

De noodzaak van omvattende raamwerken voor begrip van, én grip op, datagedreven werken

Bijdrage van dr. mr. Erik H.A. van de Sandt aan de deskundigenbijeenkomst WG AI van de Eerste Kamer*

Samenvatting

Een omvangrijke ondergrondse beveiligingseconomie faciliteert de ondermijnende criminaliteit in Nederland. Mede hierdoor is het opsporen van deze vorm van zware, georganiseerde criminaliteit in zwaar weer gekomen. Datagedreven opsporen, waarbij gebruik wordt gemaakt van datawetenschappelijke methoden en technieken, biedt een uitkomst. Randvoorwaardelijk voor datagedreven opsporen is harmonisatie op juridisch, organisatorisch en technisch gebied, zowel binnen de politie, als met binnenlandse en buitenlandse (publieke) partners. Om deze harmonisatie te faciliteren is een gemeenschappelijke taal en raamwerk ontwikkeld door de Dienst Landelijke Recherche, internationale partners en academici. Het CSAE-raamwerk – in de Engelse taal uitgesproken als ‘See-Say’ – bestaat uit een bedrijfsproces, methodologie, dataschema, beleidsagenda en publieke waardenfilosofie. Hiermee bevordert CSAE het respecteren van de juridische kaders en ethische principes behorend bij de democratische rechtsstaat én aan de versterking hiervan. CSAE draagt zo niet alleen bij aan begrip van, en grip op, datagedreven opsporen. CSAE toont verder de noodzaak aan voor een zorgvuldige overheid om omvattende raamwerken te hebben voor werkerterreinen waar datagedreven wordt gewerkt.

Empirische rechtvaardiging voor inzet datascience: de beveiliging van ondermijnende criminaliteit In de informatiesamenleving maakt de internationaal georganiseerde criminaliteit gebruik van een gevarieerd palet aan facilitators met technische beveiligingsproducten en -diensten. Het gaat niet alleen om cryptocommunicatie voor beveiligde telecommunicatie, maar ook om *bulletproof hosting* voor beveiligde servers, *cryptocurrency mixers* voor beveiligd witwassen en *dark markets* voor beveiligde online drugshandel. Deze leveranciers bedienen grote aantallen criminele klanten die samen enorme hoeveelheden data genereren. Op het eerste oog lijkt dit een goudmijn voor de opsporing, maar niets is minder waar. Zelfs wanneer opsporingsdiensten bewijs kunnen veiligstellen via digitale- en tactische opsporingsmethoden en -technieken, zijn er nog steeds veel beveiligingsmaatregelen actief – denk aan deceptie en encryptie – die van negatieve invloed zijn op de kwaliteit en kwantiteit van bewijs. Nieuwe methoden en technieken zijn noodzakelijk om deze *big data*, of beter, *big evidence*-problemen te lijf te gaan (zie Figuur 1 in de Bijlagen). Een datawetenschappelijke kijk op bewijs is de weg voorwaarts om de enorme variëteit, tijdelijke waarde, beperkte waarheidsgetrouwheid en omvang van veiliggestelde ruwe data efficiënt en effectief te verwerken tot feitelijke processen-verbaal over wie wat doet binnen de georganiseerde misdaad. Tegelijkertijd kunnen datawetenschappelijke modellen ook bijdragen aan het waarborgen van de privacy van slachtoffers, getuigen en geheimhouders/verschoningsgerechtigden [1].

De sleutel tot datagedreven opsporen is juridische, organisatorische én technische harmonisatie De overgang naar datagedreven werken is niets minder dan een paradigmaverschuiving. Binnen de strafrechtketen kan de politie deze transitie niet alleen uitvoeren. Harmonisatie tussen partners, waaronder de Trias Politica, is hierbij het sleutelwoord. Juridische en organisatorische harmonisatie op nationaal en internationaal niveau zijn altijd al speerpunten geweest van beleidsmakers, bestuurders en wetgevers. Denk aan respectievelijk de toepasbaarheid van de Wet Politiegegevens op alle opsporingsdiensten of het ratificeren van internationale verdragen, en de oprichting van één nationale politie of deelname van Nederland aan Europol. Datagedreven opsporen vereist echter een nieuwe stap in (inter)nationale samenwerking: technische harmonisatie (zie Figuur 2). Bij deze vorm van harmonisatie ligt de nadruk op gezamenlijke dataschema's, ontologieën, extractie-transformatie-laadprocedures, datawetenschappelijke modellen en *tooling* om ruwe data te verwerken tot feiten die noodzakelijk zijn in de uitvoering van overheidstaken, zoals de bestrijding van georganiseerde criminaliteit. Technische harmonisatie op haar beurt brengt weer nieuwe juridische en organisatorische uitdagingen met zich mee [1, pp.234-255].

