



Delft University of Technology

## Big data, grote vragen; een institutionele onderzoeksagenda

de Bruijn, Hans; van der Voort, Haiko

**DOI**

[10.5553/Bk/092733872016025001012](https://doi.org/10.5553/Bk/092733872016025001012)

**Publication date**

2016

**Document Version**

Final published version

**Published in**

Bestuurskunde

**Citation (APA)**

de Bruijn, H., & van der Voort, H. (2016). Big data, grote vragen; een institutionele onderzoeksagenda. *Bestuurskunde*, 25(1). <https://doi.org/10.5553/Bk/092733872016025001012>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Big data, grote vragen; een institutionele onderzoeksagenda\*

*Hans de Bruijn & Haiko van der Voort*

Met big data zijn we op weg naar een ‘datarevolutie’ (Kitchin, 2014) – niet alleen omdat er gigantische hoeveelheden data worden gegenereerd, maar ook omdat we die bovendien beter kunnen verwerken. Er komen steeds nieuwe algoritmen beschikbaar om betekenis te geven aan de data. Data vertegenwoordigen een economische waarde, dus ontstaat er ook een handel in data. En er wordt gezocht naar nieuwe toepassingsgebieden.

Ook voor de publieke sector is er de belofte dat de nieuwe beschikbaarheid van data tot betere informatie leidt en dat die weer tot betere besluiten en dienstverlening leidt (Daalhuijsen, Steenman, & Meijer, 2015). In dit afsluitende artikel reflecteren we kort op een aantal aspecten van het gebruik van big data in het publieke domein. Daarmee proberen we wat evenwicht te brengen in de discussie over big data in de publieke sector. Zoals altijd met nieuwe ontwikkelingen worden de kansen en mogelijkheden van big data sterk benadrukt – en dat is goed. Maar er behoort ook aandacht te zijn voor een aantal vervolgvragen. Krijgen die geen aandacht, dan kunnen de mogelijkheden van big data niet ten volle worden benut. We staan stil bij onder meer de neutraliteit van big data, waarden als transparantie en privacy en de impact van big data op processen van beleidsontwikkeling.

Indien achter de big data revolutie de aanname schuilgaat dat betere data leiden tot betere informatie en tot betere besluitvorming, dan is deze natuurlijk aanvechtbaar. De aanname wordt ondersteund door Gaël Kermarrec en Marc Debuschere voor hun projecten. In een aantal bijdragen in dit nummer wordt deze aanname echter betwijfeld. Maar er is meer: regelmatig komt in dit themanummer de vraag op naar de relatie tussen big data en instituties. Daar staan we in deze reflectie ook bij stil en we sluiten af met een institutionele onderzoeksagenda.

## **De ‘wie-vraag’: feilbaarheid en politiek achter big data processen**

In twee bijdragen in dit themanummer wordt het proces van dataverwerving naar datagebruik geanalyseerd. Kwakkel en Verbraeck problematiseren de verschillen tussen mogelijkheid, bruikbaarheid en wenselijkheid van big data voor de besluitvormer. Niet alles wat mogelijk is, is tevens bruikbaar en/of wenselijk. Janssen en Van der Voort presenteren het proces als een keten van activiteiten en problematiseren de transfers tussen de schakels. Deze transfers zijn op verschillende manieren complex (semantiek, deskundigheid, motivatie) en daarmee vatbaar

\* Prof. dr. J.A. de Bruijn is hoogleraar aan de Technische Universiteit Delft. Dr. H.G. van der Voort is universitair docent aan de Technische Universiteit Delft.

voor ruis en misverstanden. Dergelijke complexiteiten spelen ook binnen organisaties. Bruno Broucker richt zich op de overheid en breekt een lans voor betere coördinatie op het niveau van dataprocessen (micro), op het niveau van de organisatie (meso) en het niveau van het beleidsdomein (macro).

Deze observaties zijn voor iedere data-analist van belang. Het zijn ‘wat’-vragen (wat is mogelijk, bruikbaar etc.), maar voor een belangrijk deel ook ‘wie’-vragen. Wie gaat er over mogelijkheid, bruikbaarheid en wenselijkheid? Vaak zijn dat verschillende spelers. Wie zijn de spelers die een schakel in een keten beheren en waarschijnlijk dus ook de data? Wie bepalen de algoritmes? Emma Uprichard (2015) somt nog een aantal van die ‘wie’-vragen op. Wie genereert data? Wie is eigenaar ervan? Wie is eigenaar van welk deel van de data? Wie heeft toegang tot data en kan ze bewerken? Wie doet dat? Naar wie gaan de analyses en bevindingen en om welke redenen? Wie profiteert er het meest van? En wie het minst?

