



Big Data: een zoektocht voor beleid

Datum:

Utrecht, 11 februari 2015

Auteurs:

Dr. Frank Bongers

Jaap Veldkamp MSc.

Ir. Tommy van der Vorst

Inhoudsopgave

| | |
|---|-----------|
| Samenvatting | 3 |
| English summary | 4 |
| 1 Inleiding | 5 |
| 2 Wat is big data? | 6 |
| 3 Rollen van de overheid..... | 7 |
| 3.1 Kaderstellen en controleren..... | 7 |
| 3.2 Aanjagen | 7 |
| 3.3 Produceren, verzamelen en verstrekken | 8 |
| 3.4 Actief gebruiken | 8 |
| 4 Impact op beleidsonderzoek | 9 |
| 4.1 Informatie op basis van 'alles' en 'overal' | 9 |
| 4.2 Meer 'real time' informatie | 9 |
| 4.3 Kansen voor beleidsevaluatie | 10 |
| 5 Uitdagingen voor beleidsonderzoek | 11 |
| 6 Tot slot..... | 14 |
| Literatuur | 15 |

Samenvatting

In dit artikel verkennen we de (potentiële) impact van big data op beleid en beleidsonderzoek. Digitale hulpmiddelen genereren steeds sneller, steeds meer en steeds vluchtiger gegevens (big data sets) die middels onverwachte en creatieve combinaties van big data ingezet kunnen worden om beleid te ondersteunen en onderzoeksvragen te beantwoorden. Vooralsnog voeren voornamelijk informatiekundige en economische perspectieven op big data de boventoon. Dit betoog schetst in aanvulling op deze waardevolle perspectieven een meer beleidskundige kijk op dit fenomeen. Het is in onze optiek vooral belangrijk om verbindingen te leggen tussen eigenaars van big data sets, potentiële gebruikers van big data sets en beleidsproblemen zodat de kansen ontstaan om middels big data maatschappelijke problemen op te lossen en innovaties (ook in beleidsonderzoek) te realiseren. Het realiseren van deze verbindingen is nog een zoektocht voor bestuurders en onderzoekers.

Kernwoorden

Big data, beleid, beleidsonderzoek, digitalisering, innovatie.

English summary

This article explores the (potential) impact of big data on policy and policy research. Smart devices generate an ever-increasing volume, variety and velocity of data (so called big data). The expectation is that big data will play a significant role in policy-making processes and policy research. However, to capture the potential big data can provide for policy making, we need unanticipated and creative combinations of big data sets. As the ICT and economic perspectives on big data have predominated up till now, a more policy science perspective is required. It is in our view especially important to link the owners as well as the potential users of big data sets and policy issues in order to create windows of opportunities for solving societal problems and achieving innovations. Policy makers and researchers continue their quest to identify and make use of these opportunities.

Key words

Big data, policymaking processes, policy research, digitalisation, innovation.

1 Inleiding

Het volume, de snelheid en de variatie van data die dagelijks verzameld en bewerkt worden, groeien exponentieel. IBM (2014) schat dat 90% van alle data in de wereld in 2013 in de twee jaren daarvoor waren gecreëerd. Dagelijks worden 2,5 miljard gigabytes aan data gecreëerd, genoeg om 27.000 tablets per minuut te vullen. De betekenis van deze enorme hoeveelheid data voor samenleving, economie, innovatie en onderzoek wordt hoog ingeschat; ook door beleidsmakers en onderzoekers (zie bijvoorbeeld Australian Government, 2013; HM Government 2013; Vlaamse Raad voor Wetenschap en Innovatie, 2014). Deze grote hoeveelheden data zijn vaak een product van informatievergaring zonder vaststaand doel. Bedrijven en overheden houden bijvoorbeeld vaak grote hoeveelheden gegevens bij zonder vooraf een duidelijke toepassing daarvan voor ogen te hebben. Deze data krijgen pas meerwaarde wanneer zij bijdragen aan het oplossen van een bedrijfsmatig of maatschappelijk probleem. Daartoe is het noodzakelijk dat er een koppeling ontstaat tussen big data sets enerzijds en genoemde uitdagingen anderzijds. Het identificeren van deze koppelingen is een zoektocht die omgeven wordt met bestuurlijke, praktische en methodologische uitdagingen.