CSAE biedt één gemeenschappelijke taal voor harmonisatie Op allerlei plekken binnen de overheid worden momenteel data-ingenieurs, datawetenschappers en data-analisten aangenomen. Dit heeft geleid tot tal van op zich zelf staande, vaak individuele successen, maar helaas ook tot structurele fouten zoals de toeslagenaffaire. Ervaring zou ons moeten leren dat fouten maar moeizaam kunnen worden hersteld en successen zelden schaalbaar en structureel kunnen worden geborgd wanneer een gemeenschappelijk, omvattend raamwerk ontbreekt. Hoe kunnen wetgevers en rechters de uitvoering effectief controleren wanneer - om maar een voorbeeld te noemen - een eenduidig bedrijfsproces voor datagedreven werken bij deze laatste macht ontbreekt? Met andere woorden, voor begrip van en grip op datagedreven werken is één gedeelde taal noodzakelijk. Om deze reden is de Dienst Landelijke Recherche samen met internationale partners en academici begonnen met het ontwikkelen en implementeren van een visie op, en een gemeenschappelijk raamwerk voor, deze innovatieve vorm van opsporen. Het raamwerk is gebaseerd op verschillende standaarden uit

*Operationeel specialist bij Team High Tech Crime, Dienst Landelijke Recherche, Landelijke Eenheid van de nationale politie.

Kwartiermaker Cybercrime bij het Korps Politie Caribisch Nederland.

Research fellow in cyber security aan de University of Bristol, Verenigd Koninkrijk.

Practitioner-in-residence bij het National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online, Verenigd Koninkrijk.

E-mail: erik.van.de.sandt [at] politie.nl & ev18710 [at] bristol.ac.uk

de *cybersecurity*-, *intelligence*- en *datamining*-gemeenschap, en aangepast aan de doel- en taakstellingen van de politie. Het raamwerk – genaamd CSAE – staat voor de vier stappen uit het cyclische bedrijfsproces (zie Figuren 3 en 4). In de *Collect*-fase wordt ruwe data verzameld, in *Store* wordt de verzamelde data genormaliseerd en opgeslagen als informatie, in *Analyze* worden relevante informatiepunten bij elkaar gebracht tot clusters en kennis hierover verwerkt tot intelligence, en in *Engage* wordt deze intelligence uitgelopen totdat er interventies kunnen worden uitgevoerd die gebaseerd zijn op feiten. Immers, een zorgvuldige, geloofwaardige en verantwoorde overheid baseert interventies op feiten, of deze nu handhaving, toezicht of opsporen betreffen. CSAE is omvattend: naast een gedetailleerd bedrijfsproces bestaat het raamwerk verder uit een methodologie, dataschema, beleidsagenda's en – zeer belangrijk – publieke waardenfilosofie [2].

Publieke waardenfilosofie als kern van elk omvattend raamwerk Datagedreven werken als versterking van het opsporen van ondermijnende criminaliteit draagt direct bij aan het versterken van onze democratische rechtsstaat. Tegelijkertijd komen er in de toepassing van datagedreven opsporen vraagstukken naar voren die gerelateerd zijn aan wezenlijke onderdelen van diezelfde democratische rechtsstaat. De publieke waardenfilosofie is onlosmakelijk verbonden met alle componenten van CSAE en loopt als een rode draad door het raamwerk. Vanuit de publieke waardenfilosofie zijn bijvoorbeeld de dataschema's en bijbehorende datanormalisatiesoftware in de Opslaanfase van het bedrijfsproces sleuteltechnologieën die opsporingsdiensten beter zelf kunnen ontwikkelen om een *vendor-lock* te voorkomen. Een organisatorische beleidsagenda moet bij een veranderend personeelsbestand ook oog hebben voor cognitieve diversiteit onder data-ingenieurs, datawetenschappers en data-analisten (zie Figuur 5). Het is vanwege onze publieke waardenfilosofie dat de bijbehorende whitepaper openbaar is gepubliceerd [2]. Het vormt een eerste stap om wetenschappelijk onderzoek met én naar de politie te stimuleren, het aantal publieke en private partners uit te breiden en het algemene publiek te informeren. Het CSAE-raamwerk draagt bij aan begrip van, én grip op, datagedreven opsporen in het bijzonder en datagedreven werken in het algemeen [3].

