

ARTIKELEN

THEMA-ARTIKELEN

Big data: een zoektocht naar instituties*

Haiko van der Voort & Joep Crompvoets

Het begrip ‘big data’ heeft een zekere magie. Het herbergt een belofte van de grote hoeveelheid gegevens die we hebben, waar we bij wijze van spreken op zitten, en die we slechts hoeven te benutten om verbeteringen aan te brengen in diverse aspecten van ons leven. Denk hierbij aan effectievere fraudebestrijding, goedkopere zorg, verbeterde veiligheid, en het vergroten van duurzaamheid (zie bijvoorbeeld in Dijkman & Hiemstra, 2014). De beloften van big data duiken vaak op in kranten of magazines. De beloften kunnen commerciële uitingen of een cafébabbel behoorlijk versterken. In dezelfde magazines en cafés klinken ook waarschuwingen over de gevaren van big data, veelal over privacy. Iedereen spreekt erover, en zo ontstaat het beeld van de hype: weinigen weten goed wat met het fenomeen te doen, velen denken dat de anderen het doen en beweren het dan ook maar zelf te doen, maar slechts enkele uitverkorenen hebben er echt (en positieve) ervaringen mee.

In dit themanummer over big data willen we voorbij de hype gaan – en willen we op een toegankelijke manier een overzicht geven van wat er speelt en telkens stilstaan bij de vraag of er iets nieuws onder de zon is. We kijken kritisch naar big data door en voor onze samenleving, en specifiek de overheid. Twee kritische vragen zullen telkens klinken: Is er meer dan een hype, ofwel wat is de kern van de beloften rond big data? En getuigen de beloften niet van een al te technologisch-rationeel wereldbeeld? Met andere woorden: hoe manifesteren instituties zich in de activiteiten die worden ontplooid onder de noemer ‘big data’? Deze vragen komen terug in een groot aantal korte bijdragen van wetenschappers en ervaringsdeskundigen. Bij elkaar geven ze een intellectuele *state of the art*: een overzicht van beargumenteerde beloften, kritische reflecties en actuele initiatieven. In deze inleiding gaan we nader in op een aantal essentiële begrippen rond big data en op de relevantie van de vraag naar instituties.

* Dr. H.G. van der Voort is universitair docent aan de Technische Universiteit Delft. Prof. dr. ir. J. Crompvoets is senior researcher/consultant/lecturer aan de KU Leuven. De auteurs bedanken Martijn van Otterlo voor zijn pioniersrol bij dit themanummer.

Haiko van der Voort & Joep Crompvoets

‘Big data’ en aanverwante begrippen

Wat is big data nu eigenlijk precies? Velen, waaronder wijzelf, worstelen met de term. De definitie van big data is niet altijd duidelijk en de term big data wordt vaak ten onrechte gebruikt.

De ‘big’ van ‘big data’ refereert aan de hoeveelheid data die te groot is om met reguliere systemen – laat staan met de hand – verwerkt te worden. Maar met de steeds beter wordende systemen wordt deze definitie van big data steeds schraler. Voor de beschrijving van big data worden vaak de v’s gebruikt (*Volume, Velocity, Variety*) (bijvoorbeeld Kitchin, 2014, p. 68). Er is een veelheid (*volume*) aan diverse data (*variety*) die bovendien *real-time* (*velocity*) kunnen worden verwerkt tot bruikbare informatie. In de context van dit themanummer willen we dit gebruiken voor een werkdefinitie: het gaat om de verwerking van een zeer grote hoeveelheid data die worden gebruikt voor besluitvorming en dienstverlening. In dit themanummer focussen we op besluitvorming en dienstverlening door de overheid.