Het is de vraag hoe bestuurders en onderzoekers omgaan met de kansen die big data kan bieden. De beschikbaarheid over een grote hoeveelheid data betekent geenszins dat beleid ook beter wordt of dat het effect van beleid beter bepaald kan worden. Bovendien zijn er waarborgen nodig om big data niet te laten vervallen in "Big Brother". Het is aannemelijk dat big data ook onderdeel wordt van de politiek-bestuurlijke arena waarbinnen beleid tot stand komt. Hierover is vooralsnog beperkt op gereflecteerd. Dit artikel is een aanzet om deze leemte op te vullen. Centraal staat de vraag wat de (potentiële) impact van big data op beleid en beleidsonderzoek is.

2 Wat is big data?

Een functionele definitie van big data is “het vermogen van de samenleving om informatie op nieuwe manieren in te zetten voor het verkrijgen van nuttige inzichten of waardevolle goederen en diensten” (Mayer-Schönberger & Cukier, 2013, p. 10). We maken nadrukkelijk een onderscheid tussen big data en big data sets. Big data gaat over de slimme combinatie en analyse van zeer grote gegevensbestanden (big data sets) die leiden tot nieuwe inzichten, diensten en innovaties. Big data sets (de input voor big data) hebben drie cruciale eigenschappen, namelijk:

1. Groot volume aan data.
2. Grote snelheid waarin deze data verzameld en getransporteerd worden.
3. Grote variatie van de data.

Big data sets zijn te groot, te vluchtig en te complex om met traditionele “hands on” databasemanagement tools te bewerken. In dit perspectief zijn big data sets ook een relatief begrip, omdat technische ontwikkelingen de capaciteiten voor opslag, transport en bewerking continu doen verbeteren.

De voortschrijdende digitalisering gaat gepaard met twee belangrijke ontwikkelingen. Allereerst zijn er steeds meer apparaten om ons heen die middels sensoren, antennes, netwerken en applicaties continu gegevens verzamelen, opslaan en transporteren. Het gaat dan al lang niet meer om gewone informatietechnologie (computers, tablets, smartphones), maar ook om kleding, meubels, gebouwen, auto's en andere producten die met elektronica worden uitgerust, ondersteund en gekoppeld aan netwerken. Waar in het verleden data vooral bewust werd verzameld, hebben ontwikkelingen in de digitalisering ervoor gezorgd dat er nu grote hoeveelheden (operationele) data zijn die van grote waarde kunnen zijn, maar momenteel niet of nauwelijks gebruikt worden. Daarnaast hebben de huidige computers ook de capaciteit om complexe analyses te kunnen uitvoeren op deze data. De combinatie van deze twee ontwikkelingen lijkt onbegrensde mogelijkheden te bieden om nieuwe kennis te ontwikkelen, patronen te detecteren, voorspellingen te doen, etc. Het gaat dan om toepassingen die in het verleden helemaal niet mogelijk waren, bijvoorbeeld grootschalige netwerk- en sentimentanalyses.

Er kleven mogelijk ook nadelen aan big data. We kunnen technologisch goed omgaan met big data sets, maar het is een uitdaging om big data sets om te zetten in maatschappelijke en economische meerwaarde en dat binnen de kaders van betrouwbaarheid en veiligheid. Dat vraagt om ambitie, samenwerking en een zoektocht om kansen te grijpen. De overheid kan hier een cruciale rol spelen.

3 Rollen van de overheid

De overheid kan vier rollen vervullen in het domein van big data: (1) kaderstellen en controleren; (2) aanjagen; (3) produceren, verzamelen en verstrekken; en (4) actief gebruiken. Deze rollen staan op volgorde van de mate waarin een overheid big data adopteert. Hoe zwaarder een overheid inzet op big data, hoe meer (inhoudelijke en technische) kennis zij nodig heeft. Kaderstellende en controlerende activiteiten in het kader van big data zijn in veel gevallen een logisch vervolg op bestaande activiteiten op dat vlak en vormen daarmee een logisch instappunt voor een overheid.

