

# Small kids, big data

## TOEPASSING ANALYTICS IN DE KINDEROPVANG IN VLAANDEREN

Frank Bongers, Koen De Bruyn en Dries Verlet<sup>1</sup>

**SAMENVATTING** Datamining is het gericht zoeken naar statistische verbanden in (grote) gegevensverzamelingen of door het combineren van gegevensbronnen. Het kan gesitueerd worden in het bredere veld van big data. Deze bijdrage illustreert het gebruik van big data ter ondersteuning van de effectiviteit en efficiëntie van inspecties in de kinderopvang. Daarnaast staan we ook stil bij de (potentiële) impact van big data op beleid en onderzoek. De impact van deze big data is niet alleen een kwestie van informatiekunde en economie. Er zijn ook belangrijke beleidswetenschappelijke uitdagingen. De beschikbaarheid van grote hoeveelheden gekoppelde data is immers nog geen garantie dat het beleid maatschappelijke uitdagingen aanpakt of dat onderzoekers hiermee in staat zijn om het effect van beleid te meten. Daarenboven is er het respect voor de privacy. Dit artikel staat stil bij deze uitdagingen.

### Inleiding

Huidige toepassingen van big data zijn heel divers en zijn te vinden in wetenschappelijk onderzoek, marketing, maar evengoed in overheidsdienstverlening, zoals criminaliteitsbestrijding, preventie van suïcide en sociaal beleid (MEIJER & VAN BERLO 2013). Het is dan ook niet verwonderlijk dat steeds meer overheden en bedrijven big data beschouwen als een te ontginnen goudmijn, die economische groei, productiviteit en innovatie kan bevorderen. Zo maken overheden gewag van opportuniteiten wat betreft publieke dienstverlening, beleidsontwikkeling, statistiek, bedrijfsvoering, economische ontwikkeling, arbeidsmarkt en productiviteit (AUSTRALIAN GOVERNMENT 2013; HM GOVERNMENT 2013).

Naar aanleiding van de actualiteit ontstond bij Kind en Gezin de onderzoeksvraag om te voorspellen of een dossierbehandelaar een dossier gericht moet opvolgen, aan de hand van de

gegevens die zich in een dossier bevinden (een zogenaamd voorspellingsmodel). Initieel werd voor de realisatie van dit voorspellingsmodel een private partner ingeschakeld, maar na grondige wetswijzigingen rond de kinderopvang was dit model niet langer bruikbaar en toepasbaar en werd besloten om het project intern aan te pakken. Maar vooraleer de case te bespreken, staan we even stil bij het concept big data op zich en de verschillende rollen die een overheid hier kan spelen.

### Big data: nieuwe wijn in nieuwe zakken?

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-SCHONBERGER & CUKIER 2013: 10). Big data hebben drie cruciale eigenschappen: het gaat om een zeer groot volume aan data, er is sprake van een zeer grote

*snellheid* waarmee deze data verzameld en getransporteerd worden, en de data zijn ongestructureerd en *gevarieerd*. In het Engels duidt men op de 3 V's: "volume, velocity and variety" (McAFEE & BRYNJOLFSSON 2012).

Het geeft aan dat big data gaan over data van een specifiek kaliber: ze zijn te groot, te vluchtig en te complex om met traditionele hands-on databasemanagementtools te bewerken. Dit maakt van de big data meteen ook een relatief gegeven. Onder impuls van technische ontwikkelingen verbeteren de capaciteiten voor opslag, transport en bewerking immers continu. Het is echter onjuist en onterecht om big data dan maar af te doen als oude wijn in nieuwe zakken. 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 werden 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 (BONGERS *et al.* 2015; PENTLAND 2014).

Termen als data science en datamining worden vaak in één adem met big data genoemd. Met data science verwijst men doorgaans naar de discipline waarin het gebruik van big data verder wordt uitgebouwd. Wanneer men het heeft over datamining, ligt de focus eerder op het zoeken naar algoritmen om voorspellings- en clas-

sificatiemodellen te maken (VAN DER VOORT & CROMPVOETS 2016).

