

Dit is de geaccepteerde versie van het artikel:

Snaphaan, T., Hardyns, W., van Dijk, A. J., Spithoven, R., & Van Brakel, R. (2023). Setting the scene: Big data policing als multidimensionaal thema voor de praktijk, beleid en onderzoek. In T. Snaphaan, W. Hardyns, A. J. van Dijk, R. Spithoven & R. Van Brakel (Eds.), *Big data policing* (pp. 13–31). Gompel&Svacina.

Setting the scene: Big data policing als multidimensionaal thema voor de praktijk, beleid en onderzoek

Thom Snaphaan¹, Wim Hardyns², Auke van Dijk³, Remco Spithoven⁴ & Rosamunde Van Brakel⁵

In deze kadertekst worden de contouren geschetst van het thema dat centraal staat in dit Cahier: big data policing. Wat verstaan we hieronder, hoe bakenen we het thema af en hoe kijken we ernaar vanuit verschillende perspectieven? Op basis van een brede conceptualisering kunnen ook de verschillende bijdragen in dit verzamelwerk gekaderd worden in een ruimer perspectief. De auteurs hopen hiermee bij te dragen aan de demystificatie van big data policing, waardoor een constructief debat kan worden gevoerd, onderzoek kan worden uitgevoerd en beleid kan worden geformuleerd op een op kennis gebaseerde wijze.

1 Inleiding

In het politiewerk wordt al sinds jaar en dag gebruikgemaakt van data. Deze data komen voor in verscheidene hoedanigheden: groot en klein, hard en zacht, gestructureerd en ongestructureerd, open en gesloten. Een van de meest gebruikte grote databron is ongetwijfeld de enorme datastroom die zichzelf genereert: politiedata, in al haar verschijningsvormen, denk daarbij aan processen-verbaal, bewijsmateriaal zoals camerabeelden en beslag, analyseproducten, et cetera. Deze data worden frequent bekritiseerd, en de bias en fouten die in de data aanwezig zijn, zijn goed gekend (Ferguson, 2017; Van Brakel, 2016). Niettemin blijven ze een belangrijke databron voor praktijk, beleid en wetenschappelijk onderzoek.

Toch is er relatief kort geleden iets veranderd. Nieuwe en opkomende databronnen, ook wel big data, alsook het gebruik ervan, hebben de laatste jaren een vlucht genomen – ook binnen het politiewerk. Voor het politiewerk zijn nu ook zogenoemde *non-police databases* (Brayne, 2021) beschikbaar, die veelal groter en gevarieerder zijn, en bovendien sneller gegenereerd worden dan politiedata. Denk dan bijvoorbeeld aan data die worden voortgebracht door sociale media, data die worden bijgehouden door zoekmachines, data van Google Street View, data afkomstig van bewakingscamera's, en ga zo maar door. Nadat het gebruik van deze data een

¹ Doctoraatsonderzoeker, Institute for International Research on Criminal Policy (IRCP), Vakgroep Criminologie, Strafrecht en Sociaal Recht, Faculteit Recht en Criminologie, Universiteit Gent en Senior onderzoeker, Centre of Expertise Veilige en Veerkrachtige Stad en Omgeving (VVS0), Avans Hogeschool.

² Professor, Institute for International Research on Criminal Policy (IRCP), Vakgroep Criminologie, Strafrecht en Sociaal Recht, Faculteit Recht en Criminologie, Universiteit Gent en Gastprofessor Veiligheidswetenschappen, Universiteit Antwerpen.

³ Strategisch adviseur, Politie Nederland.

⁴ Lector Maatschappelijke Veiligheid, Hogeschool Saxion.

⁵ Docent en postdoctoraal onderzoeker, Vakgroepen Metajuridica en Criminologie, Faculteit Recht en Criminologie, Vrije Universiteit Brussel.

vlucht nam in de private context is deze overgewaaid naar de publieke context (Wetenschappelijke Raad voor Regeringsbeleid [WRR], 2016). De beschikbaarheid van grote, diverse en (bijna) in real-time beschikbare databronnen, alsook de mogelijkheid om deze ruwe data met analyse middels krachtige algoritmen te raffineren tot ‘actionable knowledge’ (WRR, 2016) heeft fundamentele gevolgen voor de manier waarop het politiewerk wordt gestuurd, uitgevoerd en geëvalueerd.

In deze kadertekst wordt de toon gezet voor de verdere uiteenzetting van het concept *big data policing*. Op basis van een brede conceptualisering kunnen de individuele bijdragen binnen dit ruimer kader worden geplaatst. Allereerst zal worden ingegaan op wat onder big data wordt verstaan. Ten tweede wordt het begrip big data policing nader uiteengezet. Vervolgens worden de verschillende facetten van big data policing voor het voetlicht gebracht. Hierbij wordt aandacht besteed aan epistemologisch-methodologische, empirische, juridische, sociaal-technologische, ethische, organisatorische en effectiviteit-evaluatieve perspectieven. Tot slot zal op basis van dit kader een blik vooruit worden geworpen: wat zal de toekomst van big data policing ons (kunnen) brengen?

2 Big data

Om de ontwikkelingen omtrent big data goed te kunnen duiden, keren we terug naar de digitale revolutie. De digitale revolutie omvat de verschuiving van mechanische en analoge elektrische technologie naar digitale en computertechnologie. In het bijzonder na de Tweede Wereldoorlog nemen de ontwikkelingen van nieuwe technologieën en rekenkracht een vlucht. Vanaf eind jaren 70 van de vorige eeuw wordt er gesproken over de opkomst van de informatiemaatschappij (Castells, 2000; Webster 2002), waarin informatie- en communicatietechnologieën steeds meer aspecten van de manier waarop we leven, werken en recreëren veranderen. Een gevolg van deze digitalisering is dat er enorme datastromen worden gegenereerd, zowel van nieuwe activiteiten, als van activiteiten die voorheen niet digitaal vastgelegd werden. Dit heeft tot een belangrijke verschuiving geleid: in veel gevallen hoeft men geen dataverzameling meer op te zetten, het wordt nu in een constante beweging gegenereerd. Deze ontwikkelingen, die we hier in een notendop beschrijven, kunnen gekenschetst worden als de oorsprong van big data. De hamvraag die hierbij rijst: wat moeten of kunnen we binnen het politiewerk met deze big data?

Wat onder big data wordt verstaan, kent vele impliciete en expliciete conceptualiseringen. Wat vandaag gebruikelijk is, is het beschrijven van dit concept op basis van enkele centrale kenmerken. Ze passeerden de revue al even: het gaat om grote datasets (*Volume*), afkomstig van een grote variëteit van bronnen (*Variety*) die snel beschikbaar zijn en geanalyseerd worden (*Velocity*) (De Mauro et al., 2015; Laney, 2001).

De meeste, zo niet alle, eerdere definities van big data waren gericht op beperkingen van middelen en de grootte van datasets (Symons & Alvarado, 2016). Deze beperkingen gaan vandaag echter veelal niet meer op: hoewel big data het gebruik van grote datasets suggereert, is omvang geen essentieel kenmerk voor het huidige gebruik van de term. Initieel waren het namelijk vooral de praktisch-technische beperkingen (bv. rekenkracht en opslag van data) die uitdagingen met zich meebrachten, terwijl het tegenwoordig meer inhoudelijke en principiële vragen zijn die prevaleren. Kitchin en McArdle (2016) stellen in dit verband, na een verkenning van 26 big-datasets: “the key boundary characteristics of Big Data, which together differentiate it from ‘small data’, are velocity (both frequency of generation, and frequency of handling, recording, and publishing) and exhaustivity” (p. 8). Met *exhaustivity* doelen de auteurs op het

idee dat big data een volledig systeem (dat op zichzelf veelal geografisch of anderszins begrensd is) capteert in plaats van op steekproeven te berusten. Met andere woorden, dat men (in principe) beschikt over alle data in plaats van een selectie daaruit. Zij stellen dat volume een bijproduct is van deze twee kenmerken: het in real time verzamelen van gegevens over een heel systeem kan – afhankelijk van de grootte van elk record – een stortvloed aan gegevens opleveren (Kitchin & McArdle, 2016).