3.1 Kaderstellen en controleren

De overheid kan als voorbereider, uitvoerder en handhaver van wetgeving regels stellen ten aanzien van het gebruik van big data sets. Dit vereist dat de wetgever duidelijke regels stelt. De verzameling van talloze gegevens over ons gedrag kan een bedreiging vormen voor onze privacy. Zeker wanneer derden deze gegevens over ons opslaan, onderling delen en gebruiken om ons leven 'aangenamer' te maken of te voorspellen. In veel gevallen kan dit leiden tot betere en op maat gemaakte diensten, maar de keerzijde is dat er ook persoonlijke informatie gebruikt wordt die men liever persoonlijk houdt.

Veel big data sets zijn niet met overheidsgebruik in het achterhoofd ontwikkeld. De overheid dient methoden te zoeken die het gebruik van externe big data sets mogelijk maken. Nagegaan moet worden of bestaande procedures op het gebied van databescherming en privacy bij verwerking van data aangepast moeten worden voor big data. Big data bestaat immers ook uit het hergebruiken van informatie die voor andere doeleinden is verzameld. Hoe kan worden voorkomen dat resultaten uit analyses van big data sets meer informatie openbaren dan strikt noodzakelijk? Dit is ook van toepassing wanneer de overheid big data sets betreft van private partijen en deze combineert met eigen (gevoelige) big data sets, en daarna de resultaten terugkoppelt.

3.2 Aanjagen

De overheid kan het gebruik van big data aanjagen. Voorbeelden van het aanjagen door de overheid zijn:

- Agenderen van het belang van big data middels het formuleren van ambities over verzamelen en ontsluiten van big data. De Australische overheid stelt bijvoorbeeld dat "big data offers organisations widespread potential opportunities and benefits. While the magnitude and nature of the value varies depending on industry sector, it is anticipated that government will be able to realise substantial productivity and innovation gains from the use of big data" (Australian Government, 2013, p. 13). Ook de Britse overheid wijst op de potentiële voordelen van big data, want "[t]he volume, velocity and variety of data being created and analysed globally is rising every day, and using data intelligently has the potential to transform public sector organisation, drive research and development, and enable market-changing products and services. The social and economic impact is significant." (HM Government, 2013, p. 5).
- Stimuleren van onderzoek en ontwikkeling in het domein van big data, bijvoorbeeld middels het co-financieren van publiekprivate onderzoeksprogramma's.

- Verzamelen en verspreiden van “best practices” als onderdeel van een bredere strategie die gericht is op het stimuleren van de bewustwording over de kansen en bedreigingen die big data bieden. Een vergelijkbare strategie is gevolgd in het kader van open data sets van het Ministerie van Economische Zaken (Dialogic & The Green Land, 2012).
- Een actieve rol vervullen in standaardisatieprocessen voor big data. Te denken valt aan het vaststellen van uitwisselingsstandaarden voor gegevens in de zorg- of transportsector. In veel gevallen is de overheid reeds betrokken bij dergelijke trajecten, zoals het Elektronisch Patiëntendossier (EPD) en de OV-chipkaart.

3.3 Produceren, verzamelen en verstrekken

De overheid beschikt over duizenden datasets die dagelijks aangevuld worden met nieuwe gegevens. Deze datasets zijn vaak wel gefragmenteerd (in beheer bij honderden verschillende overheidsinstellingen, zoals gemeenten, provincies, ministeries en agentschappen). De overheid kan deze datasets intern delen en – als onderdeel van een open data strategie – beschikbaar stellen aan externen. Dit impliceert dat de overheid big data sets – waar mogelijk – behandelt als open data. Open data zijn (1) uit publieke middelen bekostigd; (2) gegenereerd bij of voor de uitvoering van een publieke taak, (3) openbaar; (4) vrij van auteursrechten of andere rechten van derden; (5) computer-leesbaar; (6) voldoen bij voorkeur aan de ‘open standaarden’ (geen pdf, wel xml of csv); en zijn voor hergebruik beschikbaar zonder beperkingen, zoals kosten of verplichte registratie (Algemene Rekenkamer, 2014).