## De rol van overheden in het bigdata-verhaal

De voordelen en opportuniteiten die big data ons bieden, worden echter niet vanzelf gerealiseerd. Het technologische is één iets, maar het inzetten van big data in het realiseren van maatschappelijke en economische meerwaarde, is nog een ander paar mouwen. Daar is mensenwerk voor nodig. Het is hier dat de overheid ook haar rol kan spelen.

Ten eerste is de overheid verzamelaar, eigenaar en gebruiker van grote databestanden. In de dienstverlening genereert een overheid heel veel data en verwerkt men grote hoeveelheden gegevens. Denk bijvoorbeeld aan het heffen van belastingen, sociale zekerheid, mobiliteit en onderwijs. De dienstverlening op deze terreinen gaat stuk voor stuk gepaard met grootschalige gegevensverwerking en -koppeling. Ten tweede kan de overheid – net als andere organisaties – big data gebruiken om de werking en impact van de eigen organisatie en het eigen beleid te verbeteren. Denk aan het registreren van tevredenheid over dienstverlening (uitingen in sociale media, terugkerend verkeer op websites ...). Ten derde kan de overheid het maatschappelijke en economische belang van big data onder de aandacht brengen bij bedrijven, kennisinstellingen en burgers (DIALOGIC 2014). De overheid doet dan zelfs niks met big data, maar levert wel een bijdrage aan randvoorwaarden voor het gebruik van big data door derden (BONGERS *et al.* 2015).

De overheid kan dus verschillende rollen vervullen in deze domeinen. Zo is er de rol van *het stellen van kaders en het controleren*. Zo kan de overheid optreden als voorbereider, uitvoerder en handhaver van wetgeving en regels stellen ten aanzien van het gebruik van big data. Dit kan alleen als de wetgever (de volksvertegenwoordiging) duidelijke regels stelt. Nagegaan kan worden of bestaande procedures op het gebied van databescherming en privacy bij verwerking van data aangepast moeten worden voor big data. Big data bestaan immers ook uit

het hergebruiken van informatie die voor andere doeleinden is verzameld. Daarnaast is er de rol van *aanjagen*. De overheid kan onderzoek en ontwikkeling in het domein van big data stimuleren of een actieve rol spelen in standaardisatieprocessen voor big data. Zo is er ook de rol als *producent, verzamelaar en verstrekker*. De overheid kan de datasets intern delen en – als onderdeel van open data – beschikbaar stellen aan externen. Daarnaast is er de rol van *gebruiker van big data*. Overheden gebruiken big data die derden verzamelen en beschikbaar stellen.

De verzameling van talloze gegevens over ons gedrag kan impact hebben op onze privacy, zeker wanneer derden deze gegevens over ons opslaan, onderling delen en gebruiken om ons leven ‘makkelijker’ te maken of voorspellingen te doen over ons gedrag. In veel gevallen kan dit leiden tot betere en op maat gemaakte diensten. Keerzijde is evenwel dat het risico bestaat dat persoonlijke informatie vrijgegeven wordt zonder men dat wil. PENTLAND (2014) pleit in dit verband voor een nieuw beleidsontwerp met werkbare garantie dat iedereen kan beschikken over gegevens die nodig zijn voor het algemeen belang, terwijl tegelijkertijd de burgers worden beschermd.