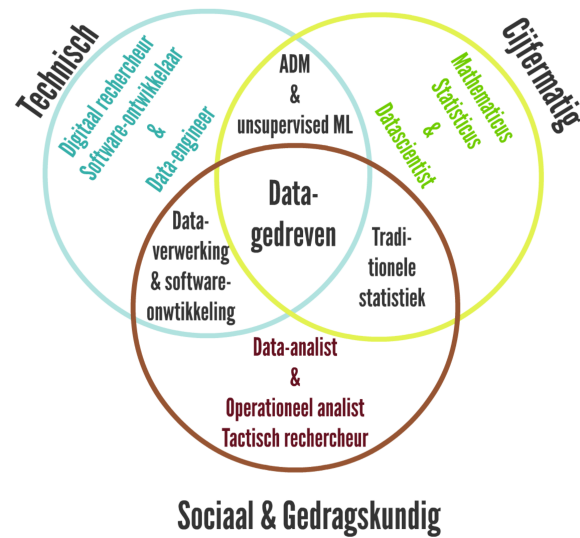
Voorbeeld: begrip van privacy Het voeren van het maatschappelijk debat over juridische en ethische principes van datagedreven werken is noodzakelijk in onze democratische rechtsstaat. Zo is privacy een onderwerp van voortdurende zorg, ook met betrekking tot de inzet van datascience bij de uitvoering van de politietoek. CSAE kan met het eenduidige bedrijfsproces een bijdrage leveren aan de verdieping van het debat door het abstracte begrip privacy meer concrete betekenis en inhoud te geven in elke fase van datagedreven opsporen. Wanneer het over privacy in de verzamelfase gaat is dataminimalisatie van belang, terwijl in de opslaafase dataretentie en -bescherming, zoals bewaartermijnen en toegangscontrole, een rol spelen. Privacy in de analysefase heeft betrekking op doelbinding wanneer politiegegevens uit verschillende onderzoeken bij elkaar worden gebracht. In de laatste fase van bestrijden heeft een verdieping van het begrip privacy betrekking op het verstrekken van politiegegevens aan derden.

Voorbeeld: grip op datawetenschappelijke modellen Een ander voorbeeld van de bijdrage van CSAE aan de rechtsstaat is het beter kunnen reguleren van de te gebruiken datawetenschappelijke methodologieën (zie Figuur 6). Met andere woorden, CSAE geeft wetgevers en rechtsprekenden grip op data, methoden en technieken en uitkomsten. Deze regulering zou kunnen bestaan uit een viertrapsraket. Allereerst moeten datagedreven methodologieën voldoen aan de grondbeginselen van onze democratische rechtsstaat, zoals het voorkomen van discriminatie. Daarna moeten methodologieën voldoen aan de grondprincipes van goed wetenschappelijk onderzoek, te weten validiteit en betrouwbaarheid wanneer een kwantitatieve aanpak wordt gebruikt en geloofwaardigheid in het geval van kwalitatief onderzoek. Verder moet over de gebruikte methodologieën op een correcte wijze normatief geredeneerd worden. Met andere woorden, empirisch onderzoek moet op een correcte wijze worden vertaald naar een juridische (in casu, strafrechtelijke) context, zoals het voorkomen van forensische drogredeneringen (denk aan *prosecutor's fallacy*). Naast deze bestaande uitgangspunten zijn er ook nieuwe juridische principes nodig, zoals *human-in-command* (de verplichting dat mensen resultaten controleren en de conclusies trekken), uitlegbaarheid en transparantie. Er zijn vormen van datawetenschap, zoals *automated decision making* (ADM), waarbij controle door mensen in theorie niet nodig is. Dergelijke datawetenschap worden door de minister van Justitie en Veiligheid momenteel als onwenselijk voor de opsporing beschouwd. Er zijn echter situaties denkbaar waarbij bijvoorbeeld ADM in de uitvoering van opsporingsbevoegdheden (zie Figuur 1) wel wenselijk is.