3.4 Actief gebruiken

De overheid zal de big data sets waarover zij beschikt ook bewerken en gebruiken. Maar ook andere datasets zijn relevant voor beleid en uitvoering. Denk bijvoorbeeld aan (particuliere) pensioenfondsen of verzekeraars die gegevens over de pensioenopbouw en inkomensverzekeringen van individuele Nederlanders beveiligd ter beschikking stellen. Ten opzichte van private partijen zijn overheden in de unieke positie dat zij desnoods ondernemingen kunnen verplichten om gegevens beschikbaar te stellen, bijvoorbeeld middels nationale bureaus voor de statistiek, de belastingdienst, burgerzaken en het Handelsregister.

In deze en de voorlaatste rol moet de overheid beschikken over een informatietechnologie(-infrastructuur) voor big data. De ongestructureerde manier waarop big data sets zijn opgeslagen zorgen ervoor dat er geen standaard tools voor verwerking beschikbaar zijn. Dit betekent dat de overheid goed moet nadenken hoe zij de tools voor verwerking binnen en tussen overheden categoriseren en bewaren, zodat analyses hergebruikt kunnen worden. Aanvullend kan de inzet van big data om aanvullende investeringen vragen in mensen en middelen. De overheid kan overigens gebruik maken van private infrastructuur (private clouds) bij het verwerken van big data. Zaken die hierbij spelen zijn (onder andere) de geografische locatie van de cloud (in verband met wetgeving) en gegevensbescherming.

4 Impact op beleidsonderzoek

Beleid is het resultante van een beleidsproces waarin allerlei maatschappelijke krachten spelen en het brengt zelf ook weer allerlei maatschappelijke krachten op gang. Dat proces kenmerkt zich door een dynamisch verloop van handelingen en interacties met betrekking tot het beleid (Hoogerwerf & Herweijer, 1998, p. 26). Het gaat daarbij om dynamiek bestaande uit een reeks van onderling samenhangende handelingen, maar ook om wederzijdse beïnvloeding tussen de factoren (macht en informatie) en actoren die tot het proces behoren. En hier zien we dus de term informatie om de hoek komen kijken en dat heeft een raakvlak met big data. Informatie is een belangrijke bron waarop beleidskeuzes berusten (naast macht, intuïtie, ervaring, ...). De traditionele kijk op informatie in het beleidsproces veronderstelt dat het beleid gevoed wordt met (beleids)onderzoeken en periodieke statistische, economische, ruimtelijke analyses door de overheid gefinancierde onderzoeksinstituten (CBS, CPB, PBL, ...), universiteiten, kennisinstellingen en particuliere onderzoeks- en adviesbureaus. Big data kan hier verandering in brengen.

4.1 Informatie op basis van 'alles' en 'overall'

Steeds meer apparaten in onze omgeving verzamelen en ontsluiten informatie over 'alles' en 'overall'. Dit zal in sommige gevallen traditioneel statistisch onderzoek, dat vanwege methodologische beperkingen en middelen zich vaak beperkt tot steekproeven, overbodig maken. Mayer-Schönberger en Cukier wijzen op drie verschuivingen in statistisch onderzoek (2013, pp. 32-33):

1. De mogelijkheid om enorme hoeveelheden gegevens over een onderwerp te analyseren in plaats van noodgedwongen te werken met kleinere verzamelingen.
2. De bereidheid de rommeligheid van gegevens uit de werkelijkheid te accepteren in plaats van de voorkeur te geven aan exactheid.
3. Toenemend respect voor correlaties in plaats van een voortgaande zoektocht naar een ongrijpbare causaliteit.

Een bijkomend gevolg is dat deze apparaten en de gegevens die zij genereren de kans bieden om nieuwe en betere indicatoren te meten, bijvoorbeeld indicatoren die een meer betrouwbaar beeld geven van kennis, gedrag of houding van een doelgroep en bovendien altijd (real time) gemeten kunnen worden. Denk bijvoorbeeld aan de schatting van het aantal demonstranten. Verkeersgegevens van mobiele operators kunnen een veel nauwkeuriger beeld geven van het aantal demonstranten (en hun verplaatsingsgedrag) dan helikoptervluchten door de politie (die overigens meer doeleinden dienen dan alleen het aantal demonstranten te schatten). De onderzoeker moet dan wel toegang krijgen tot deze data (en dat is niet altijd vanzelfsprekend, laat staan toegestaan).