Die grote datasets worden in de praktijk vaak gegenereerd zonder dat men actief data aan het verzamelen is met het oog op analyse ervan. In die zin spreekt men dan ook vaak van *found data* (Connelly et al., 2016; Kitchin, 2014a). De datasets die tegenwoordig in de big-dataliteratuur worden beschouwd of gebruikt, zijn veelal kleiner dan de datasets waarmee de term is opgekomen (Boyd & Crawford, 2012). Het gaat met andere woorden al lang niet meer alleen over de hoeveelheid data. Ook kleinere datasets kunnen vandaag deel uitmaken van een ruimer kader inzake big data policing. Er zijn immers tal van zogenaamde big-data-analysemethoden (Hardyns & Snaphaan, 2020) beschikbaar die aangewend kunnen worden op maat van de data in kwestie. Bovendien kunnen die big-datasets niet los gezien worden van de dataverwerkings- en data-analysemethoden die noodzakelijk zijn om ze om te zetten naar actiegerichte informatie, alsook de implicaties die dit heeft voor sturing, uitvoering en evaluatie van politiewerk (WRR, 2016). Het is juist deze flexibele inzet van data, methoden en technieken, samen met het in acht nemen van specifieke randvoorwaarden, die de reikwijdte van big data policing zal bepalen. Deze big-datatoepassingen en de randvoorwaarden die daaraan verbonden zijn, zijn namelijk gedeeltelijk universeel, maar voor actoren met een rechtsstatelijke functie (zoals de politiefunctie) brengt dit ook bijzondere uitdagingen met zich mee.

3 Van big data naar big data policing

Onder invloed van de digitale revolutie is ook de aard en inhoud van politiewerk ingrijpend veranderd. Data, informatie en *intelligence* zijn in toenemende mate centraal komen te staan met bijbehorende nieuwe politiemodellen als informatiegestuurd politiewerk (Janssens et al., 2020) of *intelligence-led policing* (Ratcliffe, 2016; Mali et al., 2017). Met de opkomst van de computer lag de nadruk in eerste instantie op automatisering en vervolgens ondersteunen van het klassieke politiewerk door het gebruik van data en de analyse daarvan. Met de wijdverbreide digitalisering van de samenleving verandert dat ingrijpend (Spithoven, 2020). Informatie- en communicatietechnologie wordt in toenemende mate bepalend voor sociale processen, en data en informatie worden expliciet de grondstof voor politiewerk (Landman, 2022). Deze ontwikkeling loopt op met bredere maatschappelijke ontwikkelingen, echter vormen deze voor de politie een bijzonder krachtenveld vanwege haar rol en inbedding in de democratische rechtstaat. Hierbij valt te denken aan de spanning tussen de bescherming *door* en *tegen* de overheid (van Dijk et al., 2022). Interessant is ook dat daarmee een nieuw (semi-)publiek domein ontstaat waar tal van partijen een rol spelen bij de *policing* daarvan. En, zoals aangegeven, er komt zeer veel informatie beschikbaar die een rol kan spelen bij *policing* zonder expliciet voor deze doelstelling te zijn verzameland. Daarbij staat het strafrecht niet als vanzelf centraal, en ook opsporing verandert van aard en omvang, zoals de recente voorbeelden van infiltratie (zie bijv. de operatie rond darkmarket Hansa Market; Openbaar Ministerie, 2017) en grootschalige interceptie (zie bijv. de operatie rond de aanbieder van cryptocommunicatie Sky ECC; Openbaar Ministerie, 2021) laten zien.

De betekenis van ‘de politie’ in de bredere context van een pluriforme politiefunctie lijkt nog niet uitgekristalliseerd. Van Steden en collega’s (2009) omschrijven de politiefunctie als: “een

regulatieve functie bestaande uit toezicht op en handhaving van algemeen geaccepteerde normen en regels en het beschermen van de orde binnen een sociale omgeving – indien nodig met gebruikmaking van drang en/of dwang” (p. 6). Naast dat er kritische vragen te stellen zijn over big-datatoepassingen bij de politie zelf en publieke organisaties binnen de politiefunctie (denk aan bijzondere opsporingsdiensten, Belastingdienst en gemeenten, zie Spithoven & Beerends, 2020; Spithoven & Foppen, 2021) lijkt het geen twijfel dat private partijen op verschillende manieren een grote rol spelen als het gaat om big data policing. Denk hierbij bijvoorbeeld aan de datapositie van sociaalnetwerksites en grote technologiebedrijven, de rol van banken in de aanpak van financieel-economische criminaliteit, en de rol van private partijen bij hardware- en softwareontwikkeling inzake big data policing (zie bv. Terpstra & Salet, in dit nummer). Dit brengt prangende vragen met zich mee, zoals of dit wel rechtstatelijk verantwoord en ethisch wenselijk is. Hoe verhoudt deze private, grootschalige toepassing van surveillance zich immers tot rechtstatelijke waarborgen? (Andrejevic, 2014; Brayne, 2017; Dencik et al., 2018; Joh, 2016; Sanders & Sheptycki, 2017; Van Brakel, 2016, 2020; Spithoven, 2020).

Dit leidt ook tot een zoektocht naar nieuwe rechtsstatelijke waarborgen die in toenemende mate tot stand komen op Europees niveau (Kool et al., 2017; Van Brakel, 2020; WRR, 2016). Het is interessant te zien hoe ‘markt’ en ‘staat’ zich hier tot elkaar zullen verhouden. Duidelijk is dat op Europees niveau ook wordt ingezet op het expliciet verantwoordelijk maken van private partijen voor ‘politiewerk’, bijvoorbeeld in de context van de opkomende AI-regulering (i.e., de Artificial Intelligence Act) en regulering van digitale dienstverlening (i.e., de Digital Services Act). Marc Andrejevic (2014) stelde bijna tien jaar geleden al vast dat deze ontwikkelingen ook op dit vlak uitdagingen met zich meebrengen:

Never in the history of human society have we faced the prospect of having all our actions, movements, and communications not simply monitored, but also recorded, stored and sorted by systems whose abilities to process large amounts of data far surpass that of any individual or group of individuals. The technology is moving much faster than the social norms that might regulate and shape its use. (p. 67)

Toch is er nog slechts mondjesmaat aandacht voor de potentiële impact van big data policing – door zowel publieke als private actoren – op de vrijheden van burgers. Hierover zou proactiever politiek, publiek en professioneel debat gevoerd moeten worden (Spithoven, 2020). Tevens dient de beeldvorming omtrent de thematiek verruimd te worden. Big data policing lijkt zowel een enge als ruime opvatting te hebben. De enge opvatting van big data policing relateert aan de beschikbaarheid van toenemende hoeveelheden gevarieerde data die zich met hoge snelheden aandienen. De ruimere opvatting van big data policing heeft naast de data zelf ook oog voor de verwerking, analyse en het gebruik ervan (zie bv. WRR, 2016). Big data policing is daarom bijvoorbeeld niet zomaar gelijk te stellen aan predictive policing (Schuilenburg & Soudijn, 2021) – al was het maar omdat het analytisch palet van mogelijkheden vele malen diverser is (denk aan descriptieve en diagnostische analyses). Ten behoeve van het beschrijven van voor- en nadelen is het van belang om conceptueel helderheid te verschaffen. Bij een verwerking en analyse van big data op basis van machine learning zou immers de vraag kunnen worden gesteld waar het probleem ligt: bij het algoritme of de data die gekozen worden als input? (Das & Schuilenburg, 2020; Kirkpatrick, 2017; Richardson et al., 2019). En als blijkt dat een bepaalde databron geanalyseerd middels een bepaald algoritme maatschappelijk problematische uitkomsten oplevert, moeten we dan alle andere mogelijke databronnen en algoritmen over dezelfde kam scheren? Tot slot omvat big data policing een breed thema (Snaphaan et al., 2021) dat erkend dient te worden in al haar facetten. We hopen hierbij, naar

analogie met de toepassing van artificiële intelligentie (WRR, 2021), bij te dragen aan een demystificatie van het thema.

