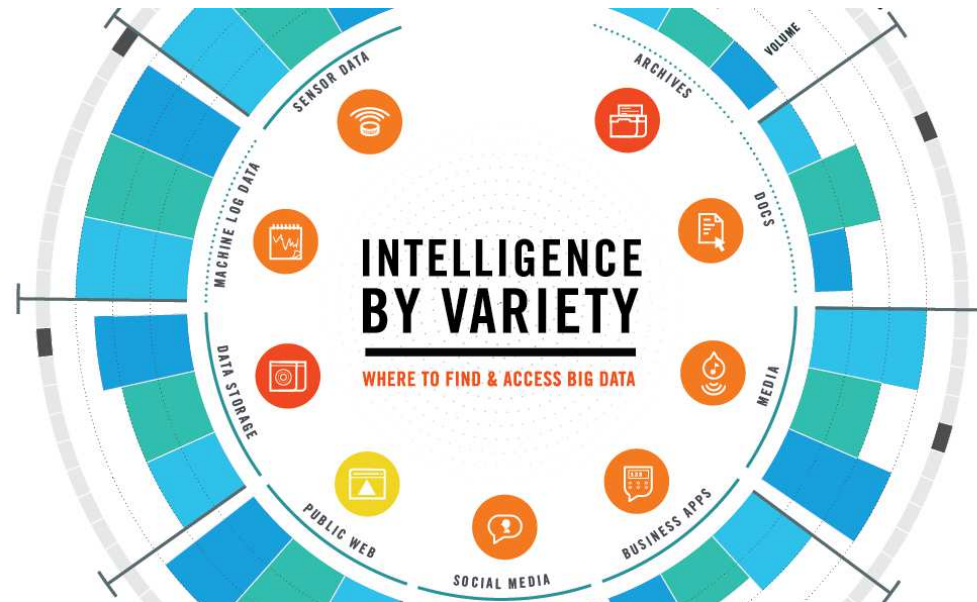


## Big Data: the art of the possible

### Samenvatting

Big Data is op dit moment een enorme hype. Zo'n veertig jaar na de start van het informatietijdperk neemt de snelheid waarmee data binnenstroomt nog steeds exponentieel toe, en lijken de mogelijkheden van wat we daarmee kunnen doen onbeperkt. Deze trendalert is gebaseerd op de vierdelige serie Big Data onderzoeksnotities van het Sogeti Verkenning-Instituut Nieuwe Technologie (VINT). We gaan in op mogelijke consequenties voor de Nederlandse waterbedrijven; kansen voor bijvoorbeeld het optimaliseren van processen, voorspellen van events en het verkrijgen van inzicht in klantwensen. Bedrijven staan op het punt om in Big Data te investeren, maar wat kunnen ze verwachten en waar moeten ze op letten?



Bron: Kapow Software  
via VINT blog (<http://vint.sogeti.com/category/big-data/>)

## Consequenties voor u

	Laag	Middel	Hoog	Beknopte uitleg
Impact				Tot nu toe verborgen patronen kunnen d.m.v. big data oplossingen zichtbaar worden gemaakt, leidend tot hogere efficiëntie en optimalisering.
Zekerheid				De hoeveelheid data groeit exponentieel en technologie is in ontwikkeling; de vraag is niet of maar wanneer we met big data aan de slag gaan.



## Trendbeschrijving en achtergrond

Zo'n veertig jaar na de start van het informatietijdperk zijn alle ogen gericht op de basis daarvan: digitale data. De hoeveelheid data die we in die periode hebben verzameld, de variatie aan datatypen, en de (exponentieel toenemende) snelheid waarmee data wereldwijd onze databases binnenstroomt zijn enorm. Dataverzameling heeft zulke opzienbarende vormen aangenomen dat we het de laatste jaren hebben over 'big data'; een alsmaar uitdijend data-universum vol met onontdekte verbanden en verborgen informatie. En het gaat niet alleen om het verzamelen van data; ook de ontwikkeling van technologieën m.b.t. het verspreiden en delen van data heeft bijgedragen aan het ontstaan van het huidige informatietijdperk. Is Big Data de belangrijkste nieuwe grondstof en de motor van compleet nieuwe bedrijfseconomische en maatschappelijke mogelijkheden?

