big-errors-people.

Taylor, C. (2013) 'Big Data's slippery issue of causation vs. correlation', *Wired* 15 juli 2013, beschikbaar op: http://insights.wired.com/profiles/blogs/big-data-s-slippery-issue-of-causation-versus-correlation#axzz45FBg1R8q.

Taylor, L. Cowls en E.T. Meyer (2014) Big Data and positive change in the developing world, *Policy and Internet* 6, 4: 418-444.

Taylor, L. en D. Broeders (2015) 'In the name of development: Power, profit and the datafication of the global south', *Geoforum* 64, 4: 229-237.

Taylor, L., L. Floridi en B. van der Sloot (red.) (2016) *Group privacy. The challenges of new data technologies*, Springer (te verschijnen).

TechAmerica Foundation (2012) *Demystifying Big Data*, Washington, DC, beschikbaar op: www1.unece.org/stat/platform/download/attachments/80053387/Demistyfying%20Big%20Data.pdf?version=1&modificationDate=1374223553898&api=v2.

Tene, O. en J. Polonetsky (2012) 'Privacy in the age of Big Data: A time for big decisions', *Stanford Law Review Online* 64: 63-69.

Tene, O. en J. Polonetsky (2013) 'Big Data for all: Privacy and user control in the age of analytics', *Northwestern Journal of Technology and Intellectual Property* 11, 5: 240-273.

Timmer, J., I. Elias, L. Kool en R. van Est (2015) *Berekende risico's. Verzekeren in de datagedreven samenleving*, Den Haag: Rathenau Instituut.

Tufekci, Z. (2014) 'Big questions for social media Big Data: Representativeness, validity and other methodological pitfalls', *ICWSM 2014: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, 2014.

US National Research Council (2015) *Bulk collection of signals intelligence: Technical options*, beschikbaar op: www.nap.edu/catalog.php?record_id=19414 [17 maart 2016].

Vedder, A., L. van der Wees, B. Koops en P. de Hert (2007) *Van privacyparadijs tot controlestaat? Misdaad- en terreurbestrijding in Nederland aan het begin van de 21ste eeuw*, Den Haag: Rathenau Instituut.

Verenigde Naties (2014) *A world that counts*, New York: VN.

Vis, F., S. Faulkner, K. Parry, Y. Manykhina en L. Evans (2013) 'Twitpicking the riots: Analysing images shared on Twitter during the 2011 U.K. riots', blz. 385-398 in K. Weller, A. Bruns, J. Burgess, M. Mahrt en C. Puschmann (red) *Twitter and society*, New York: Peter Lang.

Walt, S.M. (2013) 'The real threat behind the NSA surveillance programs', beschikbaar op: http://foreignpolicy.com/2013/06/10/the-real-threat-behind-the-nsa-surveillance-programs/

Werkgroep Verkenning kaderwet gegevensuitwisseling (2014) *Kennis delen geeft kracht. Naar een betere èn zorgvuldigere gegevensuitwisseling in samenwerkingsverbanden*, Den Haag, beschikbaar op: http://njb.nl/Uploads/2015/1/blg-442395.pdf [8 februari 2016].

WRR (2011) *iOverheid*, Rapporten aan de Regering nr. 86, Amsterdam: Amsterdam University Press.

WRR (2015) *De publieke kern van het internet*, Rapporten aan de Regering nr. 94, Amsterdam: Amsterdam University Press.

White House (2014) *Big Data: Seizing opportunities, preserving values*, Washington, DC: Executive Office of the President.

Wikipedia (2016) 'Big Data', beschikbaar op: http://en.wikipedia.org/wiki/Big_data. [30 maart 2016].

Willems, D. en R. Doeleman (2014) 'Predictive policing: Wens of werkelijkheid?', *Het Tijdschrift voor de Politie* 76, 4/5: 39-42.

WODC (2014) *Collectieve acties, Een interne rechtsvergelijking tussen privaatrecht en bestuursrecht*, L.F. Wiggers-Rust, Den Haag: WODC.