4.2 Meer 'real time' informatie

Steeds meer data die relevant (kunnen) zijn voor het beleid komen in grote getale en real-time beschikbaar en big data tools bieden kansen om sneller complexere patronen te detecteren. Dit impliceert dat beleidsmakers niet altijd hoeven te wachten op resultaten van beleidsonderzoek waarin eerst informatie wordt verzameld, bewerkt en vervolgens gerapporteerd (denk aan monitorings- en evaluatierapportages). Dit zal vooral indicatoren

raken die voortaan real-time gemeten kunnen worden in plaats van periodiek. Het inflatiecijfer – als voorbeeld – wordt meestal per maand, kwartaal of jaar bepaald aan de hand van het vergelijken van prijzen in de tijd in een groep representatieve producten en diensten. Webtools maken het echter mogelijk om continu de prijzen te verzamelen en te vergelijken of – bij wijze van spreken – *alle* producten en diensten die op internet worden aangeboden. Uiteraard moet hierbij rekening gehouden worden met welke producten relatief veel of weinig op internet worden verkocht, geografische grenzen die op internet ontbreken, maar voor inflatiecijfers wel relevant zijn. Big data is dus een stimulans voor de automatisering van beleidsonderzoek die al zeker een decennium geleden is ingezet (zie ook Van Hoesel, Leeuw & Mevissen, 2005, pp. 163 – 167). Onderzoeksvragen worden steeds meer beantwoord via gedigitaliseerde kanalen (denk aan zogenaamde management dashboards die real-time sturingsinformatie ontsluiten). Dan hoeven er geen periodieke rapporten opgeleverd te worden, maar hebben beleidsmakers altijd en overal toegang tot relevante indicatoren. De vraag blijft natuurlijk wel overeind staan hoe fluctuaties in scores op deze indicatoren verklaard kunnen worden. Daar zal beleidsonderzoek voor nodig blijven evenals het duiden van cijfers, kwalitatieve informatie, het formuleren van onderzoeksvragen en –aanpakken, etc.

4.3 Kansen voor beleidsevaluatie

Big data bieden nieuwe en vaak ook betere informatie om bijvoorbeeld beleidsevaluaties uit te voeren. Dit speelt bijvoorbeeld bij het verzamelen van meer gegevens over beleidsprocessen op basis van betere indicatoren. Triangulatie zal echter noodzakelijk blijven, omdat big data niet alle aspecten van een evaluatie zal afdekken en dat achter de verzamelde cijfers kwalitatieve duidingen en interpretaties nodig blijven. Ook wordt het attributieprobleem niet zo maar opgelost. We kunnen met steeds meer data steeds meer indicatoren beter meten, maar dat beantwoordt niet de vraag of een wijziging in een score op een prestatie-indicator veroorzaakt door het beleid of dat andere factoren een rol spelen. Daarvoor blijven geavanceerde evaluatiemethoden als natuurlijke experimenten met experimentele en controlegroepen noodzakelijk (zie Webbink, 2006).

5 Uitdagingen voor beleidsonderzoek

De genoemde impact van big data stelt beleidsonderzoekers voor een aantal nieuwe uitdagingen. Dit in de context dat we te maken hebben met een beweeglijke trend waarvan nog niet duidelijk is hoe groot de impact zal zijn. Nu zijn de verwachtingen ten aanzien van big data hoog, maar de vraag of zij echt productief kan bijdragen aan beleidsonderzoek kan nog niet helemaal goed beantwoord worden. De verandering (als daarvan sprake zal zijn) van beleidsonderzoek waarin big data een meer prominente rol zal vervullen, zal niet vanzelf verlopen. Big data vraagt ook om nieuwe kennis en het verrichten van een zoektocht om kansen van big data te detecteren. Nieuwe kennis richt zich op dataspecialisten die in staat moeten zijn met (nieuwe) hulpmiddelen meerwaarde te creëren uit de brei van ongestructureerde gegevens. De zoektocht is een iteratief proces waarbij geschakeld moet worden tussen enerzijds de vraag (wat is de informatiebehoefte en in welke mate kan big data – dat wellicht ergens opgesloten zijn maar nog niet gekend – inspelen op deze behoefte) en anderzijds het aanbod (er liggen heel veel data op de plank, maar dat leidt niet automatisch tot maatschappelijke meerwaarde).