## Big data, beleid en beleidsonderzoek

Informatie is een belangrijke bron waarop beleidskeuzes kunnen berusten (naast macht, intuïtie, ervaring ...). Hoewel big data bijdragen aan de cijfermatige onderbouwing en evaluatie van beleid (‘evidence-based policy’) creëren zij voor de beleidswetenschap ook een tegenstrijdige situatie. De stortvloed van gegevens lijkt immers vooral aan te sluiten bij het ‘rationeel lineair model van beleid’ (HOOPERWERF & HERWEIJER 2002; DUNN 1994). Hierbij veronderstelt men dat het beleidsproces op te delen is in elkaar opvolgende fasen waarin er één doorslaggevende actor is. Precies dit model hebben vele beleidswetenschappers in de afgelopen jaren achter zich gelaten. Het lineair model helpt misschien wel als analysekader, maar tegelijkertijd sluit het onvoldoende aan bij de praktijk van beleid en beleidsonderzoek. De honger naar data lijkt er nu voor te zorgen dat het rationele model via de achterdeur toch weer aan belang wint

(los van de vraag of dat terecht of onterecht is) (BONGERS *et al.* 2015).

We verwachten dat big data een substantiële impact kunnen hebben op beleidsprocessen en -onderzoek. Hiervoor zijn er meerdere redenen. Vooreerst zijn er steeds meer data die relevant (kunnen) zijn voor het beleid, ze komen in groten getale en zijn in real time beschikbaar. Dit impliceert dat beleidsmakers niet altijd hoeven te wachten op resultaten van onderzoek waarin eerst informatie wordt verzameld, bewerkt en vervolgens gerapporteerd. Meer en meer gegevens kunnen in real time gemeten worden in plaats van periodiek (McAFEE & BRYNJOLFSSON 2012). Ten tweede zijn er in onze omgeving steeds meer apparaten die altijd en overal informatie verzamelen en ontsluiten. Dit zal in sommige gevallen traditioneel statistisch onderzoek overbodig maken (PENTLAND 2014). Ten derde krijgen informatieasymmetrie, informatiemonopolie en informatieoverlast een andere betekenis. De asymmetrie en het monopolie van informatie worden kleiner, omdat data steeds meer open zijn, gedeeld worden en zelf gegenereerd kunnen worden. De informatieoverlast kan wel groter worden (BONGERS *et al.* 2015).

Big data kunnen ons eveneens helpen om bijvoorbeeld beleidsevaluaties uit te voeren. Op zich zal triangulatie noodzakelijk blijven. Big data zullen niet alle aspecten van een evaluatie afdekken en ook kwalitatieve duidingen blijven nodig. Ook is met big data het attributieprobleem niet zomaar opgelost. De beschikbaarheid van gegevens beantwoordt op zich nog niet de vraag of een wijziging in een score op een prestatie-indicator veroorzaakt wordt door het beleid of dat andere factoren een rol spelen. Hiervoor blijven geavanceerde evaluatiemethoden als natuurlijke experimenten met experimentele en controlegroepen noodzakelijk (WEBBINK 2006).

Om kansen te grijpen die big data bieden, moet er evenwel nog een ontdekkingsreis worden ondernomen, zodat de meest waardevolle datasets worden gevonden en gekoppeld aan specifieke vraagstukken. Dat is geen sinecure, want eigenaars van datasets weten misschien helemaal niet dat hun data geschikt zijn voor dergelijke vraagstukken, terwijl beleidsmakers niet altijd

weten waar geschikte datasets aanwezig zijn en waar zij dus moeten zoeken. In dit perspectief kan een analogie met het 'garbage can model' worden gemaakt (MARCH & OLSEN 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 ('T HART *et al.* 1995). Bij het effectueren van big data zien we iets gelijkaardigs gebeuren, namelijk een koppeling van maatschappelijke uitdagingen ('problemen') aan big data ('oplossingen') door interactie tussen verschillende participanten.

Big data vragen om nieuwe specialismen en tools die nog niet in de breedte voorhanden zijn. Feitelijk heeft deze ontwikkeling last van een breder probleem, namelijk een tekort aan ICT-specialisten. Effectieve inzet van big data vergt echter ook capaciteiten inzake statistiek en kennis over de domeinen die men met big data wil analyseren of verkennen. Big data leiden tot de ontwikkeling van een nieuwe vakgebied en doen ook een beroep op nieuwe vaardigheden. Het nieuwe vakgebied betreft 'big data analytics' oftewel "*the process of examining and interrogating big data assets to derive*" (YU 2012: 10).