World Economic Forum (WEF) (2014) *The global information technology report 2014. Rewards and risks of Big Data*, beschikbaar op: www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf.

Yin, C.T., Z. Xiong, H. Chen, J. Wang, D. Cooper en B. David (2015) 'A literature survey on smart cities', *Science China Information Sciences* 58, 100102: 1-18.

Zarsky, T. (2003) 'Mine your own business! Making the case for the implications of the data mining of personal information in the forum of public opinion', *Yale Journal of Law and Technology* 5, 1, beschikbaar op: http://digitalcommons.law.yale.edu/yjolt/vol5/iss1/1.

Zarsky, T.Z. (2013) 'Transparent predictions', *University of Illinois Law Review* 4.

Zarsky, T.Z. (2014) 'Understanding discrimination in the scored society', *Washington Law Review* 89, 1375: 1375-1412.

Zarsky, T.Z. (2016) 'The trouble with algorithmic decisions. An analytic road map to examine efficiency and fairness in automated and opaque decision making', *Science, Technology, & Human Values* 41, 1: 118-132.

Završnik, A. (2013) 'Blurring the line between law enforcement and intelligence: Sharpening the gaze of surveillance?', *Journal of Contemporary European Research* 9, 1: 181-202.

Zikopoulos, P. en C. Eaton (2011) *Understanding Big Data: Analytics for enterprise class Hadoop and streaming data*, McGraw Hill Professional.

Zook, M., M. Graham, T. Shelton en S. Gorman (2010) 'Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake', *World Medical and Health Policy* 2, 2: 7-33.

Zuiderveen Borgesius, F. (2015) *Improving privacy protection in the area of behavioural targeting*, Alphen aan den Rijn: Kluwer Law International.

Zwenne, G-J. En A.H.J. Schmidt (2016) 'Wordt de homo digitalis bestuursrechtelijk beschermd?', blz. 307-385 in E.M.L. Moerel, J.E.J. Prins, M. Hildebrandt, T.F.E. Tjong Tjin Tai, G-J. Zwenne en A.H.J. Schmidt, *Homo Digitalis*, Handelingen Nederlandse Juristen-Vereniging 146e jaargang/2016-I, Wolters Kluwer, Beschikbaar op: njv.nl/wp-content/uploads/2011/04/Preadviezen-NJV-2016.pdf.

ADVIESAANVRAAG

Ministerie van Veiligheid en Justitie

> Retouradres Postbus 20301 2500 EH Den Haag

De Voorzitter van de
Wetenschappelijke Raad voor het Regeringsbeleid
Prof. dr. J.A. Knottnerus
Postbus 20004
2500 EA Den Haag

Directoraat-Generaal Rechtspleging en Rechtshandhaving

Turfmarkt 147
2511 DP Den Haag
Postbus 20301
2500 EH Den Haag
www.rijksoverheid.nl/venj

Ons kenmerk
504054

Bij beantwoording de datum en ons kenmerk vermelden. Wilt u slechts één zaak in uw brief behandelen.

Datum 26 mei 2014
Onderwerp Big data, veiligheid en privacy

Op 13 december jl. heeft het kabinet de notitie "Vrijheid en veiligheid in de digitale samenleving. Een agenda voor de toekomst" aan de beide kamers aangeboden (kamerstukken II 2013/14, 26643, nr. 298; kamerstukken I 2013/2014, 33750, O).

In deze notitie heeft het kabinet aangegeven dat de ontwikkelingen en vraagstukken die in die notitie geagendeerd worden, op sommige punten nog verder doordacht moeten worden. Het gaat daarbinnen specifiek om de ontwikkelingen en vraagstukken die onder het overkoepelende thema "big data, veiligheid en privacy" vallen. De drie hoofdvragen worden hieronder uiteengezet. Daarbij past de kanttekening dat deze vragen in lijn met eerdergenoemde kabinetsnotitie geen betrekking hebben op het werk van de inlichtingen- en veiligheidsdiensten,
De eerste hoofdvraag die in de notitie wordt opgeworpen, is of er een sterker onderscheid moet worden gemaakt tussen toegang tot en gebruik van gegevens bij "big data". Traditioneel omvat gegevensbescherming zowel de toegang tot als het gebruik van persoonsgegevens. Echter, in tijden van "big data" kan het lastig zijn om de toegang tot gegevens te beschermen. Zo speelt bij regulering van de toegang bijvoorbeeld mee dat de gegevensstromen zich niet aan landsgrenzen houden en de locatie waar gegevens zich bevinden (mede als gevolg van de groei van "cloud-computing"), een steeds meer fluïde karakter krijgt. Daarom zal men sterker zijn toevlucht willen nemen tot het reguleren van het gebruik van gegevens.