Hierbij zit ook onze huidige kijk op beleid en beleidsprocessen in de weg. In de afgelopen decennia is het gangbare model waarbij beleid als een logische reeks van vervolgstappen wordt gepresenteerd en geanalyseerd in onmin geraakt. In dit model leefde de beleidsbepaler in een relatief overzichtelijke wereld en was hij 'alleswetend' en 'allesbepalend'. Rationeel-wetenschappelijk onderzoek voedde het beleidsproces en leidde tot objectief beleid (Beschrijvingen van dit type modellen zijn te vinden in Hoogerwerf & Herweijer, 1998; Dunn, 1994; John, 2005). Inmiddels is duidelijk dat beleid veel meer het resultaat is van een arena waarin verschillende actoren met uiteenlopende belangen 'strijden' om de 'juiste' beleidskeuzes waardoor een beleidsproces minder gestructureerd verloopt en informatie slechts één van de bronnen is die van invloed zijn op de richting van het beleid. Hoewel big data ongetwijfeld het beleidsproces weer beter kan voorzien van informatie betekent dat niet automatisch dat big data ook richtinggevend wordt. Andere factoren (macht, loyaliteit, historie, gewoonten, emotie, intuïtie...) zijn ook van invloed. De informatietechnische kijk op beleid – gevoed en versterkt door big data – zal blijvend "concurreren" met meer bestuurskundige perspectieven op beleid. Waarbij overigens de aanname blijft bestaan dat deze (en andere) perspectieven kunnen coëxisteren en elkaar kunnen verrijken.

Voor het effectief maken van big data in beleid zien we ons dus voor een aantal (nieuwe) uitdagingen gesteld, die variëren van meer bestuurlijk/organisatorische tot en met methodologische en praktische aspecten. Big data heeft een technische connotatie en dat maakt het lastig het belang van deze ontwikkeling in een politiek-bestuurlijke context te agenderen. Ten eerste moet dat belang voor een deel nog ontdekt en onderbouwd worden. Ten tweede is het nodig om dat belang in de taal van degenen die hiervan kunnen profiteren uit te drukken. Ondanks dat informatie- en communicatietechnologieën (ICT) (waar big data mee geassocieerd worden) grote voordelen voor de overheid en het beleidsonderzoek kunnen bieden, gaan implementatie en gebruik gepaard met tegenslagen en in de politiek met een steeds negatiever imago. Denk aan het parlementair onderzoek naar het mislukken van grote ICT-projecten. Een meer praktische kwestie is dat big data ook om nieuwe specialismen en tools vraagt die nog niet in de breedte voorhanden zijn. Feitelijk heeft deze ontwikkeling last van een meer breder probleem en dat is een tekort aan ICT specialisten. Het vraagt om investeringen in (her- en bij-) scholing en om nieuwe opleidingen (data analytics, ...). Big data leidt tot de ontwikkeling van een nieuwe

vakgebied en doet ook een beroep op nieuwe vaardigheden. Het nieuw vakgebied betreft "big data analytics" oftewel "the process of examining and interrogating big data assets to derive insights of value for decision-making" (Yu, 2012, p. 10). Specifiek is er behoefte aan vaardigheden op het vlak van datamanagement, data-analyse en bedrijf en beleid (de toepassingsgebieden) (zie Breur, 2013, p. 23 e.v.).