Er zijn allerlei andere begrippen die vaak in één adem genoemd worden met big data. *Datamining* bijvoorbeeld is het gericht zoeken naar (statistische) verbanden in gegevensverzamelingen met als doel profielen op te stellen voor bijvoorbeeld wetenschappelijk, journalistiek of commercieel gebruik. Hierbij vormen het onderhoud en het zoeken een uitdaging en speelt het analyseren van de data een steeds grotere rol. Deze data bevatten immers een potentieel van een schat aan informatie voor veelvoudige doeleinden. Hiervoor is een aantal *big data analytics* tools voor ontwikkeld. Door het analyseren van big data kunnen verborgen patronen ontrafeld, onbekende correlaties, trends, klantgedrag en andere informatie afgeleid worden.

De basis van *datamining* en *big data analytics* zijn de ‘algoritmen’. De datavolumes zijn zo groot, dat deze niet meer ‘met de hand’ kunnen worden verwerkt. Dit gebeurt door computers. Alle geautomatiseerde handelingen – analyse van data, clusteren van data, profielen maken op basis van data, et cetera – gebeuren op basis van algoritmen. Algoritmen bevatten een eindige reeks instructies voor de analyse van data. Hier begeven we ons op het welhaast exclusieve terrein van mathematici en data-analisten. Hoe slimmer zij de algoritmen maken, hoe beter de uitkomsten.

Open data gaat over het ter beschikking stellen van data aan eenieder die er gebruik van wil maken. *Linked data* betreft een methode voor het ontsluiten van datasets ten opzichte van elkaar, zodat ze linkbaar en daarmee bruikbaar worden.

Deze opsomming is uiteraard niet uitputtend. De wereld van big data heeft in een korte tijd een bijzonder rijke, vooral Engelstalige, vocabulaire ontwikkeld.

Big data, grote beloften

Tot zover klinkt het niet zo revolutionair. De vraag is waarom de veelheid, de verbeterde verwerkingscapaciteit en het vermogen tot *real-time* verwerking van data leiden tot een revolutie. Door de spectaculaire toename van de verwerkingscapaciteit van data ontstaan er veel toepassingen die eerst niet mogelijk waren (Meijer & Van Berlo, 2014, p. 131), maar de ontwikkelingen zouden wellicht eerder een incrementele verandering aankondigen. Toch worden er forse claims gepubliceerd. Zo zou big data *the end of theory* inluiden, omdat met behulp van statistische algoritmen patronen beter kunnen worden gevonden dan met onze wetenschappelijke theorieën (Graham, 2012). Uit deze inmiddels beroemde quote spreekt distantie van een oude, kennelijk obsoleete wetenschap. Een nieuwe rationaliteit wordt aangekondigd. Deze nieuwe rationaliteit laat zich als volgt kenmerken:

Correlationisme

Feitelijk wordt een deel van het operationele onderzoekswerk geautomatiseerd. Slimme algoritmen bieden meer mogelijkheden om snel heel veel verbanden tussen variabelen te leggen en hier ook betekenis aan te geven. Het menselijke onderzoekswerk verschuift deels van dataverwerving en -analyse naar het ontwikkelen van algoritmen die computers in staat stellen dit te doen. De correlaties die de computer aangeeft, worden vervolgens door mensen geïnterpreteerd. Het leidt tot een andere benadering van onderzoek, die door Bowker (2013) wel ‘correlationisme’ wordt genoemd. ‘You don’t need to know whether someone is male or female, queer or straight, you just need to know his or her patterns of purchases and find similar clusters’ (Bowker, 2013). Met andere woorden: er wordt niet meer gestreefd naar het begrijpen van een kenobject. Causaliteit wordt minder nagestreefd, maar men neemt genoeg met sterke correlaties. Andrejewic (2013, p. 26) beschrijft deze ambitie als ‘a state of knowing without understanding’.