In een wereld waarin steeds meer data gegenereerd worden, is het des te belangrijker om hiermee te kunnen omgaan. Niet alleen de hoeveelheid aan data, maar ook de variëteit, verwerkingssnelheid en de kwaliteit vormen een uitdaging. Wil men als overheid deze mogelijkheden optimaal benutten, vergt dat niet enkel en alleen ICT-vaardigheden. Daarnaast is ook statistische en inhoudelijke kennis van de materie onontbeerlijk in het uitbouwen van big data (DAAS 2017).

## **En dan nu naar de praktijk: toepassing analytics in de kinderopvang en de rol van Kind en Gezin**

### **a. Kinderopvang**

Kind en Gezin is een intern verzelfstandigd agentschap van de Vlaamse overheid met rechts-

persoonlijkheid. Het heeft als opdracht om actief bij te dragen tot het welzijn van jonge kinderen en hun gezinnen door dienstverlening op de beleidsvelden preventieve gezinsondersteuning, kinderopvang, adoptie en groeipakket (vroeger kinderbijslag). Het kinderopvangbeleid wordt bepaald door de Vlaamse minister van Welzijn, Volksgezondheid en Gezin. Kind en Gezin voert het beleid uit.

Kind en Gezin organiseert zelf geen opvang, maar treedt op als algemene regisseur van de kinderopvang in Vlaanderen en in Brussel. Kind en Gezin is zowel bevoegd voor de opvang van de allerkleinsten als voor sommige opvang van kinderen uit het basisonderwijs (<https://www.kindengezin.be/>).

Zodra iemand baby's en peuters beroepsmatig en tegen betaling opvangt, moet men aan Kind en Gezin een vergunning aanvragen en daarvoor aan de wettelijke verplichtingen voldoen. Hierbij maakt men een onderscheid tussen twee soorten kinderopvang: gezinsonopvang (met maximaal 8 opvangplaatsen) en groepsopvang (minimaal 9 opvangplaatsen).

### **b. Toezicht en controle**

Kind en Gezin heeft naast zijn preventieve en ondersteunende rol ook een controlerende rol. En ondanks het opvolgen van de preventie maatregelen en de kwaliteit van de dienstverlening van de partners van Kind en Gezin, loopt het soms fout en kunnen in de voorzieningen van de partners in kinderopvang gevaarsituaties ontstaan. Om deze controlerende rol mee in te vullen, doet Kind en Gezin beroep op het agentschap Zorginspectie, dat eveneens resideert onder het Departement Welzijn, Volksgezondheid en Gezin (WVG).

Zorginspectie heeft twee kerntaken: inspecteren en rapporteren. De voorzieningen die door het departement of door de andere agentschappen van het beleidsdomein WVG worden erkend, geattesteerd, vergund of gesubsidieerd, onderwerpt Zorginspectie regelmatig aan een controle. Men inspecteert op eigen initiatief of op vraag van Kind en Gezin. Na elke inspectie rapporteert Zorginspectie haar bevindingen aan

het departement of aan een van de agentschappen van het beleidsdomein VVG. Die beslissen of er gevolgen zijn voor de verdere erkenning, vergunning of toekenning van subsidies.

Zorginspectie spreekt zich niet uit over de mogelijke gevolgen van een inspectie voor een voorziening. De Vlaamse overheid heeft bewust gekozen om de inspectiefunctie te scheiden van de erkennings- en subsidiëeringsfunctie. Dat betekent dat Zorginspectie inspecteert, en vervolgens over deze inspectie rapporteert aan het Departement VVG, en *i.c.* Kind en Gezin.