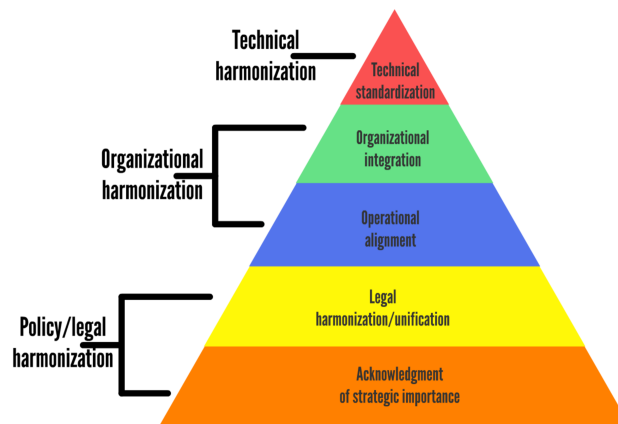
Voorbeeld: toezicht op datagedreven werken CSAE stimuleert nadrukkelijk (denken over) toezicht op datagedreven werken door bijvoorbeeld de opname van verschillende evaluatiemomenten in het bedrijfsproces. CSAE roept ook vragen op over verantwoordingsmechanismen (*accountability*) binnen de opsporing. Deze richten zich nu op de integriteit - waaronder kennis en kunde - van politiemedewerkers middels bijvoorbeeld toetsing en certificering, digitaal-forensische software en bewijs, maar zouden zich ook kunnen richten op datamodellen (denk hierbij aan het Kwaliteitskader Big Data van het Openbaar Ministerie en de nationale politie [4]). Omdat de datawetenschappelijke methodologie van CSAE in elke fase van het bedrijfsproces kan worden ingezet roept dit ook de vraag op of dergelijke algoritmeverantwoording niet alleen door de rechter aan het einde bij de fase van bestrijden, maar ook in overige fases moet worden ingevoerd, en wanneer dit wenselijk is, in welke mate, met welke mechanismes en door wie.

Tot slot Bij het geleidelijk invoeren van datagedreven werken binnen verschillende werkterreinen van de overheid zullen doorlopend nieuwe vraagstukken opkomen die betrekking hebben op rechtsethiek en (implementatie van) wet- en regelgeving. Een gemeenschappelijk, omvattend raamwerk als CSAE biedt helderheid en ondersteuning bij het voeren van het maatschappelijke debat over deze onderwerpen. Zo bezien draagt het CSAE-raamwerk, vanuit de publieke waardenfilosofie, bij aan het respecteren van de juridische kaders en ethische principes behorend bij de democratische rechtsstaat én aan de versterking hiervan.