Waarom doen die vragen ertoe? De wereld van big data wordt zo af en toe wel wat al te paradijselijk voorgesteld. Er zijn oneindig veel data. Dataverwerking, analyse, betekenisgeving en predictie – het verloopt allemaal ‘machinaal’, alsof menselijke interventies er niet meer toe doen. De werkelijkheid is een andere. Er is ruis en chaos. Complexe transfers beïnvloeden de betrouwbaarheid van data. Veel data worden niet gedeeld. Er is een soms problematische bruikbaarheid. Er zijn de impliciete, politieke keuzes van de data-analist. De wereld van big data is daarmee veel rommeliger dan we als buitenstaanders wellicht denken. Die rommelige wereld kunnen we scherper in beeld krijgen als we de wie-vraag vaker stellen. Ten eerste, omdat de wie-vraag iets zegt over de kwaliteit van de data en data-analyse. Wie heeft de eigendom van welke data? Wie beslist welke data met wie mogen worden gedeeld? Wie kan dus beslissen bepaalde data achter te houden? Wie is de analist en welke politieke opvattingen heeft hij of zij? Wie beslist over bruikbaarheid? Het antwoord op deze en dergelijke wie-vragen kan ons helpen een beter zicht te krijgen op de kwaliteit van data en data-analyses die we krijgen voorgeschoteld. Ten tweede, big data is goud, dus de vraag welke data waarvoor worden gebruikt, impliceert ook een machtsspel waarbij de belangen van de betrokken spelers een rol spelen. Inzicht in dit machtsspel kan helpen om tot een beter oordeel te komen over data-analyses.

### **Big data vergt nieuwe checks and balances – maar ‘who cares’?**

Als data goud zijn, dan zijn ze macht. Als ze macht zijn, dan ontstaan nieuwe machtsconcentraties. Binnen organisaties wordt hier vaak gewezen op de macht van data-analisten. Zij nemen al analyserend allerlei beslissingen, die politiek van aard zijn, maar vaak onzichtbaar voor de echelons boven hen. De vergelijking met de *quants* in de wereld van de banken dringt zich op: analisten die op operationeel niveau ingewikkelde producten maakten en onvoldoende *countervailing powers* hadden – met alle desastreuze gevolgen van dien. De essentie van hun werk was te onzichtbaar, te complex, ogenschijnlijk succesvol – en dus bleek het onmogelijk om tegenkrachten te organiseren.

Hans de Bruijn & Haiko van der Voort

Er wordt ook veelvuldig gewezen op organisaties als Google en Yahoo, die beschikken over een grote hoeveelheid data en een enorme verwerkingscapaciteit. Wie biedt hen tegenwicht? Individuen? Nationale overheden? Maatschappelijke organisaties? Nu hebben Google en Yahoo een hoge zichtbaarheid, maar daarnaast bestaan er veel en minder zichtbare internationaal opererende bedrijven die data vergaren, opslaan en verhandelen (zie het interview met Michel van Eeten). Het beeld is ook hier: onvoldoende *countervailing powers*.

Dit leidt tot een volgende, belangrijke vraag: *who cares?* Wie heeft er belang bij de kwaliteit van data? Wie investeert in het begrip van dataprocessen, opdat de vinger op de zere plek gelegd kan worden bij de aanbieders van data? Ronald Leenes problematiseert een gebrek aan transparantie. Deels gaat het hier om klassieke vragen over transparantie en representatie – zijn kwetsbare groepen voor wie besluiten worden genomen voldoende gerepresenteerd? – maar deels ook om nieuwe vragen. Zo wordt het klassieke denken over privacy volledig overhoop gegooid door de big data revolutie (zie de bijdrage van Hans de Bruijn en het interview met Michel van Eeten).

### **Causaliteit wordt minder belangrijk, behalve bij high-impact besluiten**

Data vragen om een betekenisgeving. Het is daarbij een misverstand dat betekenisgeving altijd uit de data volgt. Die kan ook worden ingegeven door de waarden, belangen of morele opvattingen van degene die betekenis geeft aan data.