### c. Proefproject

Men kampte al lang met de vraag op welke manier de toezichtcapaciteit voor de kinderopvangvoorzieningen op de meest optimale wijze kan worden ingevuld of opgevolgd. Er werden systemen ontwikkeld met 'aandachtsvoorzieningen' en er werd een analyse-instrument gemaakt. Zeker bij dit laatste was er geen draagvlak om dit algemeen in te voeren, en dit omwille van de complexiteit en de moeilijke hanteerbaarheid. Eind 2013, begin 2014 startte Kind en Gezin samen met de private partner IBM een proefproject om de toegevoegde waarde van het gebruik van datamining te peilen, om zo via indicatoren in kaart te brengen welke kinderopvangvoorzieningen meer aandacht verdienen vanuit het oogpunt van de klantenbeheerders, alvorens te worden onderworpen aan een inspectie ter plaatse.

De vraag was dus of het mogelijk was om aan de hand van de gegevens die zich in een dossier bevinden, te voorspellen of een dossierbehandelaar een dossier gerichter moet opvolgen. De dossiergegevens besloegen zowel het administratief dossier van een kinderopvangvoorziening, een mogelijk bestaand klachtendossier als de gegevens uit inspectie. IBM slaagde erin om een model te bouwen dat meer dan 90 % zekerheid gaf dat aandacht nodig zou zijn, gebaseerd op 11 tot 13 kenmerken in de dossiergegevens. Deze kenmerken waren geen kwalificaties door klantenbeheerders, maar waren kenmerken gebaseerd op (min of meer) objectieerbare gegevens. IBM gebruikte voor de bouw van zijn model software uit zijn eigen rekken: SPSS.

Vanaf 1 april 2014 wijzigde de regelgeving en impliciet de structuur rond de kinderopvang grondig, waardoor het model van IBM niet meer toepasbaar was. De studie toonde wel aan dat het mogelijk is om een computemodel te bouwen met een redelijk voorspellende waarde dat dienstig is voor preventieve doeleinden. Eind 2014 werd dan ook beslist om de analyse in de kinderopvang als proefproject datamining een vervolg te geven. Dit was het resultaat van een oefening waarbij beschikbaar budget, mankracht en expertise in de weegschaal lagen. Het resulteerde in de opzet van een overkoepelende samenwerking van Kind & Gezin met het departement VVG, waar binnen de afdeling Beleidsinformatie en Kennisbeheer reeds ervaring rond data science aanwezig was. Er werd beslist om geen commerciële partner in te schakelen bij de ontwikkeling van het nieuwe voorspellingsmodel, maar het gehele traject aan te vatten met eigen mankracht.

Om de benodigde data te ontsluiten uit de verschillende databronnen, werd gebruikgemaakt van een ETL-softwarepakket dat reeds binnen Kind & Gezin aanwezig was. Voor de ontwikkeling van het voorspellingsmodel zelf is gebruikgemaakt van de opensourceprogrammeeromgeving *RStudio*, waarbinnen met *R* geprogrammeerd kan worden. *R* is een opensourceprogrammeertaal ontwikkeld voor statistiek en data-analysedoeleinden (<https://www.r-project.org/>). Voor de visualisatie van de output die wordt gegenereerd door het voorspellingsmodel, kan worden gebruikgemaakt van *Microsoft PowerBI*.

### d. Datamining gefaseerd

Typisch voor datamining is dat de ontwikkeling volgens een iteratief proces verloopt. Een model om de praktijk qua datamining te duiden, is de zogenaamde CRISP-DM-methodologie<sup>2</sup>. Het geeft een gestructureerde benadering van dataminingprojecten, zoals geïllustreerd wordt in figuur 1.