Deze ontwikkeling roept ook een aantal vervolgvragen op:
- Welke gevolgen zullen ontstaan voor de regievoering door de staat ten aanzien van zowel de bescherming van persoonsgegevens als het gebruik daarvan voor het bevorderen van de veiligheid?
- Waar liggen kansen om "big data" zo te gebruiken dat dit gebruik zowel de effectiviteit van het veiligheidsbeleid vergroot als de bescherming van persoonsgegevens beter waarborgt?
- Hoe kunnen de beginselen van doelbinding en dataminimalisatie, die aan het huidige Europese en Nederlandse gegevensbeschermingsrecht ten grondslag liggen, hun functie behouden?
- Hoe verhoudt het reguleren van het gebruik van gegevens op nationaal of Europees niveau zich toch het verschijnsel dat datastromen niet binnen de

landsgrenzen blijven. Wat voor toegevoegde waarde levert het reguleren van het gebruik van data tegen deze achtergrond? En is er aanvullend beleid nodig dat voldoende technologische soevereiniteit waarborgt zodat het reguleren van het gebruik van data op nationaal of Europees niveau zinvol wordt?
- Als het enkel gaat om verzamelen en opslaan van persoonsgegevens, zonder dat van die gegevens kennis wordt genomen, welk gewicht moeten we dan toekennen aan deze beperking van het recht op bescherming van persoonsgegevens? Bij deze vraag is ook van belang op te merken dat bij "big data" men van tevoren soms niet kan uitsluiten of een gegeven op een later moment een persoonsgegeven zal worden.
- Hoe verhoudt zo'n beperking, gelegen in het verzamelen en opslaan van persoonsgegevens, zich tot de beperking van het recht op bescherming van persoonsgegevens die gelegen is in het verder verwerken van een deel van die gegevens, waarbij van de inhoud van de gegevens wèl kennis wordt genomen?

De tweede hoofdvraag uit de notitie is hoe bij het gebruik van "big data" ervoor kan worden gezorgd dat het proces van "profiling", "datamining" en andere analyse-technieken ten behoeve van de veiligheid voldoende transparant zijn. Het vermogen om uit een enorme hoeveelheid digitale gegevens snel en precies relevante patronen in kaart te brengen, vormt een belangrijk kenmerk van de huidige technieken voor "big data"-analyses. Daarbij kan zowel *ongericht* worden 'gegraven' naar datapatronen in grote hoeveelheden gegevens (ook wel: "datamining") als *gericht* worden "gegraven" naar dergelijke datapatronen ("profiling").

Bij het gebruik van dergelijke technieken rijst de vraag op welke wijze in het analyseproces afwegingen plaatsvinden met betrekking tot beginselen als doelbinding en dataminimalisatie De vragen met betrekking tot transparantie, doelbinding en dataminimalisatie gelden temeer, indien personen bij zowel het dataminen als het profileren worden gecategoriseerd en indien op grond daarvan bepaalde beslissingen ten aanzien van individuele personen worden genomen.

Ten aanzien van de processen van "profiling" en "datamining" speelt een aantal vragen:
- In hoeverre en op welke wijze kunnen deze processen transparant zijn?
- Kan de transparantie van dergelijke processen op zo'n manier worden geborgd dat het niet het belang van een effectieve uitvoering van bijvoorbeeld de politietaak doorkruist?
- In hoeverre kan daaraan bijdragen dat de technologie die voor "profiling" en "datamining" wordt gebruikt, open source software is?
- Is het noodzakelijk dat persoonsgegevens eerst moeten worden opgeslagen, voordat je tot een vorm van "profiling" en "datamining" kan komen en, zo ja, op welke wijze zou dit kunnen worden voorkomen?
- Voor zover dat bij "big-data-analyses" niet voorkomen kan worden, zijn daarvoor dan effectieve waarborgen te creëren?