Bij big data is dat niet anders. Eén van de beloftes van big data is dat er minder noodzaak is om op zoek te gaan naar causale relaties tussen verschijnselen. In de zee aan data kan de analist op zoek gaan naar patronen en correlaties die zich telkens weer manifesteren. ‘Big data helps answer what, not why, and often that’s good enough’ (Cukier & Mayer-Schoenberger, 2013). Een eenvoudig voorbeeld: wanneer zich een orkaan aandient, wordt in de dagen daarvoor heel veel taart gekocht in de supermarkten. Bij ‘normaal’ onderzoek zouden er, als het gaat om de betekenisgeving hiervan, twee vragen zijn:

- Waarom neemt de verkoop van taarten toe? Met andere woorden, is er een causale relatie tussen orkanen en taarten?
- Wat betekent die toename van vraag naar taarten voor de besluitvormers bij de supermarkten?

Die beide vragen naar betekenisgeving zijn, als we de data-analist hier moeten geloven, niet zo interessant, omdat de betekenisgeving rechtstreeks uit de data volgt. Wat je als supermarkt moet doen als je dit weet? Meer taarten inslaan als er een orkaan aankomt. Waarom leiden orkanen tot meer inkoop van taarten? Big data vertellen ons dat dat zo is – en dat is genoeg. Wie in de grote hoeveelheid data telkens weer een correlatie aantreft, hoeft niet meer op zoek naar een antwoord op de waarom-vraag.

Het inkopen van taarten is echter geen *high-impact* besluit. Mocht de correlatie waarop werd vertrouwd, zich een volgende keer niet voordoen, dan is de schade daarvan beperkt. Maar wat als het besluit dat op grond van de big data analyse wordt genomen, wel een hoge impact heeft – als er bijvoorbeeld veel kosten mee gepaard gaan, of het besluit verstrekkende gevolgen heeft voor burgers?

Een voorbeeld is hier de verspreiding van een gevaarlijke griep. Big data analisten kunnen de verspreiding van die griep wellicht voorspellen – denk aan *Google Flu Trends*, dat een correlatie vond tussen zoekgedrag op het internet en de verspreiding van een griep. Een minister van volksgezondheid kan voor de keuze staan of een land massaal dure vaccins inkoopt. Dat is een high impact besluit, en dan wordt de vraag naar de betekenisgeving toch wat prangender. Zijn de uitspraken van Google gebaseerd op oorzaak-gevolganalyses of niet? En als dat niet zo is, hoe zeker zijn we er dan van dat de analisten het bij het rechte eind hebben? Eerst voorspelde Google het verloop van een griep goed, daarna zat Google er helemaal naast (Lazer, Kennedy, King, & Vespignani, 2014). Het laat het bestuurlijke risico van dit soort analyses zien. Laten op correlaties gestoelde beslissingen zich goed verantwoorden?

### **De politiek van betekenisgeving: een nieuwe positie van data-analisten**

Een analist maakt chocola van data met behulp van een algoritme. Goede data-analisten zijn schaars. Cukier (2010) geeft aan dat de mogelijkheden van big data niet volledig zullen worden benut bij een gebrek aan menselijk talent. Hij schetst een *data scientist* die de vaardigheden van een programmeur, een statisticus en een *storyteller* combineert. *Data scientists* moeten, willen zij hun maatschappelijke rol kunnen waarmaken, dus ook zijn ingevoerd in communicatie en de politieke kant van besluitvorming (zie ook Davenport, Barth, & Bean, 2012). Die brede opvatting van de rol van een *data scientist* zien we vaker. Kitchin (2014, p. 162) geeft een – hilarisch – lange tabel met gewenste kwaliteiten, waaronder domeinkennis, datakennis, analytische vaardigheden en projectmanagement-vaardigheden, waarbij iedere categorie weer is onderverdeeld in minimaal vijf vaardigheden. Een dergelijk overzicht van gewenste kwaliteiten geeft vooral weer waar de schoen knelt: blijkbaar is de brug tussen analyse en besluitvorming problematisch. Big data bereikt de besluitvorming niet of heeft onvoldoende impact op de besluitvorming – of, anders geformuleerd, wordt door besluitvormers onvoldoende benut. Dat is niet nieuw, maar wel meer problematisch als big data aan belang toeneemt.