Men maakt het onderscheid tussen zes fasen, hoewel het op zich geen lineair proces is, maar eerder een samenspel van en een terugschakeling tussen de verschillende fasen. De eerste fase

in het overzicht is die van het begrijpen van de business, *i.e.* zicht krijgen op de noden, behoeften en de context van het thema dat men op grond van datamining wil aanpakken. Een tweede fase is die van het begrijpen van de data, een proces van exploreren en leren op grond van de data. Een derde is die van het klaarmaken ('prepareren') van de data. Het is in deze fase dat de data-architectuur vorm krijgt, waarbij ruwe data worden klaargezet om te gebruiken in de daaropvolgende fase van het modelleren. In deze modelleringsfase worden de technieken

geselecteerd, toegepast en geoptimaliseerd om bijvoorbeeld voorspellingen te doen, zoals ook het geval was in de case van Kind & Gezin. Een essentieel onderdeel van een dataminingproject is de fase van het evalueren van de verschillende stappen in het project. Pas als het model grondig is geëvalueerd, is er de fase van het toepassen van het model in de zogenaamde business. Dit is op zich geen eindfase van een dataminingproject, gezien een constante alertheid nodig is omtrent de onderscheiden fasen van het dataminingproject.

**Figuur 1** De fasen in een dataminingproject



Bron: <https://www.the-modeling-agency.com/crisp-dm.pdf>

Deze fasering kunnen we ook toepassen in het project van Kind & Gezin. Naast het uitklaren van de businessvragen, is er ook een technische kant die moet worden geanalyseerd. Hierbij wou men inschatten of het mogelijk was om op eigen kracht en met een nieuwe tool een vergelijkbaar resultaat te behalen.

### **e. Businessuitdagingen**

Alles start met het uitklaren van de businessvragen waarop de datamining een antwoord moet vinden.

Hierbij een kort overzicht van deze vragen:

- Kan de bestaande businessexpertise vergroot worden met een geautomatiseerd proces dat op basis van data aangeeft dat aandacht nodig is voor een kinderopvanglocatie, nog vooraleer een inspectie heeft plaatsgevonden?
- Omwille van gewijzigde regelgeving is er een breuk in de historiek van de beschikbare data. Welke impact kan dit hebben op het voorspellingsmodel?
- Kunnen en mogen alle databronnen ontsloten worden die nodig zijn voor de bouw en werking van het voorspellingsmodel?

- Is er steeds toegang tot de meest recente, operationele data?
- Kan de business een inzicht geven in de belangrijke predictoren?
- Wat is een haalbare en bruikbare frequentie van het predictieve proces?
- Kan de output van het model op makkelijk te interpreteren wijze worden gevisualiseerd?

Het was cruciaal om de business van meet af aan bij het project te betrekken, niet alleen om het doel van het project te kaderen, maar ook om te benadrukken dat dit predictieve model geen instrument zou zijn ter vervanging van hun expertise, maar veeleer als aanvulling en ondersteuning daarvan.

In een eerste fase werden dan ook de 'oude' databronnen gebruikt, waarop met *R* verschillende modellen werden gebouwd. Het resultaat hiervan was bevredigend en vanuit statistisch oogpunt zelfs een verbetering te noemen in vergelijking met de *oude* opzet. Dit was een belangrijke mijlpaal voor het verdere verloop van het project. In een tweede fase werd een model ontwikkeld op de nieuwe databronnen. Enerzijds betreft dit data die beschikbaar zijn in het datawarehouse van Kind en Gezin, data van de klachtendienst van Kind en Gezin en anderzijds externe data van Zorginspectie.

Ook nu kon statistisch aangetoond worden dat de output van het model een goede voorspellende waarde biedt. Correlaties tussen variabelen konden via 'plotgrafieken' in beeld worden gebracht en aan de hand van statistische meetwaarden kon de kwaliteit van het model worden gescoord.

Als finale stap werd dan de output gevisualiseerd op een voor de business duidelijk te interpreteren manier. Hier werd gekozen om te werken volgens het systeem van sterren al naargelang de kwaliteit. Zo krijgen de kinderopvanglocaties sterren van 1 tot 5. Hoe meer sterren, hoe groter de kans op service van topkwaliteit, hoe minder sterren, hoe groter de kans dat een dossier gerichter opgevolgd moet worden. Bovendien wordt ook een kleurenschaal gehanteerd: 1 ster = rood ... 5 sterren = groen, wat het visueel

aantrekkelijker maakt. Op maandelijkse basis worden de predicties aangeboden.