Naast de technologische kant verdient bij "profiling" en "datamining" ook de maatschappelijke kant aandacht:
- In hoeverre kan aan de transparantie van deze processen worden bijgedragen door inzicht te bieden in de gedragswetenschappelijke vooronderstellingen die aan een specifieke vorm van "profiling" of "datamining" ten grondslag liggen?
- En hoe kan dit inzicht geboden worden zonder dat dit de effectiviteit van bijvoorbeeld het optreden van de politie nadelig beïnvloedt?
- In relatie tot deze processen komt ook de vraag op welke betekenis dient te worden gehecht aan het feit dat vele gegevens die daarvoor kunnen worden gebruikt, geplaatst zijn op internet in een openbare omgeving.

De derde en laatste hoofdvraag die verder doordacht moet worden, is wat de komst van quantum-computers voor het proces van gegevensverwerking ten behoeve van de veiligheid betekent:
- Op welke wijze kunnen we zowel de mogelijkheden die deze computers bieden, goed benutten als een adequaat niveau van gegevensbescherming handhaven?
- En hoe moeten we aankijken tegen de verwachting dat een quantumcomputer veel sneller dan nu encryptie-sleutels kan kraken?
- Hoe verhoudt zich dat tot de verwachting dat de komst van quantumcomputers ook een enorme impuls aan encryptietechnieken kan geven, die zelfs met behulp van quantumcomputers niet gemakkelijk te doorbreken zijn?

Het gaat hier om drie belangrijke hoofdvragen voor de Agenda voor de toekomst, zoals die in eerdergenoemde kabinetsnotitie is gelanceerd. Het gaat hier om een agenda met een dynamisch karakter: er kunnen andere vraagstukken aan worden toegevoegd, als bepaalde ontwikkelingen dat vergen.

Met dat uitgangspunt in het achterhoofd is in een debat in de Eerste Kamer over de Staat van de rechtsstaat op 11 maart jl. een vraagstuk aan de orde geweest dat in het teken staat van het thema "big data" en zich daarom goed leent voor opneming in deze agenda. Dit vraagstuk houdt verband met het verschijnsel dat het volgen en het beïnvloeden van gedrag met behulp van technologie, ook zonder dat we het merken, steeds gemakkelijker wordt.

Dit verschijnsel roept in dit tijdperk van "big data" de volgende vragen op:
- Hoe zorgen we ervoor dat, nu de informatie over personen in databases steeds belangrijker wordt, de kwaliteit van de informatie in gegeven context op een navenant niveau wordt gewaarborgd?
- Hoever strekt de eigen verantwoordelijkheid van de burger voor de kwaliteit van zijn gegevens in databases?
- Hoe slagen we er dan in de burger zelf meer effectieve controle te geven over zijn gegevens?
- Is het voor de burger mogelijk om steeds op basis van "informed consent" te beslissen?

Gelet op de brede expertise van uw raad op onder meer het terrein van technologie, veiligheid en grondrechten, leg ik mede namens de Staatssecretaris van Veiligheid en Justitie en de Minister van Binnenlandse Zaken en Koninkrijksrelaties deze vragen graag aan u voor, met het verzoek ons uiterlijk 1 juli 2015 van advies te dienen. Tegelijkertijd verzoeken wij bij de opstelling van uw advies optimaal gebruik te maken van elders aanwezige bijzondere expertise op genoemde terreinen. Tot slot zouden wij het op prijs stellen, indien u bij de opstelling van uw advies de inzichten betrekt die voortvloeien uit vergelijkbare discussies en onderzoeken in het buitenland.

De Minister van Veiligheid en Justitie,

I.W. Opstelten