Bijlagen



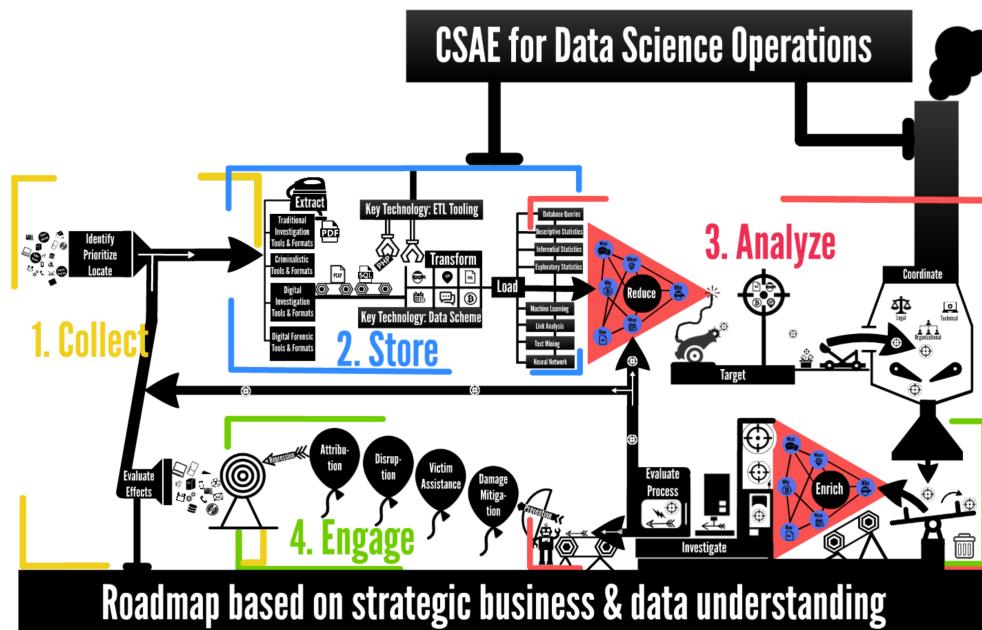
Figuur 1: Naast de sociale en gedragskundige kijk van tactisch (traditioneel) opsporen en de technische kijk van digitaal opsporen is ook een meer cijfermatige kijk op bewijs nodig om de kwantiteit van rechtmatig verkregen data te verminderen en de kwaliteit van bewijs te verhogen. Een cijfermatige kijk betekent niet alleen de inzet van traditionele statistiek (zoals bij *intelligence-led policing* uit de jaren '90 van de vorige eeuw), maar ook het gebruik van meer geavanceerde algoritmes op basis van – onder andere – kunstmatige intelligentie. Deze nieuwe methoden en technieken staan niet op zichzelf, maar leiden juist in combinatie met bestaande tactische en digitale perspectieven tot een nieuw paradigma in opsporingsland: datagedreven opsporen. De digitaal onderzoekers verzamelen data, de data-ingenieurs normaliseren de data, de software-ontwikkelaars ontwikkelen tools om de data inzichtelijk te krijgen. De tactische onderzoekers kunnen met deze tools de opgeschoonde informatie bekijken door handmatige zoekslagen uit te voeren, terwijl operationeel analisten relaties leggen. Ook kunnen deze onderzoekers en analisten input geven aan *datascientists* die middels geavanceerde algoritmes patronen in datasets blootleggen. De resultaten van deze modellen worden gecontroleerd en geïnterpreteerd door data-analisten die de resultaten vertalen naar concrete acties voor tactisch of digitaal onderzoekers.



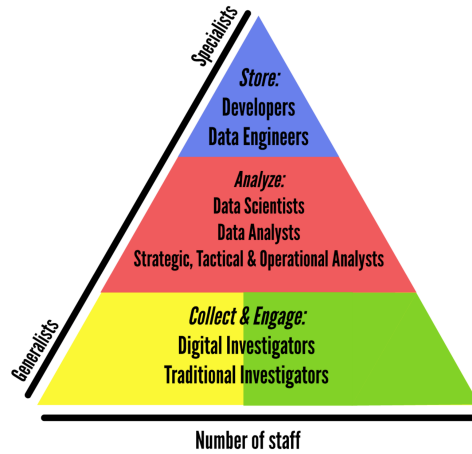
Figuur 2: Harmonisatie impliceert afstemming en samenwerking met respect voor de eigen cultuur van organisaties. De term impliceert ook een mate van flexibiliteit en wendbaarheid die nodig is in snel veranderende omgevingen waar wetgevers tegelijkertijd invloed moeten kunnen uitoefenen met betrekking tot regulering en toezicht. Uiteindelijk kan harmonisatie leiden tot standaardisatie – lees: bindende afspraken.



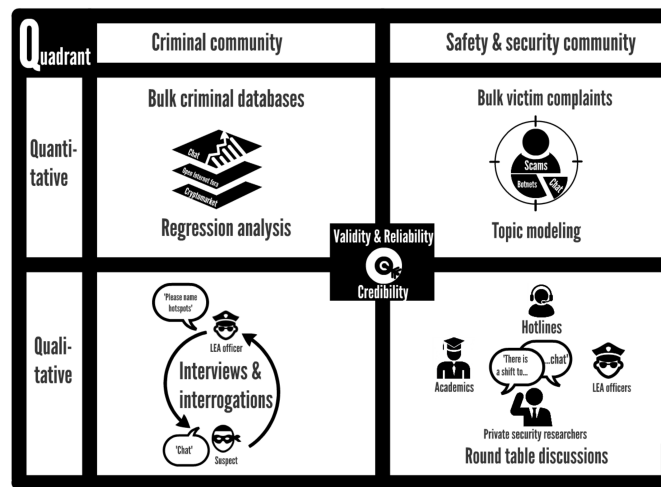
Figuur 3: Het cyclische bedrijfsproces van datagedreven opsporen is verder gedetailleerd en visueel uitgewerkt in een praatplaat, zie Figuur 4. Deze vier stappen zullen ook opgaan voor andere terreinen binnen de overheid waar datagedreven wordt gewerkt. Immers, een zorgvuldige, geloofwaardige en verantwoorde overheid baseert interventies op feiten, of deze nu handhaving, toezicht of opsporen betreffen. Vanuit de externe controletoelaak van de wetgevende en rechtsprekende macht en interne controlemechanismes binnen de uitvoerende macht dragen deze vier fasen bij aan kwalitatieve en kwantitatieve verantwoording. Wat voor en hoeveel datasets zijn er verzameld en waarom? Wat voor en hoeveel informatiepunten zijn er opgeslagen en waarom? Wat voor en hoeveel analytische intelligenceproducten zijn er gemaakt en waarom? Wat voor en hoeveel interventies zijn er uitgevoerd en waarom?



