

“Big data zijn een
verrijking voor de
omgevingsanalyse”



Big data gebruiken bij uw strategiebepaling

Steeds meer bedrijven en organisaties houden vele gegevens bij en gebruiken deze om hun dienstverlening te verbeteren. U kent het wel van websites als Bol.com: “Mensen die dit artikel bestelden, bestelden ook vaak...” Ook de overheid verzamelt veel gegevens en maakt deze steeds vollediger en sneller online beschikbaar. Onderwijsorganisaties kunnen deze data goed gebruiken bij strategische vraagstukken. Puk Witte en Sandra Beekhoven verkennen de potentie van big data bij strategievorming.

Wat zijn ‘big data’?

Big data. Het is een veelgebruikte, maar moeilijk te definiëren term. Het is een ‘parapluterm’ voor data, die de volgende kenmerken hebben:

- Ze hebben een groot volume.
- Ze worden in een hoge snelheid geproduceerd (velocity).
- Ze hebben een grote variëteit aan vormen (variety).
- De betrouwbaarheid van de data is niet altijd hoog (veracity).

We staan nader stil bij deze kenmerken en geven aan hoe data die voor het onderwijs zijn te gebruiken zich verhouden tot deze aspecten.

Volume

Het gaat om een grote hoeveelheid data, maar ‘een grote hoeveelheid’ is een subjectieve aanduiding. Bij big data denken we aan een enorme hoeveelheid gegevens, grote zware datasets, te groot om op één enkele server te

kunnen draaien. Hoewel de hoeveelheid onderwijsgerelateerde data is toegenomen, is het volume nog vrij beperkt in vergelijking met beschikbare data op andere terreinen. Maar de ontwikkelingen op dit vlak gaan wel erg snel.

Velocity

Velocity is de snelheid waarmee de data doorkomen. Real life streaming past meer bij big data dan periodieke updates. Een voorbeeld van life streaming is betalingsverkeer in een winkel. Bij de openbare data over onderwijs is nog vrijwel geen sprake van live streaming. Leerlingen tellen we nog steeds op 1 oktober, waarna er in het volgend kalenderjaar een update plaatsvindt. Digitale leermiddelen, zoals Snappet, slaan de leerlingenprestaties op en produceren daardoor wel data met een hoge ‘velocity’.

Variety

Er zijn verschillende datavormen, die zowel gestructureerd als ongestructureerd kunnen zijn. Hoewel het volume van big data vaak als een

probleem wordt ervaren, is de ongestructureerdheid van big data de werkelijke uitdaging. Ongestructureerde data zijn data die niet zomaar in een database met rijen en kolommen passen, bijvoorbeeld tekst, foto's en socialemediaberichten. Waar mogelijk en legaal, kan wat er op sociale platforms wordt uitgewisseld, worden omgezet in data. Denk bijvoorbeeld aan Twitter: hoeveel wordt er getwitterd, door wie, welke hashtags gebruikt men, welke woorden worden er gebruikt en wat is het sentiment van de berichten? Voor schoolbesturen kan het interessant zijn te weten hoe er in de regio wordt gereageerd op bijvoorbeeld de gevolgen van passend onderwijs of de komst van een vluchtelingencentrum.

Veracity

We moeten rekening houden met de bias, ruis en abnormaliteit, ofwel de onzekerheid van data. Net zoals alle data, bevatten ook big data onjuistheden en foutjes. Dit wordt bij de analyse echter niet altijd in de eerste stap gecontroleerd; de snelheid maakt dat moeilijk en de veelheid maakt het minder nodig.

Er is geen regel die stelt dat we pas mogen spreken van big data als aan alle vier de voorwaarden is voldaan. Sommige auteurs voegen zelfs nóg een v toe aan de kenmerken van big data: value (waarde, waarover later meer). Hieronder zoomen we vooral in op het gebruik van openbare databronnen met gestructureerde en ongestructureerde data, en op methodes om big data te analyseren.