4 Facetten van big data policing

In deze paragraaf worden de verschillende facetten van big data policing in de vorm van perspectieven uitgewerkt. Hierbij wordt naast de beschrijving en contourschets ervan, eveneens een inkijk gegeven in enkele actuele inzichten uit onderzoek en de praktijk. Daarnaast wordt waar mogelijk verwezen naar relevante bijdragen in deze bundel.

4.1 Epistemologisch-methodologisch perspectief

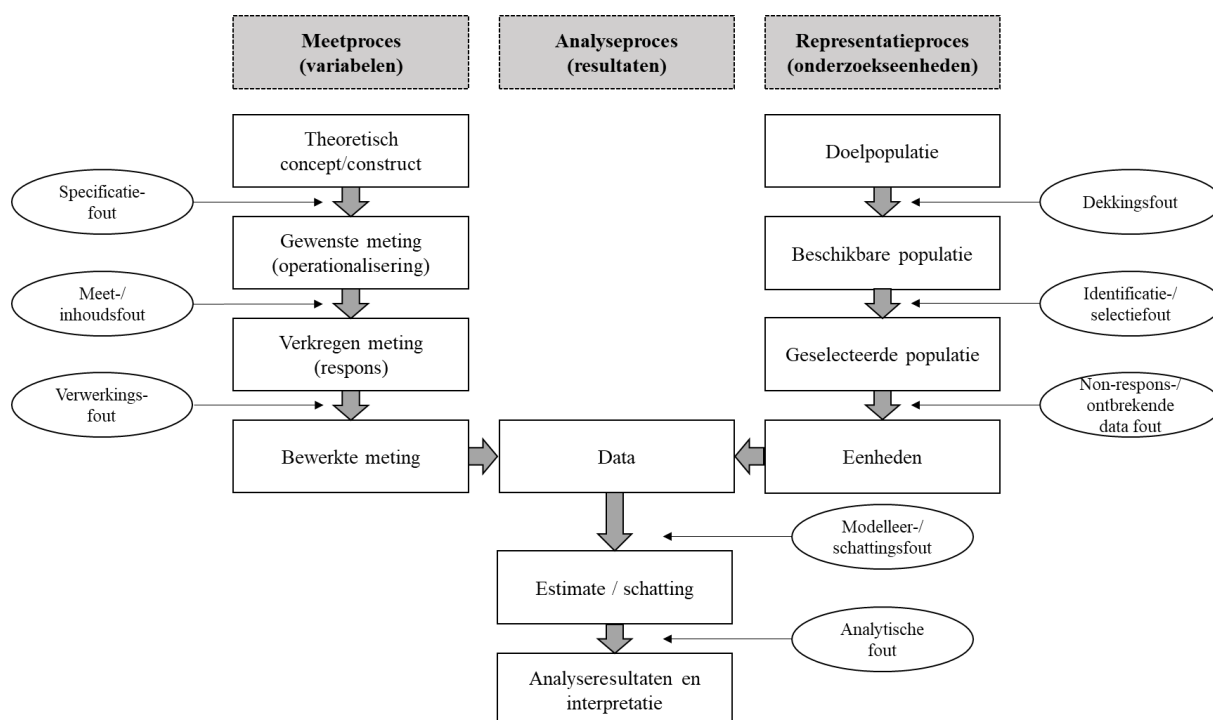
Een van de meest fundamentele aspecten van big data policing richt zich op het epistemologisch-methodologisch vraagstuk. Epistemologie en methodologie zijn niet hetzelfde, maar wel onlosmakelijk met elkaar verbonden. Epistemologie, ook wel kennistheorie, betreft de tak van de (filosofische) wetenschap die de oorsprong, aard, voorwaarden en reikwijdte van kennis en ‘het weten’ onderzoekt. Methodologie omvat de wetenschappelijke methoden, procedures en werkwijzen die moeten worden gebruikt om deze kennis te verwerven.

Het epistemologisch vraagstuk gaat in dit verband terug naar de vraag welke kennis überhaupt uit data te ontginnen is, met inbegrip van hoe dat dan zou moeten gebeuren. In grote lijnen zijn daar twee manieren te onderkennen: een inductieve en een deductieve aanpak, die algemeen bekend zijn in de uitvoering van wetenschappelijk onderzoek. Simpel gesteld vertrekt inductie van observatie om tot hypothese- en theorievorming te komen, terwijl bij deductie theoriegestuurde hypothesen worden getoetst middels observaties. De toenemende beschikbaarheid en het toenemend gebruik van big data houdt een epistemologisch risico in om te vervallen in vormen van empirisme, oftewel dat ‘de data voor zichzelf kunnen spreken’ en dus louter inductieve analyses prevaleren. Er bestaan echter ook epistemologieën die ontvankelijk zijn voor een constructieve omarming van big data, die relevante data dus niet negeren noch verwerpen, en die (sociale) wetenschappers in staat stellen om betekenisvolle inzichten te verkrijgen uit deze data (zie hiervoor Snaaphan, Pauwels & Hardyns, in dit nummer).

De aspecten van het methodologisch vraagstuk vallen uiteen in twee grote lijnen: welke methoden en technieken kunnen gebruikt worden om big data te verwerken en analyseren, en welke criteria kunnen gehanteerd worden om de methodologische kwaliteit van de databronnen te beoordelen? Symons en Alvarado (2016) stellen dat het gebruik van computationele methoden een onderscheidende eigenschap is voor big data. Hiermee wordt ook verwezen naar de eerdere definities van big data die wezen naar de beperkingen van de middelen om ze te verwerken. De data zijn te talrijk, te snel en/of te gevarieerd om manueel te verwerken en analyseren, waardoor de noodzaak voor computationele en ook algoritmische methoden zich aandient. In de sociaalwetenschappelijke disciplines vertalen deze ontwikkelingen zich in een nieuwe subdiscipline: *computational social science*, waarbij methoden en technieken afkomstig van de computerwetenschappen toegepast worden op onderzoeksvragen uit de sociale wetenschappen. Een fundamenteel verschil tussen de methoden uit de sociale wetenschappen en de methoden uit de computerwetenschappen is dat deze niet op dezelfde leest geschoeid zijn: methoden in de sociale wetenschappen baseren zich voornamelijk op datamodellen (ontleend aan de exacte wetenschappen) en methoden in de computerwetenschappen baseren zich voornamelijk op algoritmische modellen. In dit verband wordt ook weleens gerefereerd aan het

verschil in benadering tussen respectievelijk de statistiek (waarbij gebruik wordt gemaakt van datamodellen) en *data science* (waarbij gebruik wordt gemaakt van algoritmische modellen; Nantais, 2019; Oberski, 2017). Berk (2013) spreekt in dit verband over respectievelijk conventionele regressiemodellen en *machine learning* modellen. Zonder hier uitgebreid op in te gaan, kan gesteld worden dat deze toepassing van machine learning en algoritmische modellen een *game changer* zijn (zie Campedelli, 2022 voor een ruimere uiteenzetting). Deze ontwikkelingen gaan hand in hand met de ontwikkelingen inzake big data en raken direct aan de andere perspectieven die uiteengezet worden in deze kadertekst. Zo heeft het gebruik van methoden en technieken uit andere disciplines directe implicaties voor organisatorische en governance vraagstukken.