Nieuw functionalisme

Er zijn enorm veel data en de computer genereert enorm veel correlaties. De correlaties hoeven echter niet per se te duiden op causaliteit. Dit is niet voor iedereen problematisch. Een voorbeeld is de ontwikkeling van formules die sociale invloed kunnen benaderen. Grote bedrijven als Google en Yahoo hebben elk hun eigen formules, met hierin bijvoorbeeld het aantal links met andere personen, het aantal reacties op *posts* van personen, et cetera. Het identificeren van sociale invloed heeft uiteraard grote commerciële waarde. Potentieel effectieve formules worden onmiddellijk gepatenteerd. Er is bovendien een prikkel om steeds ambitieuzere formules te ontwikkelen. Zo ontstaat het beeld van functionalisme: ‘Werk het?’ lijkt het belangrijkste criterium, of wellicht een commerciële variant hiervan, zoals ‘kunnen we het verkopen?’ (Polonski, 2015). Ondertussen is er een hoge tolerantie voor imperfectie. Vooral in een commerciële setting is bruikbaarheid al snel belangrijker dan onderbouwing.

Haiko van der Voort & Joep Crompvoets

Nieuw positivisme

We hebben veel nieuwe toepassingen aan dit functionalisme te danken. Successen kunnen er echter toe leiden dat het bescheiden ideaal van bruikbaarheid verschuift naar maakbaarheid. Big data is altijd omgeven geweest door een air van *evidence based* toepassingen. Meijer en Van Berlo (2014, p. 142) schetsen dat de big data beweging werkt vanuit rationalistisch ideaal van techno-rationaliteit, kwantificatie en voorspelbaarheid, ofwel een nieuw positivisme. De nieuwe mogelijkheden van big data zouden nu eindelijk ertoe kunnen leiden dat we kunnen vaststellen wat is.

Een zoektocht naar instituties

Ondertussen is er nog steeds een verschil tussen de manier waarop we de wereld ervaren en de wereld die middels ICT kan worden gekend (Posner, 2015). Nieuw positivisten wijten dit misschien aan kinderziektes. Maar zijn zij wellicht op zoek naar de heilige graal? De nieuwe rationaliteit lijkt voor veel bestuurskundigen een terughal naar vroege tijden. Zij zullen zich vragen stellen over de instituties rond big data, zoals bijvoorbeeld:

Waarden en machtsverhoudingen

Waar zijn de waarden en machtsverhoudingen die, naast kennis en informatie, essentieel zijn voor het maken van beleid (Meijer & Van Berlo, 2014, p. 143)? Besluitvormers zijn geen passieve actoren aan het einde van een keten. Zij zijn juist actief in het zoeken en articuleren van vragen, vaak in interactie met andere actoren die wellicht concurrerende waarden aanhangen. Articulatie van vragen is daarmee een dynamisch proces.

Waar is de verantwoording?

De complexiteit van algoritmes bemoeilijkt inzicht in en controle op de werking en uitkomsten van big data analyses. De automatische algoritmische handel op de beurs zorgde in 2010 al voor een snelle en onvoorziene beurscrash. Overheden die gebruikmaken van voorspellende software, horen aan burgers te kunnen uitleggen waarom het systeem hen selecteert voor een nadere inspectie. De complexiteit van algoritmes, die soms werken met duizenden variabelen, maakt dit bijna onmogelijk. Maar transparantie kan de effectiviteit van het systeem ook ondermijnen. Als duidelijk is waar een fraudeopsporend systeem op selecteert, zullen burgers daar hun gedrag op aanpassen.

Waar is de overheid?

De overheid is in discussies over big data in verschillende gedaanten betrokken. De overheid genereert data en is als zodanig leverancier. De overheid is ook gebruiker voor bijvoorbeeld toezicht. Daar waar het gebruik van big data schuurt aan de juridische grenzen, is de overheid toezichthouder en beleidsmaker. Vanuit al deze rollen zal de overheid de mogelijkheden en beperkingen, alsmede de gevaren en onwenselijkheden – deels – nog moeten leren kennen. Een overheid die

inzet op vernieuwing van big data, kan dat alleen maar legitiem doen als de elementaire rechten van burgers niet worden geschonden. Dit is niet eenvoudig. Er zijn veel vragen over bijvoorbeeld privacy en democratie. Deze vragen worden in dit themanummer verder geëxploreerd.