Eerder in dit artikel hebben we al gewezen op de nood om koppelingen te leggen tussen eigenaars, bewerkers en gebruikers van big data. In sommige gevallen liggen ze allemaal in één hand, maar dat zal heel vaak niet het geval zijn. Ter illustratie: een zoekmachine registreert gebruikte zoektermen, maar doet daar – als eigenaar – niets mee. Andere kunnen dat juist wel zeer waardevolle informatie vinden. Zij zijn potentieel gebruiker van data die zij niet bezitten (of bewerken). Om kansen te pakken moet er dus een zoektocht worden ondernomen zodat de meest waardevolle datasets gevonden worden en gekoppeld worden aan specifieke beleidsvraagstukken. Dat is geen sinecure, want eigenaars van datasets weten misschien helemaal niet dat hun data geschikt zijn voor dergelijke vraagstukken, terwijl beleidmakers niet altijd weten waar geschikte datasets aanwezig zijn en waar zij dus moeten zoeken. In dit perspectief kan een analogie met "garbage can model" worden gemaakt (March & Olson, 1976). Dit model leunt op de gedachte dat besluitvorming is opgebouwd uit drie stromen, namelijk problemen, oplossingen en participanten. Deze stromen zijn continu in beweging en aan verandering onderhevig. Voortgang in besluitvorming ontstaat wanneer deze drie stromen elkaar op eenzelfde moment "raken", meestal door inzet van een participant (Zie voor een meer uitgebreide bespreking van dit model 't Hart, Metselaar & Verbeek, 1995). Bij het effectueren van big data geschiedt iets vergelijkbaars, namelijk een koppeling van maatschappelijke uitdagingen ("problemen") aan big data ("oplossingen") door interactie tussen verschillende participanten.

Rondom big data kleeft een zweem van 'objectiviteit'. Het zijn immers apparaten (van sensoren tot en met tablets) die informatie verzamelen en technici die de data bewerken. Dat geeft geen garantie dat big data objectief zijn. Een klein voorbeeld: Sommige big data tools putten uit informatie die via social media wordt verspreid (Facebook, Twitter, ...). Men zou daarmee de stemming van gebruikers over een bepaald onderwerp kunnen 'meten' en volgen. Echter, net als bij enquêtes, is er sprake van zelfselectie en sociaalwenselijk gedrag. Ten eerste is niet iedereen aanwezig op sociale media, en ten tweede stelt men het eigen leven vaak als positiever, spannender en meer inspirerend voor dan het feitelijk is (zeker wanneer men weet dat men geobserveerd wordt, het zgn. Hawthorne effect). We zouden wat dat betreft nog veel kunnen leren of gebruiken van de sociologische studies over hoe de mens zichzelf (graag) presenteert in het publieke – en nu dus het digitale – domein (cf. Goffman, 1959). Dit doet overigens geen afbreuk aan de waarde van veel big data, zeker wanneer het niet gaat om verzameling en analyse van uitspraken, opvattingen en attitudes. Denk bijvoorbeeld aan verplaatsingsgedrag van individuen op basis van locatiegegevens die mobiele apparaten registreren of de inventarisatie van pagina's waar de meeste lezers van een bepaald e-book afhaken. Het gaat echter te ver om te stellen dat (big) data de belangrijkste bouwsteen is voor 'evidence-based' beleid.

Ook als de data 100% objectief is, zijn de algoritmen die gebruikt worden voor het analyseren van deze data samengesteld door mensen. Het is dan belangrijk in acht te nemen dat ook algoritmen (bijna) nooit 100% objectief en waarheidsgetrouw zijn. Als voorbeeld kunnen wij kijken naar Spotify en Bol.com. Spotify geeft de gebruiker muzieksuggesties, Bol.com laat zien welke boeken mogelijk ook interessant zijn. Op basis van de algoritmes van Spotify en Bol.com wordt ons lees- en luistergedrag beïnvloed. Echter, de algoritmen die Spotify en Bol.com gebruiken voor het analyseren van de big data sets zijn niet 100% objectief. Met andere woorden: op basis van hetzelfde luister-

(muziek) en leesgedrag (boeken) zal een concurrent hoogstwaarschijnlijk nét andere suggesties doen.

6 Tot slot

Big data is een relevante ontwikkeling die vanuit het beleid(sonderzoek) de volle aandacht geniet. Hoewel technologieën het werken met big data sets steeds eenvoudiger maakt, betekent dat geenszins dat vraag en aanbod elkaar weten te vinden. Laten we ons sturen door big data of laten we big data voor ons werken? Mogelijk doen de beste kansen zich voor op het snijvlak van organisaties en sectoren die thans nog weinig met elkaar te maken hebben, maar die in hun zoektocht naar een effectief gebruik uiteenlopende big data sets koppelen. Het is dus niet alleen een zoektocht naar de mogelijkheden van big data. Het is evenzeer een zoektocht naar de onverwachte combinaties en de maatschappelijke en economische behoeften waar big data potentieel een grote rol kan spelen.