De andere kant van de methodologische medaille ziet op de kwaliteitscriteria die gebruikt kunnen worden voor big-databronnen. De mate van methodologische kwaliteit wordt doorgaans uitgedrukt in termen van *bias* en *error*, die een invloed hebben op de accuraatheid van de meting. Deze accuraatheid weerspiegelt de mate waarin een geobserveerde waarde afwijkt van de werkelijke waarde. Deze afwijkingen kunnen zowel random (variabiliteit) als systematisch (*bias*) zijn. In surveyonderzoek is het al langere tijd gebruikelijk om de mate van fouten in te schatten aan de hand van een raamwerk: het Total Survey Error raamwerk. Voor (big) data in het algemeen was een dergelijk raamwerk tot voor kort niet voorhanden. Daar is met de introductie van het Total Error raamwerk (zie Figuur 1) verandering in gekomen (Amaya et al., 2020; Snaphaan & Hardyns, 2021a). In dit raamwerk is specifieke aandacht voor de verschillende soorten fouten die zich kunnen voordoen in verschillende fasen en aspecten van het onderzoeksdesign: in het meetproces, het representatieproces en het analyseproces. Binnen die respectievelijke processen zijn een aantal generieke processtappen te onderscheiden (zie vierkanten in Figuur 1), waar bij elk van die processtappen ook gerelateerde foutcomponenten te onderscheiden zijn (zie ovalen in Figuur 1). Het doel van dit raamwerk is niet om elke fase van het proces zo foutloos mogelijk uit te voeren. Dat zou immers betekenen dat, rekening houdend met de immer relatieve schaarste van middelen, het beschikbare budget en/of de timing buitenproportioneel verruimd dient te worden. Bovendien kunnen er zelfs onder de beste omstandigheden, met een onbeperkt budget en onbeperkte tijd nog altijd fouten (zowel *bias* als variabiliteit) kunnen optreden. In plaats van ‘foutloos’ is het doel om de meest flagrante fouten te vermijden en andere fouten tot een aanvaardbaar niveau terug te brengen (Biemer, 2010). Dat aanvaardbaar niveau wordt bepaald door de personen die instaan voor het studiedesign, waarbij het wordt afgewogen ten opzichte van beschikbare middelen, en waarbij ook gekeken wordt naar kwaliteitscriteria die zich richten op de bruikbaarheid, zoals tijdigheid, toegankelijkheid en vergelijkbaarheid.

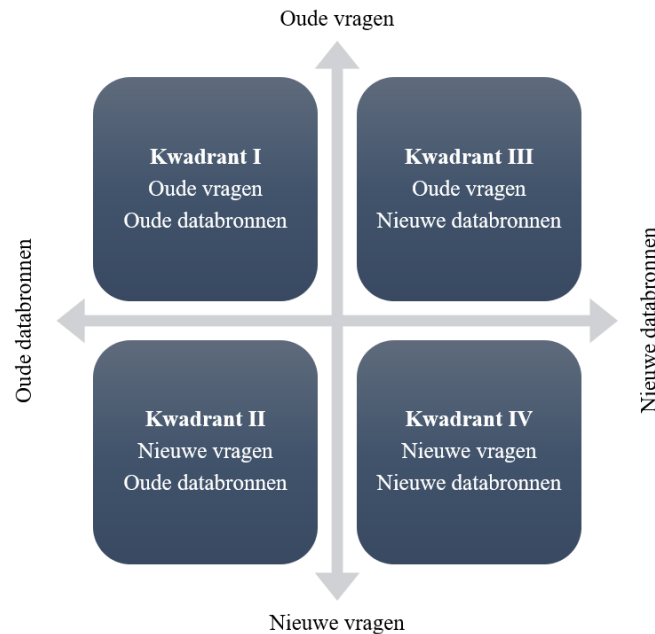


Figuur 1. Total Error raamwerk: Total Error componenten (ovalen) in relatie tot een generiek proces (vierkanten) van theoretisch concept en doelpopulatie tot en met analyse en interpretatie van de resultaten (Snaphaan & Hardyns, 2021a).

4.2 Empirisch perspectief

Naast de theoretische en methodologische vraagstukken, rijst de vraag in welke mate data überhaupt empirisch beschikbaar zijn om te gebruiken in functie van big data policing. Empirische kennis is kennis die voortvloeit uit wetenschappelijk onderzoek gebaseerd op waarnemingen. Big data vormen potentiële databronnen die gebruikt kunnen worden in wetenschappelijk onderzoek om empirische kennis te genereren (Chan & Bennett Moses, 2016). Het is immers niet het volume, de omvang of de hoeveelheid data die revolutionair zijn, maar veeleer de toenemende beschikbaarheid van nieuwe en gevarieerde databronnen die voorheen niet beschikbaar waren voor sociaalwetenschappelijk onderzoek (Connelly et al., 2016). In het verleden was het soms uitdagend om data te verkrijgen om sociaalwetenschappelijk onderzoek uit te voeren en de beschikbaarheid van die data is primordiaal voor de groei van wetenschappelijke kennis in een bepaald domein binnen de sociale wetenschappen: data bieden namelijk een onafhankelijk middel om theoretische stellingen of voorafgaand onderzoek te bevestigen of te falsificeren (Schroeder, 2018).

De manier waarop deze nieuwe databronnen kunnen bijdragen aan onderzoek, het verkrijgen van nieuwe inzichten en daarmee het opbouwen van kennis, kan gezien worden door twee assen met elkaar te kruisen: er zijn zowel oude als nieuwe vragen, en er zijn zowel oude als nieuwe databronnen. Dit creëert vier kwadranten met ieder een bepaald type van onderzoek (zie Figuur 2). Het onderzoek waarin gebruik wordt gemaakt van big data valt in de twee rechterkwadranten.



Figuur 2. Kwadranten oude en nieuwe vragen versus oude en nieuwe databronnen (naar idee van Bowers & Snaphaan, 2022).

Bij het beantwoorden van oude vragen met nieuwe databronnen (kwadrant III) wordt gebruikgemaakt van de nieuwe meetopportunities die big data bieden. Voorbeelden hiervan zijn het gebruik van Google Street View om fysieke overlast te meten (Snaphaan & Hardyns, 2021b) of het gebruik van mobiele-telefoondata om menselijke mobiliteit te meten (Rummens et al., 2021). De beantwoording van nieuwe vragen met nieuwe databronnen (kwadrant IV) is in twee opzichten innovatief, omdat het op beide assen om nieuwe ontwikkelingen gaat. Deze twee gaan vaak ook hand in hand, zeker wanneer de (eerder beschreven) toenemende digitalisering in de basis aan de grondslag ligt van een (relatief) nieuwe fenomeen, zoals cybercriminaliteit. Voorbeelden hiervan zijn onderzoek naar forumcommunicatie op darkmarkets in het verhandelen van cybercrimepakketten (van Wegberg et al., 2018) of het gebruik van blockchaintransacties om phishing van cryptocurrencies te detecteren (Chen et al., 2020).

4.3 Juridisch, sociaal-technisch en ethisch perspectief

Wanneer nieuwe technieken worden geïmplementeerd – zoals big-datatoepassingen in het politiewerk – is er doorgaans slechts oog voor de directe effecten die men ermee tracht te realiseren, bijvoorbeeld efficiëntie en effectiviteit. Maar naast de beoogde effecten treden ook vaak neveneffecten op, zowel op juridisch, ethisch als sociaal-technisch vlak. Deze waren op voorhand niet altijd verwacht, maar kunnen wel in positieve en negatieve zin optreden. Wanneer we kijken naar de maatschappelijke impact van toepassingen van nieuwe technologie, zeker wanneer het gaat over een assemblage van eerder losstaand ontwikkelde technieken zoals vandaag de dag in hoog tempo gebeurt, is het noodzakelijk dat er ex ante, ex durante en ex post wordt geëvalueerd op de effecten en neveneffecten van deze verandering (Spithoven, 2020, Van Brakel, 2020; zie ook Testerink et al., in dit nummer). In dit kader is het overzicht van Kool en collega's (2017) ten aanzien van de maatschappelijke en ethische dilemma's rondom de digitalisering van de samenleving bijzonder illustratief (zie Figuur 3).