## Opvolging

Het onderhoud en de opvolging van een dergelijk opzet is vanzelfsprekend noodzakelijk. Wetgeving kan immers wijzigen, functionele structuren kunnen wijzigen, gebruikte variabelen kunnen verdwijnen of er kan net nieuwe informatie beschikbaar worden die voor verdere verfijning kan zorgen. Op geregelde tijdstippen moet dan ook worden teruggekoppeld met de business. Op die manier kan het predictieve model optimaal als extra, complementair instrument ingezet worden, waarbij het individuele oordeel van de businessmedewerker kan afgetoetst worden aan de output die gebaseerd is op statistische meetwaarden. Finaal ligt de eindbeslissing om al dan niet over te gaan tot controle of opvolging van de kinderopvangvoorzieningen, immers bij de business.

## Conclusie

De toepassing uit de praktijk van Kind & Gezin is een illustratie van het gegeven dat overheden zich bewust zijn van het belang van big data. Het is een voorbeeld van hoe een overheid op dit vlak ook het verschil kan maken. Door samenwerking over de grenzen van disciplines en agentschappen heen, werden kennis en ervaring gebundeld en meerwaarde gecreëerd.

De verzameling en ontsluiting van grote databestanden is een relevante ontwikkeling die vanuit het beleid en het beleidsonderzoek de volle aandacht moet genieten. Het grijpen van kansen die deze ontwikkelingen bieden, is echter een onzekere ontdekkingsreis. Hoewel technologieën het verzamelen, ontsluiten en bewerken van big data steeds eenvoudiger maken, betekent dat geenszins dat vraag en aanbod elkaar weten te vinden. Wellicht doen de beste kansen zich voor op het snijvlak van organisaties en sectoren die thans nog weinig met elkaar te maken hebben. Het is dus niet alleen een ontdekkingsreis naar de mogelijkheden van big data. Het is evenzeer een zoektocht naar de onverwachte combinaties en de maatschappelijke en economische behoef-

ten waar big data potentieel een grote rol kunnen spelen.

Er lijkt sprake te zijn van een hype, maar er zijn voldoende redenen om aan te nemen dat big data een vaste plek zullen verwerven in het beleid en beleidsonderzoek. De explosie van gegevens lijkt vooralsnog geen grenzen te kennen. Bovendien zullen er steeds meer apparaten en applicaties onderling verbonden raken en steeds meer informatie uitwisselen en genereren. De grote vraag hierbij is hoe daar verstandig mee om te gaan. Dat wil zeggen de gegevens op een effectieve manier bewerken en benutten en dat binnen de wetgevende kaders, die zorgen voor betrouwbaarheid, veiligheid en bescherming van persoonsgegevens. De overheid kan hier een cruciale rol spelen. Enerzijds als aanbieder en gebruiker van big data, anderzijds als partij die derden stimuleert en bewust maakt in dit domein, maar ook regels durft te stellen ten aanzien van het gebruik.

Beleidsmakers moeten zich regelmatig de vraag stellen of big data een rol kunnen spelen. Voor

de onderzoeker geldt dat big data het nut en de noodzaak van 'traditionele' onderzoeksmethoden onder druk zetten. Zij vormen daarmee een uitdaging voor het eigen vakgebied.

Big data worden vaak gezien als een rationeel ideaal waarin het technologische op de voorgrond treedt. Niettemin, de benutting van kansen op grond van big data in de publieke dienstverlening vereist ook overheidsstrategieën, beleid dat door mensen gemaakt wordt. In dezelfde context kunnen we meegeven dat beleid niet alleen een kwestie is van informatie, maar even goed een kwestie van waarden en machtsverhoudingen (MEIJER & VAN BERLO 2013).