Figuur 4: Bovenstaande praatplaat faciliteert één gemeenschappelijke taal voor verschillende disciplines - denk aan bestuurders, academici en inhoudelijk specialisten - om op een gedetailleerd niveau begrip van, en grip op, datagedreven opsporen te krijgen. Naast dat elke fase verschillende stappen, kernwaarden en doelstellingen heeft, zijn de vier fasen verder gebaseerd op gangbare organisatiestructuren, bestaande juridische kaders en nieuwe en bestaande functieprofielen (zie bijvoorbeeld Figuur 5). Ook is er overlap die de samenwerking tussen fasen weergeeft. Rechercheurs die data verzamelen moeten immers overleggen met collega's die deze data gaan normaliseren en opslaan als informatie. Wanneer de opgeslagen informatie nodig is voor analysedoeleinden dient er overleg te zijn tussen data-ingenieurs (Store) en datawetenschappers (Analyze) welke informatiebronnen er zijn, hoe en op welke wijze deze bij elkaar kunnen en mogen worden gebracht, et cetera.



Figuur 5: CSAE geeft per fase van het bedrijfsproces overzichtelijk de aard en omvang weer van bestaande en nieuwe politiefuncties. *Nota bene*, hoewel datawetenschappers en data-analisten maken en interpreteren modellen in alle fases leert de praktijk dat zij hoofdzakelijk werkzaam zijn in de analysefase en daaraan gerelateerde politieafdelingen.



Figuur 6: Uitgangspunt van de datawetenschappelijke methodologie van CSAE - genaamd Quadrant - is dat kwantitatieve en kwalitatieve methoden en technieken gecombineerd toegepast worden op het criminele domein én het veiligheidsdomein. Het veiligheidsdomein moet hierbij breed beschouwd worden en omvat partijen als private (beveiligings)organisaties, universiteiten, slachtoffers en getuigen. Deze gemengde methodenbenadering (*mixed-methods approach*) creëert strategisch, tactisch en operationeel zicht - meer specifiek: inzicht, overzicht, vooruitzicht en terugblik - in elke fase van het bedrijfsproces. Zo creëert de Quadrant-methodologie onder andere zicht op wat moet worden verzameld, wat moet worden opgeslagen, wat moet worden geanalyseerd en wat moet worden bestreden.

References

- [1] E. Van De Sandt, *The Deviant Security Practices of Cyber Crime*. Leiden, The Netherlands: Brill | Nijhoff, 2021. [Online]. Available: <https://brill.com/view/title/60184>
- [2] E. Van de Sandt, A. Van Bunningen, J. Van Lenthe, and J. Fokker, "Towards Data Scientific Investigations: A Comprehensive Data Science Framework and Case Study for Investigating Organized Crime Serving the Public Interest," in *Third INTERPOL-UNICRI Global Meeting on AI for Law Enforcement*. Virtual conference: INTERPOL-UNICRI, 2020. [Online]. Available: <https://cpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/1/670/files/2021/03/White-Paper-Towards-Data-Scientific-Investigations.pdf>
- [3] E. Van de Sandt, S. Van der Maden, P. De Bruijn, R. Westerhof, and M. Den Hengst, "Datagedreven opsporen: Nieuwe loot aan de stam in de bescherming van de rechtsstaat." *Politie Cahiers (verwacht)*.
- [4] Openbaar Ministerie and Nationale Politie, "Kwaliteitskader Big Data," Tech. Rep., 2020. [Online]. Available: <https://www.rijksoverheid.nl/documenten/rapporten/2020/05/29/tk-bijlage-2-kwaliteitskader-big-data>