Thema	Vraagstukken
Privacy	Gegevensbescherming, privacy, digitaal huisrecht, mentale privacy, surveillance, doelverschuiving
Autonomie	Keuzevrijheid, vrijheid van meningsuiting, manipulatie, paternalisme
Veiligheid	Informatieveiligheid, identiteitsfraude, fysieke veiligheid
Controle over technologie	Controle en inzicht in algoritmen, verantwoordelijkheid, onvoorspelbaarheid
Menselijke waardigheid	Dehumanisatie, instrumentalisering, <i>de-skilling</i> , de-socialisatie, werkeloosheid
Rechtvaardigheid	Discriminatie, uitsluiting, gelijke behandeling, stigmatisering
Machtsverhoudingen	Oneerlijke concurrentie, uitbuiting, relatie consument-bedrijf

Figuur 3. Maatschappelijke en ethische dilemma's rondom de digitalisering van de samenleving (Kool et al., 2017, p. 47).

Een van de voornaamste juridische neveneffecten wat big data policing betreft zijn de risico's voor gegevensbescherming en andere grondrechten die in het Europees Handvest van de Grondrechten⁶ zijn opgenomen zoals privacy en non-discriminatie. Meer concreet moet het gebruik van big data door politiediensten in overeenstemming zijn met de nationale vertaling van de Europese Politie en Justitie Richtlijn⁷ (zie voor een kritische bespreking van deze wetgeving en Europese rechtspraak in het kader van big data policing de bijdrage van De Hert en Sajfert, in dit nummer). Naar de toekomst toe zal de voorgestelde AI-regulering⁸ een grote impact hebben op het gebruik van big data in politiewerk en vooral de ontwikkeling van de algoritmen die gebruikt worden om de big-data-analyses uit te voeren door technologiebedrijven strenger te reguleren. Deze wetgeving is echter nog volop in ontwikkeling. Ten slotte heeft het gebruik van big data ook implicaties voor het straf(proces)recht (zie voor een bespreking van het gebruik van big data in de opsporing en het strafprocesrecht de bijdrage van van Schendel en Cuijpers, in dit nummer).

Een belangrijk kenmerk van het gebruik van big-data-analyses is dat deze op geaggregeerd niveau plaatsvinden en dat het niet enkel gaat om gevolgen voor het individu maar ook collectieve en sociale gevolgen inhoudt. Hierdoor is discriminatie minstens een even groot risico als privacy (Van Brakel, 2020, 2022; WRR 2016). Meer specifiek, wordt in de literatuur de volgende sociaal-technologische neveneffecten van big data policing geïdentificeerd: discriminatie, sociale stratificatie, toegenomen ongelijkheid, stigmatisering en *overpolicing* als gevolg van onder meer bias en feedback loops (Ferguson, 2017; Lum & Isaac, 2016; Richardson et al., 2019; Van Brakel, 2016; Williams & Kind, 2019). De manier waarop big data worden toegepast in het politiewerk zal dus een impact hebben op de juridische, ethische en sociale vraagstukken (zie voor een reflectie over big data in de praktijk de bijdrage van Schuilenburg, in dit nummer).

Wanneer naast de directe effecten van big-datatoepassingen in het politiewerk, ook de indirecte effecten op bijvoorbeeld bovenstaande thema's worden gewogen, ontstaat een complex en meer gebalanceerd overzicht van de werkelijke socio-technische impact van deze verandering. Daarbij is het bijzonder raadzaam deze evaluaties door een interdisciplinair team te laten

⁶ Handvest van de Grondrechten van de Europese Unie van 7 december 2000, *Pb.L.* 18 december 2000, 364, https://www.europarl.europa.eu/charter/pdf/text_nl.pdf

⁷ Europese richtlijn van 27 april 2016 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 en tot intrekking van Kaderbesluit 2008/977/JBZ van de Raad, *Pb.L.* 4 mei 2016, 119, <https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=celex%3A32016L0680>

⁸ Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende Artificiële Intelligentie (Wet op Artificiële Intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Europese Unie, 21 april 2021, COM(2021)206, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

uitvoeren van bijvoorbeeld executieve politiemensen, juristen, ethici, bestuurskundigen, economen en sociaal-wetenschappers (Spithoven & van de Pas, 2021). Gezien de complexiteit van de materie is het noodzakelijk om effecten en neveneffecten vanuit verschillende, complementaire disciplines te beschouwen om de effecten en neveneffecten in een verantwoorde balans te houden.

4.4 Organisatorisch en governance-perspectief

De implementatie van nieuwe technologie in een organisatie, leidt tot verandering van onder andere werk, werkprocessen en dagelijkse praktijken. Dit geldt evenzeer voor de politieorganisaties in België (zie Forceville, in dit nummer) en Nederland (zie den Hengst & Wijsman, in dit nummer). Terpstra en Salet (2020) roepen op om meer aandacht te besteden aan de sociale processen die de invloed van technologische vernieuwing van het politiewerk beïnvloeden. Daarbij gaat het naast de ervaren impact op de werkvloer en de wenselijkheid daarvan volgens professionals, bij big data policing ook nadrukkelijk om de verhouding tussen mens en computer: hoe verhoudt big data policing zich tot professioneel vakmanschap? Daarbij blijken de verwachtingen van technologie bij nader inzien vaak (te) hooggespannen (zie, ter illustratie, de bijdrage van Terpstra en Salet, in dit nummer). In het kritisch debat over digitalisering en de kansen van big data en artificiële intelligentie op complexe praktijken als die van het politiewerk, is doorgaans de consensus dat technologie de menselijke professional kan ondersteunen in het dagelijks werk. Wanneer het gaat over intelligence-led policing, wreekt daar zich immers al snel de complexiteit van intelligence, zoals deze in Figuur 4 illustratief is opgenomen.



Figuur 4. Het cyclisch proces van signaal tot intelligence (Kop & Klerks, 2009, p. 8).

Om van data naar actiegerichte informatie te komen, zullen deze – evenals bij wetenschappelijk(e) onderzoek en methodologie – immers geïnterpreteerd moeten worden en causale relaties moeten worden gelegd om tot praktisch, betekenisvol handelen te komen. Daarbij kan een computer en kunnen big data de menselijke professional helpen signaleren van potentiële samenhangen, maar zal de menselijke professional de interpretatie op basis van causale uitleggen in combinatie met de professionele kennis, intuïtie en wijsheid moeten komen (Spithoven, 2020). Daarmee is de praktijk van big data policing dus vele malen complexer dan louter de implementatie van een big-datatoepassing. Überhaupt de afweging van de eventuele ontwikkeling van een dergelijke toepassing vormt idealiter het startschot om met uitvoerende politiemensen en verschillende experts (zie paragraaf 4.3) naar de professionele,

organisatorische en maatschappelijke impact te kijken en deze op voorhand al op verantwoorde wijze te ontwerpen.

We zien vandaag een evolutie richting multidisciplinaire en geharmoniseerde modellen waar organisatorische, technische en juridisch-beleidsmatige aspecten samenkomen. Politie Nederland ontwikkelde bijvoorbeeld een raamwerk om omkadering en invulling te geven ten behoeve van een datagedreven opsporing van misdrijven (i.e. het zogenoemde CSAE-raamwerk, zie van de Sandt et al., 2022). Als gevolg daarvan blijken bestaande (lijn)organisatiestructuren niet altijd afdoende om te beantwoorden aan deze noden, wat resulteert in flexibele netwerkorganisaties (zie bijvoorbeeld Roest, in dit nummer). Dit alles brengt ook nieuwe noden ten aanzien van sturing met zich mee (zie den Hengst & Wijsman, in dit nummer). Naast nieuwe noden in governance ontstaan er ook nieuwe rollen binnen teams. De relevantie van en nood aan data scientist en data engineers in dit verband behoeft weinig toelichting, maar vandaag zien we ook het belang van rollen als die van analytics translator (Henke et al., 2018) of algoritmische kennismakelaars (Waardenburg, 2021) meer en meer opkomen, omdat ze complementaire competenties beslaan:

“Het is van belang om vooral bij besluitvormers de misvatting te voorkomen dat analyses in de naaste toekomst geheel door computers kunnen worden verricht, die door enkele data scientists worden bediend (...). Dat is niet zo. Naakte gegevens zonder nadere menselijke analyse zijn zelden bruikbaar om tot tactische of strategische besluitvorming te komen.” (Klerks & Vink-Teeven, 2020, p. 166).