Die problematische relatie tussen analyse en besluitvorming heeft nog een andere component. Voor veel besluitvormers zijn big data processen moeilijk te bevatten. Hoe data zijn ontstaan en in een eerder stadium zijn verwerkt, is voor hen vaak onbekend. Het kost vaak te veel tijd en geld om de kwaliteit van dat proces te traceren. Dit is gevaarlijk, omdat aan de algoritmen allerlei politieke keuzes ten grondslag liggen, die de ontwerper van het algoritme maakt. Achter de uitkomst van de analyse kan dus een politieke, morele of persoonlijke keuze schuilgaan, die

we niet zien. Analisten construeren met behulp van hun algoritmen een werkelijkheid – en dat is voor besluitvormers vaak niet zichtbaar. Deze problematiek van de politiek van algoritmen is des te belangrijker omdat hier een venijnig paradoxje speelt. Hoe meer data, hoe meer correlaties, hoe meer we onze besluiten op feiten kunnen baseren, zo lijkt het. Edoch, er is een risico: hoe meer data en correlaties, hoe makkelijker het ook kan worden voor een analist om zijn eigen werkelijkheid te construeren, zeker als de analist is ontslagen van de verplichting om causale relaties aan te tonen. Sterker, er zijn zo ontstellend veel data, dat er voor iedere theorie waarschijnlijk wel correlaties zijn te vinden, die die theorie ondersteunen. Het is dus van belang dat de keuzes die achter de algoritmen schuilgaan, worden geëxpliciteerd.

### **De politiek van big data gedreven besluitvorming: een omkering van het beleidsproces?**

Nu doen bovenstaande problemen zich natuurlijk niet altijd voor. Ten eerste kunnen er veelal wel causale verbanden worden gevonden. Ten tweede worden data veelal gebruikt om meer eenduidige publieke dienstverlening op operationeel niveau te verbeteren, zoals in de gezondheidszorg en toezicht. Een interessante vraag is dan ook of het grotere belang van data-analisten de verhouding tussen analyse en beleid gaat veranderen.

In veel gevallen zal de volgorde der dingen niet zijn dat er een vraag van boven komt, uit de top van de organisatie, die vervolgens door big data analisten onderin wordt beantwoord. Het is niet: er is een probleem, big data leveren een oplossing. Maar het is veelal andersom: big data analisten komen met een innovatie of oplossing, waar vervolgens een probleem bij wordt gezocht. De innovaties en probleemoplossing komen in dergelijke gevallen van onderuit.

Op terreinen waar big data ertoe doet, kan zo een radicale decentralisatie van beleidsontwikkeling optreden. Wie bijvoorbeeld een interventie wil plegen op een beleidsterrein, doorloopt niet de sequentie van doelen stellen, informatie verzamelen en vervolgens de interventie plegen. Het is veeleer andersom: uit de vele informatie kunnen interventies worden gemaakt, die vervolgens leiden tot het vaststellen van doelen. Doelen stellen – traditioneel het terrein van bestuurders en managers – wordt dus meer een *ex post* activiteit. Dit suggereert een voorname positie van data-analisten, die immers *de facto* een grotere invloed krijgen op doelbepaling. Soms zullen we dat toejuichen, soms niet (zie de waarschuwingen hierboven). Dat roept een tweeledige sturingsvraag op. Hoe sturing van big data zo vorm te geven, dat er maximaal gebruikgemaakt kan worden van de kansen die big data biedt? En tegelijk: hoe sturing zodanig vorm te geven, dat er niet een te grote afhankelijkheid bestaat van analisten, die immers onderin belangrijke politieke keuzes maken? Het antwoord op beide vragen lijkt het tolereren van variëteit te zijn – een variëteit aan analisten, methoden, thema's. Variëteit biedt ruimte voor ondernemende analisten en kan tegelijk zorg dragen voor *checks and balances* en voorkomt afhankelijkheid van slechts enkele analisten.

## Een institutionele onderzoeksagenda

Het zal duidelijk zijn – big data kan een enorme impact hebben, maar er lijkt sprake te zijn van een *institutional void* rond big data. Wat zou een institutionele onderzoeksagenda kunnen zijn? Er zijn in het themanummer al legio belangrijke vragen genoemd. Wij zien de volgende vragen:

Allereerst lijkt de ‘wie’-vraag nog niet gezaghebbend beantwoord, ondanks de diverse tools voor actor-, keten- en netwerkanalyses die we tot onze beschikking hebben. Een eerste investering geldt de ontwikkeling van een institutionele kaart rond big data. Welke spelers hebben welk rol bij dataverzameling, -analyse en -toepassing? Welke posities nemen zij in en welke belangen hebben ze? Wat zijn de belangrijkste prikkels die hun gedrag bepalen?