Dit themanummer is een zoektocht naar instituties. Hoe manifesteren instituties zich in de nieuwe activiteiten die worden ontplooid onder de noemer 'big data'? Verschillende deskundigen uit Nederland en Vlaanderen hebben hier vanuit hun eigen perspectief hun licht op geworpen. Jan Kwakkel en Alexander Verbraeck identificeren een procesketen van big data voorafgaand aan datagebruik. Zij zien allerlei logisch-technische complexiteiten die veel (overheids)gebruikers niet zien, maar die er wel zijn. Marijn Janssen en Haiko van der Voort beschrijven ook een procesketen van dataverwerking tot datagebruik. Zij richten zich echter meer op sociale complexiteiten die coördinatie bemoeilijken. Bruno Broucker vraagt zich af hoe overheden zich moeten organiseren om van big data te kunnen profiteren. Hiervoor bestaan er vele randvoorwaardelijke vragen, die op micro-, meso- en macroniveau kunnen worden beschouwd. Martijn van Otterlo en Frans Feldberg werken een voorbeeld op macro-niveau uit voor een case in Alkmaar. Zij beschrijven een 'triple helix' samenwerkingsverband tussen overheid, wetenschap en bedrijfsleven. Zij gaan in op zowel de beloften als de randvoorwaarden voor deze samenwerking rond big data. Ronald Leenes beschrijft hoe de belofte dat big data leidt tot meer voorspellend vermogen voor de overheid, kan schuren met waarden van transparantie en verantwoording. Hans de Bruijn gaat in op een andere waarde, namelijk privacy. Hoe wordt de privacy bedreigd door big data? Welke kenmerken van big data leiden tot een dergelijke schending en tot welke governance-vraagstukken leidt dit? Ten slotte vraagt Dhoya Sniijders zich in een filosofische bijdrage af hoe big data de verhouding tussen mens en machine verandert. Wie domesticeert wie: de mens de data of de data de mens? De conceptuele bijdragen worden aangevuld met drie interviews met beleidsmakers en onderzoekers voor wie big data dagelijkse praktijk is en die als zodanig 'met de voeten in de modder' staan.

Dit themanummer is geen afrekening. 'Nieuw functionalisme', 'nieuw positivisme' en 'de institutiële redenering' lijken voor bestuurskundigen wellicht diskwalificaties. Dit hoeven ze geenszins te zijn. Big data leidt al tot een nieuwe manier van kennisverwerking en verandert al de manier waarop we tegen mensen en hun interactie aankijken, conform het adagium van Latour: 'Change the instruments, and you will change the entire social theory that goes with them' (2009, p. 9; geciteerd door Boyd & Crawford, 2011). De zoektocht naar instituties levert een fascinerende multidisciplinaire onderzoeksagenda op, die verder gaat, maar ook voortbouwt op de rationalistische ideeën vanuit de big data beweging. Deze presenteren we in de slotparagraaf.

Literatuur

- Andrejewic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. New York: Routledge.
- Bowker, G.C. (2013). The theory/data thing. *International Journal of Communication*, 8, 1795-1799.
- boyd, d., & Crawford, K. (2011). *Six provocations for big data*. Paper gepresenteerd aan het Oxford Internet Institute 'A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society', 21 September 2011.
- Dijksman, M., & Hiemstra, J. (2014). Overheid moet werk maken van Big data. *Gemeente.nu*, 16 april 2014.
- Graham, M. (2012). Big data and the end of theory? *The Guardian*, 9 maart 2012. <http://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. London: Sage.
- Latour, B. (2009). Tarde's Idea of quantification. In M. Candea (Ed.), *The social after Gabriel Tarde: Debates and assessments* (pp. 145-162). London: Routledge.
- Meijer, A., & Berlo, D. van. (2014). *De nieuwe overheid*. Den Haag: Boom Lemma uitgevers.
- Polonski, V. (2015). *Hacking the system of social influence: How can we use the mechanics of influence to drive behaviour for public good?* Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/08/07/hacking-the-system-of-social-influence/>
- Posner, M. (2015). *The radical potential of the digital humanities: The most challenging of computing is the interrogation of power*. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2015/08/12/the-radical-unrealized-potential-of-digital-humanities/>