4.5 Effectiviteits- en evaluatieperspectief

Een laatste, en soms vergeten, facet van big data policing heeft te maken met het effectiviteitsvraagstuk. Als we te weten willen komen of big data policing in de praktijk tot de gewenste effecten leidt, is er nood aan evaluatie. Dit is allerm minst een makkelijke, maar wel een noodzakelijke opdracht. Bij het evalueren van big data policing toepassingen komt – idealiter – vrijwel elk van bovenstaande perspectieven aan bod. Een holistische aanpak is daarbij aan te bevelen door rekening te houden met verschillende van de hierboven besproken benaderingen van big data policing (zie voor een voorstel van waar zo een holistische aanpak allemaal mee rekening zou moeten houden bijvoorbeeld Van Brakel, 2021). In de praktijk worden evaluaties veelal beperkt tot het kijken naar het effect van een of meerdere big-datatoepassingen op criminaliteit en spelen juridische, ethische en sociale vragen geen centrale rol. Stellen we minder criminaliteit en/of onveiligheid vast in een bepaald gebied? Wordt de pakkans verhoogd? Is er sprake van een preventief of afschrikwekkend effect? Dit lijken op het eerste zicht makkelijk te beantwoorden vragen, maar de realiteit is complexer. Er moet immers zeer goed nagedacht worden over hoe men een vastgesteld patroon of effect (louter en eenduidig) kan toewijzen aan een toepassing die zich situeert binnen het brede spectrum van big data policing. Hiervoor zijn geavanceerde experimentele designs noodzakelijk (zie hiervoor Khalfa & Hardyns, in dit nummer).

Daarbij mag men niet vergeten ook oog te hebben voor zogenaamde secundaire effecten: hoe wordt het toepassen van big data policing ervaren binnen de politieorganisatie? Wordt er voldoende rekening gehouden met sociaal-maatschappelijke vraagstukken? Worden de ethische en juridische waarborgen voor mens en maatschappij gegarandeerd? Zijn de technologische investeringen ook kosteneffectief? Leggen we het ook vast wanneer bepaalde toepassingen niet het verhoopte resultaat opleveren in bepaalde contexten? Dit zijn slechts

enkele van de vele vragen die naast het primaire effect ten aanzien van criminaliteit belangrijk kunnen zijn.

Naast de evaluatie van big-datatoepassingen in het politiewerk brengen digitaliserings- en dataficatieprocessen ook opportuniteiten met zich mee om het politiewerk zelf te evalueren. Sporen van het politiewerk zijn immers ook steeds vaker digitaal beschikbaar (denk aan GPS-tracers in surveillancevoertuigen en logbestanden van databasebevragingen), waardoor ze bijvoorbeeld gebruikt kunnen worden als onderwerp van onderzoek of als controlevariabelen in evaluaties van big-datatoepassingen.

Evaluatiestudies die vanuit een dergelijke holistische aanpak vertrekken, zijn vandaag – zo niet onbestaand – bijzonder schaars. Dit heeft veelal te maken met de beschikbare middelen voor evaluatieonderzoek: als er al geld wordt vrijgemaakt voor evaluatie, zijn de vrijgemaakte middelen vaak te beperkt om zowel primaire als secundaire effecten in kaart te brengen. Bovendien is voor een holistische aanpak een verscheidenheid aan expertise nodig die in de praktijk moeilijk bij elkaar te brengen en gezamenlijk te activeren is. Nochtans zullen investeringen in doorgedreven en volledig evaluatieonderzoek, zowel wetenschappelijk als maatschappelijk, op termijn altijd de meeste ‘winst’ opleveren. Op die manier kan men immers mét kennis van zaken beslissen bepaalde toepassingen verder uit te bouwen, bij te sturen of af te schaffen.

5 Besluit

Deze kadertekst heeft de contouren geschetst van wat big data policing inhoudt. Dit concept wordt concreter en door diverse auteurs nader uitgewerkt in dit Cahier. De contouren van deze conceptualisatie laten zien dat het gaat om een breed en divers thema, waar veel rond gebeurt, maar dat onverminderd aandacht vereist van beleidsmakers, praktijkmensen en academici. Op basis van deze contouren, waarbinnen nog veel valt ‘in te kleuren’, wensen we ook een ruwe schets te maken van de weg voorwaarts: wat zijn de routes voor beleid, praktijk en wetenschap? In wat volgt zullen in de respectievelijke structuur van de hierboven beschreven vraagstukken enkele routes voorwaarts worden beschreven.

Voor wat betreft *epistemologische en methodologische aspecten* verdient het de aanbeveling om te vertrekken vanuit een constructief epistemologisch kader dat aansluit bij het gegeven dat onder invloed van digitaliseringsprocessen en dataficatie er gewoonweg meer (diverse) data beschikbaar zijn. Het is zaak deugdelijke vertrekpunten te kiezen om deze data op een betekenisvolle manier in te zetten (zie bijdrage Snaphaan, Pauwels & Hardyns, in deze bundel). Vanuit methodologisch perspectief kan teruggegrepen worden op bestaande raamwerken om zowel de intrinsieke kenmerken van big data op waarde te schatten (Snaphaan & Hardyns, 2021a; cfr. big data policing in enge zin), alsook raamwerken of kwaliteitskaders die handvatten bieden voor het ruimer proces waarin big data een rol hebben (zie bijv. het Kwaliteitskader Big Data, OM en Politie, 2020; cfr. big data policing in ruime zin).

Met betrekking tot de *empirische aspecten* dient allereerst het belang onderstreept te worden van een empirische verkenning van de meerwaarde van verschillende big-databronnen. Door de waarde hiervan empirisch vast te stellen, middels theorievormend en theorietoetsend onderzoek, kan op basis daarvan gestuurd worden waarvoor, in welke mate en hoe de data ook gebruikt worden. Op dit vlak kan ook gesteld worden dat de relatie tussen (fundamenteel en praktijkgericht) onderzoek en de praktijk aan belang wint. In een tijd waar op kennis gebaseerd beleid steeds belangrijker wordt, wordt de rol van onderzoek dominanter en de relatie tussen

beleid, praktijk en onderzoek intensiever. Specifiek met betrekking tot dit thema is het raadzaam om een collectieve denktank in te richten die zich buigt over de vraagstukken gerelateerd aan big data policing. Deze denktank dient multidisciplinair en onafhankelijk ingericht te worden. Een onderzoeksprogramma dat specifiek gericht is op big data policing zou een impuls kunnen geven aan een versterking van de relatie tussen onderzoek en beleid c.q. praktijk. Hierbij zouden praktijk- en beleidsvragen, maar ook fundamentele onderzoeksvragen in verschillende termijnen en met verschillende doorlooptijden opgepakt kunnen worden. De inzet van publiek-private samenwerkingen als middel om tot gebalanceerde resultaten te komen, verdient een plaats in dit programma. Tot slot dient dit alles – los van de betrokkenheid bij bovenstaande initiatieven – te resulteren in borging in het (wetenschappelijk en praktijkgericht) onderwijs, waardoor de inzichten ook worden doorgegeven aan de onderzoekers en professionals van morgen.

Als het gaat om de *juridische, sociaal-technische en ethische aspecten* kan een impuls gegeven worden aan het publieke, politieke en professionele debat rond big data policing. Dit debat dient de breedte van het thema te beslaan. Het is hierbij van belang, gegeven het maatschappelijk belang van de vraagstukken, alle partijen uit de zogeheten *quadruple-helix* (overheid, wetenschap, industrie en burgers) die een belangrijke rol innemen binnen dit thema te betrekken. Tevens dient een continue evaluatie te worden gemaakt of huidige waarborgen toereikend zijn voor big data policing, en of er dus met andere woorden voldoende aandacht is voor multidisciplinaire en geharmoniseerde richtlijnen en toezichtsorganen om de rechtsstatelijke principes en beginselen van behoorlijk bestuur te waarborgen.