De vraag blijft echter of big data ook tot beter onderzoek leiden (en tot betere besluiten). Discussies over betrouwbaarheid van data zullen blijven bestaan. Het gaat daarbij zowel om wat big data weergeven (welke metingen eraan ten grondslag liggen), hoe compleet deze data zijn, op welke wijze zij geïnterpreteerd en bewerkt worden. Daarom moeten eens te meer statistiek, inhoud en IT elkaar vinden.



## Referenties

AUSTRALIAN GOVERNMENT (2013). *The Australian Public Service Big Data Strategy Improved understanding through enhanced data-analytics capability*. Canberra.

BONGERS, F., VELDKAMP, J. & VAN DER VORST, T. (2015). *Big data: een zoektocht voor Beleid*. Utrecht.

DAAS, P. (2017). "Van Big Data naar Officiële Statistiek. Spanning tussen theorie en data gedreven manier van werken". Lezing op de studiedag '20 jaar peilen in Vlaanderen', Leuven, 31 januari 2017.

DE KOOL, D. (2014). *De kansen en risico's van Big Data voor de overheid*. Paper politicologenetmaal, Maastricht, 12-13 juni 2014.

DIALOGIC (2014). *Big Data in onderwijs en wetenschap. Inventarisatie en essays*. Utrecht.

DUNN, W.N. (1994). *Public Policy Analysis: An Introduction*. Englewood Cliffs: Prentice Halls.

HM GOVERNMENT (2013). *Seizing the data opportunity. A strategy for UK data capability*. Londen.

'T HART, P., METSELAAR, M. & VERBEEK, B. (1995). *Publieke besluitvorming*. Den Haag: VUGA.

HOOGERWERF, A. & HERWEIJER, M. (2002). *Een inleiding in de beleidswetenschap*. Alphen a/d Rijn: Samsom.

<https://www.ibm.com/products/spss-modeler>.

<https://www.kindengezin.be/>.

<https://www.rproject.org/>.

MAYER-SCHONBERGER, V. & CUKIER, K. (2013). *De Big Data Revolutie. Hoe de data- explosie al onze vragen gaat beantwoorden*. Amsterdam: Maven Publishing BV.

MARCH, J.G. & OLSEN, J.P. (1976). *Ambiguity and choice in organizations*. Bergen: Universitetsforlaget.

McAFEE, A. & BRYNJOLFSSON, E. (2012). "Big data: de managementrevolutie". In *Management Executive*, 10 (6), 18-24.

MEIJER, A. & VAN BERLO, D. (2013). "Big Data : overheidsbeleid in De Gekende Samenleving". In *Bestuurswetenschappen*, 67 (5-6), 163-178.

PENTLAND, A. (2014). *Sociale big data. Opkomst van de data-gedreven samenleving*. Amsterdam: Maven Publishing.

VAN DER VOORT, H. & CROMPVOETS, J. (2016). "Big Data: een zoektocht naar instituties". In *Bestuurskunde*, 25 (1), 3-7.

YU, C. (2012). *The Big Data Opportunity Making government faster, smarter and more personal*. Londen.

WEBBINK, H.D. (2006). "Causale effecten van beleid. Over gecontroleerde en natuurlijke experimenten". In *Justitiële Verkenningen*, 31 (8), 72-84.

## EINDNOTEN

- 1 *Frank Bongers is principal consultant bij Dialogic te Utrecht, bongers@dialogic.nl. Koen De Bruyn is projectleider ICT Datawarehouse/Analytics bij Kind & Gezin, koen.debruyne@kindengezin.be. Dries Verlet is adviseur voor Statistiek Vlaanderen. Tussen 2007 en 2017 was hij adviseur beleidsevaluatie binnen de Studiedienst van de Vlaamse Regering. Verder is hij als gastdocent actief binnen de opleiding Bestuurskunde en Publiek Management van de UGent, Dries.Verlet@vlaanderen.be.*
- 2 *Deze afkorting staat voor: Cross-Industry Standard Process for Data Mining (<https://www.the-modeling-agency.com/crisp-dm.pdf>).*