Met behulp van de nieuwe kaartkennis kunnen we concreter belangrijke sturingsvragen beantwoorden. Onder welke institutionele connecties kunnen de mogelijkheden van big data maximaal worden benut? Wat betekent variëteit hier en kan variëteit bijdragen aan een maximale benutting? Zijn er nieuwe sturings- en managementconcepten noodzakelijk?

Binnen publieke organisaties geldt de vraag welke nieuwe *checks and balances* binnen publieke organisaties moeten worden ontwikkeld, gegeven de mogelijkheid dat big data tot omkering van beleidsprocessen kan leiden. Wordt het ‘de analist leidt, de besluitvormer volgt’? En hoe wordt de kwaliteit van bestuurlijke verantwoording in een dergelijke situatie geborgd?

Met de verandering van methoden van (beleids)onderzoek zal ook de genoemde kloof tussen analyse en besluitvorming veranderen. Er is een grote behoefte aan data-analisten met bestuurlijke intelligentie, en bestuurders die de processen achter big data kunnen doorgronden. Wie springt er in het gat? En hoe kunnen opleiding en advisering inspelen op de nieuwe behoeften?

Analisten kunnen, onder de politieke radar, politieke besluiten nemen. Dat vraagt niet alleen om *checks and balances* – maar wellicht ook om professionalisering van de beroepsgroep. Als data-analist een professie is, als we als samenleving meer en meer moeten vertrouwen op data-analisten, als ze besluiten nemen die veel impact kunnen hebben – dan dringt de vergelijking zich op met professies als artsen en advocaten. En dan is er ook hier een vraag naar institutionalisering: zou deze beroepsgroep niet met professionele codes moeten komen, zoals andere beroepsgroepen die ook hebben? Hoe kunnen condities worden gecreëerd waarbinnen een dergelijke institutionalisering kan plaatsvinden?

Daarnaast is er natuurlijke de grote vraag hoe er *checks and balances* kunnen ontstaan tussen de bedrijven die data in eigendom hebben en gebruiken enerzijds en burgers en overheden anderzijds – in een wereld die zich van jurisdicties niets aantrekt. Dat vergt nieuwe instituties, want hier is echt sprake van een *institutional void*. Voor veel arrangementen die we kunnen ontwerpen om deze leegte op te vullen, geldt dat moeilijk te voorspellen is welk effect ze zullen hebben. Het zal



Hans de Bruijn & Haiko van der Voort

veelal *trial and error* zijn. Bovendien, die arrangementen worden door een speler ontworpen (bijvoorbeeld Europa) en het is de vraag welke impact ze zullen hebben op spelers buiten Europa.

De institutionele leegte is hier zo groot, dat wellicht minder interessant is welke instituties die leegte kunnen vullen, en meer interessant hoe we het proces van *trial and error* zodanig vorm kunnen geven, dat er maximale kansen zijn op succes. Hier kunnen we weer terug naar een nationaal of lokaal niveau. Hoe handelen betrokkenen – bestuurders, data-analisten, publiek – in de wetenschap dat instituties er niet zijn of niet geloofwaardig zijn? En om met een optimistischer vraag te eindigen: welk lokaal tegenwicht kan er ontstaan tegen een mondiale institutionele leegte?

## Literatuur

- Cukier, K.N. (2010). Data, data everywhere. *The Economist*. <https://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>
- Cukier, K.N., & Mayer-Schoenberger, V. (2013). The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*. <https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data>
- Daalhuijsen, T., Steenman, S., & Meijer, A. (2015). Big Data: een revolutie in gemeentelijk beleid? *Bestuurswetenschappen*, 69, 6-24.
- Davenport, T.H., Barth, P., & Bean, R. (2012). How 'big data' is different. *MIT Sloan Management Review*. <http://sloanreview.mit.edu/article/how-big-data-is-different/>
- Kitchin, R. (2014). *The data revolution; Big data, open data, Data Infrastructures & Their Consequences*. London: Sage.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343, 1203-1205. <http://j.mp/1vWuOVp>
- Uprichard, E. (2015). Most big data is social data – The analytics need serious interrogation. <http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-data-science-emma-uprichard/>