Kijkend naar de *organisatorische en governance-aspecten* wordt vastgesteld dat de klassieke politiefunctie en -organisaties aan verandering onderhevig zijn. Niet alleen het speelveld wordt pluriformer, ook het politiewerk als zodanig verandert – wat invloed heeft op de organisaties en de professionals die daarbinnen werken. Zo zijn er naast de diversiteit van politiewerk op zich, ook subcategorieën en specialismen, die soms transversaal zijn. Dit maakt het een complex samenstel van werkzaamheden, functies, teams, organisaties, samenwerkingsverbanden en specialismen. Er kan bijvoorbeeld technisch naar de verschillende databronnen gekeken worden (bv. gestructureerde van ongestructureerde data onderscheiden, of data indelen naar hun aard), maar de vraag is ook hoe bijvoorbeeld financiële of forensische data – waarvan de analyse veelal toekomt aan specialistische diensten – te integreren zijn. De komst van nieuwe functies (bv. data scientists en data engineers) en de randvoorwaarden die nodig zijn om deze werkzaamheden te ondersteunen en te borgen zijn van een andere aard dan in het traditionele politiewerk. De structurele omkadering van deze functies (bv. training en bijscholing, technologische vernieuwing, infrastructuur) brengt dus andere uitdagingen met zich mee. Daarnaast is het maar de vraag of deze personen met verschillende disciplinaire achtergronden dezelfde taal spreken. Zowel binnen de wetenschap als in de praktijk is ‘elkaars taal spreken’ een belangrijke randvoorwaarde om tot betekenisvolle samenwerking te komen (Keuschnigg et al., 2018; Waardenburg, 2021). Wanneer de kloof te groot is, kunnen ‘analytische vertalers’ een nieuwe maar noodzakelijke oplossing bieden.

Tot slot de *aspecten van effectiviteit en evaluatie*. Zoals in deze kadertekst beschreven ontbreekt het aan een grondige en holistische evaluatie van big data policing in de praktijk. Hierbij dient oog te zijn voor zowel primaire als secundaire (bv. sociale, juridische en ethische) effecten. Het is daarom van belang om tijdelijke experimenten mogelijk te maken om gecontroleerd maar volledig te kunnen testen wat implicaties van big-datatoepassingen zijn. Door dit getrapt in te richten (en daarmee ook te voorkomen dat toepassingen zonder meer geïmplementeerd worden zonder evaluatie vooraf) kan bepaald worden wat potentiële gevolgen zijn, hoe ongewenste

neveneffecten weggenomen of tot een acceptabel niveau gemitigeerd kunnen worden, en welke waarborgen ingeregeld moeten worden. Dergelijke experimenten worden bemoeilijkt door diverse regelgeving op verschillende niveaus (Europees, nationaal, etc.) – regels die op zichzelf noodzakelijk zijn, doch de doorgedreven evaluaties van effectiviteit bemoeilijken. Experimentele proeftuinen (conform de principes van *regulatory sandboxing*, zie bv. Kenniscentrum Data & Maatschappij, 2021) die, met inachtneming van een aantal basisvoorwaarden, worden toegelaten en beperkt zijn in tijd en omvang, zouden hier soelaas kunnen bieden. Naast deze experimenten in de werkelijkheid, kunnen methodologische innovaties eveneens soelaas bieden. Hierbij kan gedacht worden aan (adaptieve) simulatietechnieken, augmented en virtual reality of vormen van *gamification* om effecten voorafgaand aan eventuele implementatie te simuleren en evalueren.

Uit deze kadertekst mag blijken dat de evoluties op vlak van data-ontwikkeling, data-extractie en dataverwerking ons voor enorme wetenschappelijke en maatschappelijke uitdagingen stellen. Evaluatie is daarbij een cruciaal element, omdat dit de enige weg is die ons uiteindelijk kan doen besluiten of bepaalde big data policing toepassingen werkzaam en verantwoord zijn. Het zou in die zin een geruststellende gedachte zijn wanneer zowel voor- als tegenstanders zich kunnen verenigen en in het belang van wetenschap en maatschappij tot onderbouwde en door empirie ondersteunde argumenten kunnen komen.

Referenties

- Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology*, 8, 89–119. <https://doi.org/10.1093/jssam/smz056>
- Andrejevic, M. (2014) Surveillance in the big data era. In: K.D. Pimple (ed.), *Emerging Pervasive Information and Communication Technologies (PICT)* (pp. 55-69). Springer.
- Berk, R. (2013). Algorithmic criminology. *Security Informatics*, 2, Article 5. <https://doi.org/10.1186/2190-8532-2-5>
- Biemer, P. P. (2010) Total survey error: Design, implementation, and evaluation. *The Public Opinion Quarterly*, 74(5), 817-848.
- Bowers, K., & Snaphaan, T. (2022, 24 september). *New methods for old problems: Opportunities and challenges in the big data era*. Presented at the 22th Annual Conference of the European Society of Criminology. Málaga, Spain.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679.
- Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977-1008.
- Brayne, S. (2021). *Predict and surveil. data, discretion, and the future of policing*. Oxford University Press.
- Campedelli, G. M. (2022). *Machine learning for criminology and crime research: At the crossroads*. Routledge.
- Castells, M. (2000). Towards a sociology of the network society. *Contemporary Sociology*, 29(5), 693–699.
- Chan, J., & Bennett Moses, L. (2016) Is big data challenging criminology? *Theoretical Criminology*, 20(1), 21-39. <https://doi.org/10.1177/1362480615586614>
- Chen, W., Guo, X., Chen, Z., Zheng, Z., & Lu, Y. (2020). Phishing Scam Detection on Ethereum: Towards Financial Security for Blockchain Ecosystem. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (pp. 4506–4512). <https://doi.org/10.24963/ijcai.2020/621>
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1–12.
- Das, A., & Schuilenburg, M. (2020). Garbage in, garbage out: Over predictive policing en vuile data. *Beleid en Maatschappij*, 47(3), 254-268
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data. A consensual definition and a review of key research topics. *AIP Conference Proceedings*, 1644(1), 97–104. <https://doi.org/10.1063/1.4907823>
- Dencik, L., A. Hintz & Z. Carey (2018). Prediction, pre-emption and limits to dissent: Socialmedia and big data uses for policing protests in the United Kingdom. *New media & society*, 20(4), 1433-1450.
- Ferguson, A.G. (2017). *The Rise of Big Data Policing*. New York University Press.
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology*, 25(4), 435–437.
- Hardyns, W., & Snaphaan, T. (Eds.). (2020). *Big data en innovatieve methoden voor criminologisch onderzoek*. Boom.
- Henke, N., Levine, J., & McInerney, P. (2018, 1 februari). Analytics translator: The new must-have rol. *Harvard Business Review*. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/analytics-translator>
- Janssens, J., Broer, W., Crispel, M., & Salet, R. (2020). *Informatiegestuurde politie*. Gompel&Svacina.