Big data lijkt thans een hype, maar we hebben voldoende redenen aan te nemen dat big data een vaste plek zal verwerven in het beleid(sonderzoek). De 'explosie' van gegevens lijkt vooralsnog geen grenzen te hebben (en dat maakt elke afbakening van big data in termen van aantallen bytes per definitie achterhaald). De grote vraag hierbij is hoe we verstandig omgaan met big data en welke randvoorwaarden ingevuld moeten worden om big data voor ons te laten werken. De beleidsonderzoeker en de -maker zullen zich dan ook regelmatig de vraag moeten stellen of big data een rol kan spelen in hun werk. De beleidswetenschap kan verkennen welke positie big data in de overheid en in beleidsprocessen inneemt en welke impact dat heeft op onze gebruikelijke methoden om het verloop en de impact van beleid te analyseren en te verklaren. De beleidspraktijk vraagt om meer praktische kennis over big data, zoals de relatie met wetgeving, investeringen in vaardigheden en het articuleren van vragen die de zoektocht naar en inzet van big data sturen.

6.1 Over de auteurs

Dr. F.J. Bongers, J. Veldkamp MSc en ir. T. van der Vorst zijn werkzaam als consultant bij Dialogic innovatie & interactie te Utrecht. Dit artikel is een bewerking van een *working paper* dat op het Politicologenetmaal 2014 onder de titel 'Big data: een ontdekkingsreis voor bestuurders en onderzoekers' is gepresenteerd.

Literatuur

- Algemene Rekenkamer (2014), *Tendrapport Open Data*, Den Haag.
- Australian Government (2013), *The Australian Public Service Big Data Strategy Improved understanding through enhanced data-analytics capability*, Canberra.
- Breur, T. (2013), *Big Data. De nieuwe goedkoorts?*, Academic Service, Den Haag.
- Dialogic & The Green Land (2012), *Open data, open doel. Verkenning van de kansen van datasets*, Utrecht.
- Dunn, W. N. (1994), *Public Policy Analysis: An Introduction*, Prentice Halls, Englewood Cliffs.
- Goffman, E. (1959), *The Presentation of Self in Everyday Life*, Anchor Books.
- 't Hart, P., M. Metselaar & B. Verbeek (1995), *Publieke besluitvorming*, VUGA, Den Haag.
- Van Hoesel, P.H.M., F.L. Leeuw & J.W.M. Mevissen (2005), *Beleidsonderzoek in Nederland. Kennis voor beleid; ontwikkeling van een professie*, Van Gorcum, Assen.
- HM Government (2013), *Seizing the data opportunity. A strategy for UK data capability*, Londen.
- Hoogerwerf, A, & M. Herweijer (2002), *Overheidsbeleid. Een inleiding in de beleidswetenschap*, Samsom, Alphen a/d Rijn.
- IBM (2014), What is Big Data?, <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- John, P. (2005), *Analysing Public Policy*, Continuum, Londen.
- March, J.G. & J.P. Olsen (1976), *Ambiguity and choice in organizations*, Universitetsforlaget, Bergen.
- Mayer-Schönberger, V. & K. Cukier (2013), *De Big Data Revolutie. Hoe de data-explosie al onze vragen gaat beantwoorden*, Maven Publishing BV, Amsterdam.
- Webbink, H.D. (2006), "Causale effecten van beleid. Over gecontroleerde en natuurlijke experimenten", in: *Justitiële Verkenningen*, jrg. 31, nr. 8, blz. 72-84.
- Yu, C. (2012), *The Big Data Opportunity Making government faster, smarter and more personal*, Londen.



Contact:

Dialogic
Hooghiemstraplein 33-36
3514 AX Utrecht
Tel. +31 (0)30 215 05 80
Fax +31 (0)30 215 05 95
www.dialogic.nl