Big data en strategiebepaling

Elke onderwijsorganisatie bevindt zich in een omgeving met bepaalde kenmerken en ontwikkelingen, bijvoorbeeld krimp of immigratie. Het zijn meestal factoren waarop een bestuur geen invloed heeft, maar die wel van belang zijn voor de onderwijsorganisatie. Om een gedegen strategie te formuleren, is het van belang om de omgeving en de ontwikkelingen in die omgeving in kaart te brengen. Dit gebeurt vaak middels een omgevingsanalyse, ook wel een DESTEP-analyse genoemd (Demografie, Economie, Sociaal, Technologie, Ecologie, Politiek (juridisch)).



“Hoe kun je met big data toegevoegde waarde creëren?”



Er zijn tegenwoordig vele databronnen die feitelijke informatie geven over de verschillende aspecten van de omgevingsanalyse en die dus bruikbaar zijn om de omgeving van de onderwijsorganisatie gedegen in beeld te brengen. De belangrijkste bronnen van openbare onderwijsdata zijn DUO en het CBS. De databestanden die op de websites van deze organisaties staan, zijn tevens de bron voor vele andere dashboards en informatiewebsites, bijvoorbeeld de website www.scholenopdekaart.nl.

Voor een omgevingsanalyse zijn deze twee openbare databronnen nuttig, maar ze geven niet voldoende inzicht in alle aspecten van de omgeving. Onderstaand kader bevat een aantal relevante bronnen die schoolorganisaties kunnen raadplegen om meer inzicht te krijgen in de omgeving waarin de organisatie opereert. De infographic op pagina 24 en 25 geeft een voorbeeld van een omgevingsanalyse voor een fictief schoolbestuur.

Relevante beschikbare openbare data

Gestructureerde data

Veel bronnen leveren informatie die relatief weinig bewerking vergt als voorbereiding op verdere analyse. Relevante gestructureerde data zijn onder andere te vinden op:

⇒ www.duo.nl/open_onderwijsdata

Naar welk type mbo-opleidingen stromen vmbo-ers uit?

⇒ www.cbsinuwbuurt.nl

Hoeveel kinderdagverblijven zijn er in de wijk waar de school staat?

⇒ www.arbeidsmarktincijfers.nl

Hoeveel werkzoekenden zijn er in onze regio per beroep?

⇒ www.verkiezingsuitslagen.nl

Wat was de opkomst en uitslag bij de laatste verkiezingen?

Ongestructureerde data

Ook fora, nieuwssites en sociale media bevatten informatie die relevant kan zijn voor een schoolbestuur. Relevante ongestructureerde data zijn onder andere te vinden op:

⇒ www.ouders.nl

Een forum waar ouders met elkaar in gesprek gaan over wat hen bezighoudt.

⇒ www.twitter.com

The-talk-of-the-day over vrijwel elk denkbaar onderwerp.

Data analyseren

Big data worden op andere manier geanalyseerd dan in de traditionele (sociaal)wetenschappelijke data-analyses gebruikelijk is. Bij traditionele wetenschappelijke data-analyses stelt men op basis van wetenschappelijke literatuur een klein aantal hypothesen op, die men vervolgens toetst met de beschikbare data. Bij big data-analyses gaat men in de data op zoek naar een breed palet aan verbanden, die op het eerste oog interessant kunnen zijn. Men veronderstelt van tevoren geen causaliteit of effectgrootte. Maar om niet te 'verdrinken' in het moeras van data, formuleert men wel richtinggevende vragen. Het meer 'onbevangen' benaderen van de data levert vaak verrassende inzichten op. Er worden bijvoorbeeld verbanden gevonden die eerder niet werden gelegd.