- Joh, E. E. (2016). The new surveillance discretion: Automated suspicion, big data, and policing. *Harvard Law & Policy Review*, 10, 15–42.
- Kenniscentrum Data & Maatschappij. (2021). *brAInfood: Wat is 'legal sandboxing' voor AI?* <https://data-en-maatschappij.ai/publicaties/brainfood-wat-is-legal-sandboxing-voor-ai>
- Keuschnigg, M., Lovsjö, N. & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, 1, 3–14. <https://doi.org/10.1007/s42001-017-0006-5>
- Kitchin, R. (2014a). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Kitchin, R. (2014b). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1–12. <https://doi.org/10.1177/2053951714528481>
- Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 1–10.
- Klerks, P., & Vink-Teeven, K. (2020). De inzet van data-analysetechnologie ter bevordering van informatiegestuurde opsporing. De zoektocht naar de optimale balans tussen mens en machine. In J. Janssens, W. Broer, M. Crispel, & R. Salet (Eds.), *Informatiegestuurde politie* (Cahiers Politiestudies nr. 54, 163–175). Gompel&Svacina.
- Kool, L., Timmer, J., Royakkers, L. & van Est, R. (2017). *Opwaarderen. Borgen van publieke waarden in de digitale samenleving*. Rathenau Instituut.
- Kop, N., & Klerks, P. (2009). *Doctrine intelligencegestuurd politiewerk*. Politieacademie.
- Landman, W. (2022). De opkomst van de politiemachine. Over de invloed van opkomende technologieën. *Tijdschrift voor de Politie*, 2022(3), 6–10.
- Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- Lum, K., & Isaac, W. (2016). To predict and serve? *Significance Magazine Royal Statistical Society*, October 2016, <https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1740-9713.2016.00960.x>.
- Mali, B., Bronkhorst, C., & den Hengst, M. (2017). *Predictive policing: Lessen voor de toekomst. Een evaluatie van de landelijke pilot*. Politieacademie.
- Nantais, J. (2019, 27 september). Data science or statistics. *Towards Data Science*. <https://towardsdatascience.com/data-science-or-statistics-9e826ebf7fe2>
- Oberski, D. (2017). Sexy data science. *STAtOR*, 18(1), 34–36. <https://www.vvsor.nl/wp-content/uploads/2018/01/STAtOR-2017-1-LR-spreads.pdf>
- OM en Politie. (2020, 29 mei). *Kwaliteitskader big data*. <https://www.rijksoverheid.nl/documenten/rapporten/2020/05/29/tk-bijlage-2-kwaliteitskader-big-data>
- Openbaar Ministerie. (2017, 20 juli). *Ondergrondse Hansa Market overgenomen en neergehaald*. <https://www.om.nl/actueel/nieuws/2017/07/20/ondergrondse-hansa-market-overgenomen-en-neergehaald>
- Openbaar Ministerie. (2021, 9 maart). *Nieuwe klap voor georganiseerde misdaad*. <https://www.om.nl/actueel/nieuws/2021/03/09/nieuwe-klap-voor-georganiseerde-misdaad>
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
- Ratcliffe, J. (2016) *Intelligence-led policing*. Routledge.
- Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review*, 94, 192–233.

- Rummens, A., Snaphaan, T., Van de Weghe, N., Van den Poel, D., Pauwels, L. J. R., & Hardyns, W. (2021). Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? *ISPRS International Journal of Geo-Information*, 10(6), Article 369. <https://doi.org/10.3390/ijgi10060369>
- Sanders, C. B., & Sheptycki, J. (2017). Policing, crime and ‘big data’. Towards a critique of the moral economy of stochastic governance. *Crime, Law and Social Change*, 68(1-2), 1-15.
- Schroeder, R. (2018). Big data: Shaping knowledge, shaping everyday life. In R. Schroeder (Ed.), *Social Theory after the Internet: Media, Technology, and Globalization* (pp. 126–148). UCL Press.
- Schuilenburg, M., & Soudijn, M. (2021). Big data in het veiligheidsdomein: Onderzoek naar big data-toepassingen bij de Nederlandse politie en de positieve effecten hiervan voor de politieorganisatie. *Tijdschrift voor Veiligheid*, 20(4), 44–62. <https://doi.org/10.5553/tvv/4001>
- Snaphaan, T., & Hardyns, W. (2021a). Handvatten voor een kwaliteitsbeoordeling van big data: De introductie van het Total Error raamwerk. *Tijdschrift voor Veiligheid*, 20(4), 63–88. <https://doi.org/10.5553/TvV/.000033>
- Snaphaan, T., & Hardyns, W. (2021b). Utilizing geo-referenced imagery for systematic social observation of neighborhood disorder. *Computers, Environment and Urban Systems*, 90, Article 101691. <https://doi.org/10.1016/j.compenvurbsys.2021.101691>
- Snaphaan, T., Hardyns, W., & Spithoven, R. (2021). Big-datatoepassingen bij de politie: Een verkenning van een actueel en veelomvattend thema. *Tijdschrift voor Veiligheid*, 20(4), 3–8. <https://doi.org/10.5553/TvV/.000031>
- Spithoven, R. & Foppen, E. (2021). Big data, kleine rechtsstaat? Over de roep van uitvoerende professionals om rechtstatelijke bezinning bij big-datatoepassingen in de Nederlandse politiefunctie. *Tijdschrift voor Veiligheid*, 20(4), 24-43.
- Spithoven, R. (2020). *Verbonden risico 's. Maatschappelijke veiligheid in de black box society*. Boom criminologie.
- Spithoven, R., & Beerends, S. (2020) Veiligheid uit de glazen bol? Naar verantwoorde toepassingen van big data in het veiligheidscomplex. *Tijdschrift voor Veiligheid*, 18(3-4), 75-92.
- Spithoven, R., & van de Pas, J. (2020). Bij voorbaat effectiever? Over de noodzaak van het herwaarderen van vakmanschap en de onvermijdelijkheid van actieonderzoek bij het gebruik van big-datatoepassingen door de politie. In W. Hardyns & T. Snaphaan (Eds.), *Big data en innovatieve methoden voor criminologisch onderzoek* (pp. 131-147). Boom.
- Symons, J., & Alvarado, R. (2016). Can we trust big data? Applying philosophy of science to software. *Big Data & Society*, 3(2), 1–17.
- Van Brakel, R. (2016). Pre-emptive big data surveillance and its (dis)empowering consequences: The case of predictive policing. In B. van der Sloot, D. Broeders and E. Schrijvers (Eds.), *Exploring the boundaries of big data* (pp. 117–141). Amsterdam University Press.
- Van Brakel, R. (2020). Een reflectie over het huidig toezicht van het gebruik van surveillancetechnologie door de lokale politie in België. *Cahiers Politiestudies*, 55, 139–160.
- Van Brakel, R. (2021). Rethinking Predictive Policing. Towards a holistic framework of democratic algorithmic surveillance. In M. Schuilenburg & R. Peeters (Eds.), *The Algorithmic Society. Technology, Power and Knowledge*. Routledge.
- Van Brakel, R. (2022) De controle op het gebruik van algoritmische surveillance- onder druk? Een exploratie door de lens van de relationele ethiek. *Tijdschrift voor Mensenrechten*, (1), 23-28.

- Van Dijk, A., Hoogewoning, F., & Welten, B. (2022). Politie en rechtsstaat in de gedigitaliseerde samenleving: Filosofische bespiegelingen met praktische betekenis. In A. van Dijk, P. De Baets, L. Gunther Moor, E. Devroe & S. Zouridis (Eds.), *Politie en rechtsstaat in de gedigitaliseerde samenleving* (pp. 15–34). Gompel&Svacina.
- Van Steden, R., Roelofs, M., & Heijnen, M. (2009). *Pluriforme politiefunctie. Inventarisatie van en burgerpercepties over beveiligers, toezichthouders en handhavers*. Vrije Universiteit Amsterdam.
- van Wegberg, R., Tajalizadehkhoob, S., Soska, K., Akyazi, U., Ganan, C. H., Klievink, B., Christin, N., & van Eeten, M. (2018). *Plug and Prey? Measuring the Commoditization of Cybercrime via Online Anonymous Markets*. Paper presented at the 27th USENIX Security Symposium.
- Waardenburg, L. (2021). *Behind the scenes of artificial intelligence: Studying how organizations cope with machine learning in practice* [Doctoraal proefschrift, VU Amsterdam]. Haveka.
- Webster, F. (2002). *Theories of the Information Society* (2nd Ed.). Routledge.
- Wetenschappelijke Raad voor het Regeringsbeleid (2016). *Big Data in een vrije en veilige samenleving*. WRR.
- Wetenschappelijke Raad voor Regeringsbeleid. (2021). *Opgave AI. De nieuwe systeemtechnologie*. <https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie>
- Williams, P., & Kind, E. (2019). Data-driven policing: The hardwiring of discriminatory policing practices across Europe, ENAR, <https://www.enar-eu.org/IMG/pdf/data-driven-pro!lingweb-!nal.pdf>.