De besproken kenmerken (de vier v's), ofwel de veelheid, snelheid, diversiteit en onzekerheid van big data, maken dat bestuurders soms geen beslissingen durven te nemen op basis van big data: *rubbish in* is immers *rubbish out*. Daarom zijn er auteurs die een vijfde kenmerk bespreken: de v van 'value'. Hoe kun je met big data toegevoegde waarde creëren? Op dit moment speelt dit punt voor het onderwijs nog niet zo, omdat de veelheid en de snelheid van data beperkt zijn en de meeste databronnen behoorlijk betrouwbaar zijn.

Maar ook wij vinden een kritische blik bij de interpretatie van de analyse noodzakelijk. We gebruiken big data daarom ondersteunend, of als aanleiding om verder te onderzoeken waarom er bepaalde verbanden bestaan.

Inzichten uit de onderwijskunde, sociologie en bedrijfskunde zijn belangrijk voor het stellen van de juiste vragen bij het uitvoeren van de analyses. Ook bij het interpreteren van de uitkomsten is wetenschappelijke kennis onontbeerlijk. Zo is bijvoorbeeld niet alles wat met elkaar samenhangt, te vertalen naar een oorzaak-en-gevolg-relatie. Vaak wordt samenhang tussen twee factoren veroorzaakt door een derde factor. Bijvoorbeeld: we lezen in een krant dat kinderen met bepaalde voornamen later succesvoller zijn dan kinderen met andere namen. Verschillen in succes worden onder meer veroorzaakt door verschillen in de sociaal-culturele achtergrond van gezinnen. In de verschillende sociaal-culturele bevolkingsgroepen zijn verschillende namen populair. Het verband tussen de naam en het succes van kinderen, wordt dus veroorzaakt door hun sociaal-culturele achtergrond.

Big data zijn een verrijking voor de omgevingsanalyse, maar de omgevingsanalyse is slechts een onderdeel van de integrale analyse die nodig is voor strategiebegrip. Om te bepalen hoe de organisatie kan of moet reageren op de omgeving, is het belangrijk dat ook de eigen organisatie goed in kaart wordt gebracht. Voor deze interne analyse zijn de data van de organisatie zelf cruciaal. Het gaat om kwantitatieve data uit het managementinformatiesysteem, zoals de opbouw van het personeelsbestand, maar vooral om kwalitatieve data over onder meer de historie en de cultuur van de organisatie. Deze informatie wordt uit documenten gehaald, maar vooral uit gesprekken met sleutelfiguren. Zo krijgen meer cijfermatige data kleur en diepgang. Als reactie op de ontwikkeling van big data, wordt deze informatie soms omschreven als 'small data'.

Tot slot

Wetenschappers zijn volop aan het ontdekken hoe zij big data kunnen gebruiken en welke analysetechnieken zij daarvoor kunnen inzetten. Voor steeds meer mensen is het gewoon om met data te werken. In het onderwijs is dit nog relatief nieuw, maar er komen steeds meer bruikbare databronnen en wij verwachten dat dit er alleen nog maar meer zullen worden.

Sardes is volop in ontwikkeling op het terrein van 'datagedreven werken' en het analyseren van veel en verschillende soorten data. Een innovatieve benadering die veel, en vaak verrassende, inzichten oplevert. Daarbij verliezen we nooit de mensen, de context en de strategische vragen uit het oog. Welke data zet u als bestuurder om in zinvolle informatie?

Meer lezen

Schouten, G. (2016). *Big Data? Natuurlijk!* Lectorale rede lector Big Data. Eindhoven: Fontys Hogeschool ICT.

Davenport, T. (2015). *Big Data aan het werk. Wat je als manager moet weten*. Alphen aan den Rijn: Vakmedianet.

Onderzoek naar discussies via twitter over onderwijs in de VS:
<http://www.hashtagcommoncore.com/#5-2>

Meer bekijken

Animatie 'Strategisch besturen Sardes' op:
www.sardes.nl/strategischebesturen of
www.sardes.nl/strategischebesturen-vo.
Of zoek op YouTube naar: Sardes strategische besturen