Deze trendalert is gebaseerd op de vierdelige serie onderzoeksnotities over Big Data van het Sogeti Verkenningsinstituut Nieuwe Technologie (VINT), dat onderzoek doet naar de praktische waarde van nieuwe informatietechnologieën voor organisaties. In de notities over Big Data worden antwoorden gegeven op de vragen wat Big Data eigenlijk is, welke mogelijkheden Big Data biedt, welke gevaren op de loer liggen en welke Big Data scenario's denkbaar zijn.

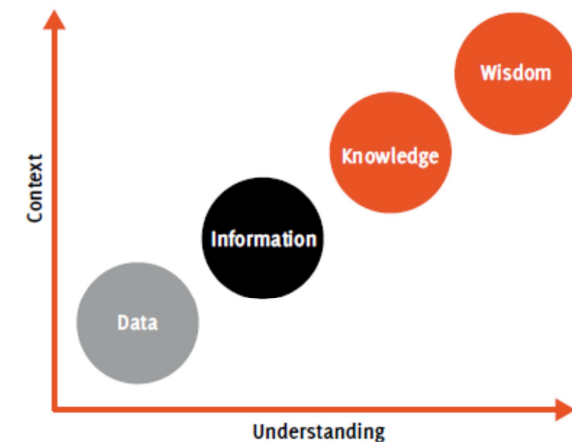
### Hoezo 'Big' Data?

Als we kijken naar onze tijdlijn volgens Wikipedia<sup>1</sup> leven we momenteel in het 'informatietijdperk', gekenmerkt door een digitale revolutie en explosieve datagroei. Maar wat is data? En wat betekent het als data 'big' is?

In het model dat aan de basis ligt van de VINT onderzoeksnotities (figuur 1) wordt het concept *data* gedefinieerd als een serie gegevens die ten grondslag ligt aan *informatie*. Informatie wordt daarbij gezien als geïnterpreteerde data. Hoe meer (relevante) data we hebben, hoe meer informatie we daaruit kunnen halen, en hoe beter we in staat zijn om de meest uiteenlopende dingen te begrijpen. Die begripsvorming is het punt waar informatie wordt omgezet naar (persoons- en contextafhankelijke) *kennis*. Hoe meer die kennis wordt geïnternaliseerd en door ervaring toepasbaar op zeer specifieke contexten, hoe meer er sprake is van intrinsieke *wijsheid*.

Naar aanleiding van de exponentiële datagroei werd in 2005 de term 'Big Data' geïntroduceerd door uitgever O'Reilly Media: een dataset die van zodanige grootte en complexiteit is dat de traditionele informatietechnologieën niet in staat zijn tot goed datamanagement. Het woord 'big' wordt doorgaans op

twee manieren gebruikt: enerzijds als het gaat over letterlijke afmetingen ('groot'), en anderzijds als het gaat over figuurlijke afmetingen ('groots'). In Big Data heeft het woord 'big' beide betekenissen; het gaat heel concreet over een gigantische hoeveelheid data, maar die data is wel ongrijpbaar. De VINT onderzoeksnotities verwijzen met Big Data naar het geheel van data en methoden en technologieën dat datamanagement op extreme schaal economisch rendabel maakt, waarbij het woord 'big' ook verwijst naar de grootsheid van de mogelijkheden die Big Data toepassingen met zich meebrengen.



Figuur 1. Data, informatie, kennis en wijsheid (Bron: VINT, 2013)

<sup>1</sup> [http://en.wikipedia.org/wiki/History\\_by\\_period](http://en.wikipedia.org/wiki/History_by_period)



Waarom we al die data willen hebben en bekijken is duidelijk volgens de VINT onderzoeksnotities. De ontwikkelingen rond Big Data worden ook wel gezien als 'de nieuwe microscoop', resulterend in het aanbreken van een nieuw tijdperk, met een fundamentele omslag in de wetenschap en praktijk. Sociale media, web analytics en allerlei logfiles en sensoren leveren waardevolle data op, de kosten van IT-oplossingen om die data te verwerken blijven dalen, en de computerkracht neemt toe. Over het algemeen wordt beweerd dat die ontwikkelingen organisaties in principe in staat stellen om de informatievloed te managen en wel tot 20% beter te presteren dan de concurrentie die de voordelen van Big Data niet benut. De wens om data naar relevante informatie te vertalen is niet nieuw, maar de complexiteit die de enorme groei in mogelijkheden rond dataverzameling, -opslag, en -analyse met zich meebrengt is dat wel. Daarnaast observeren de VINT trendonderzoekers dat het behalen van voordelen uit Big Data niet alleen afhankelijk is van de technologie, maar van de algehele mindset binnen een organisatie; de bedrijfsvoering zal minder gedreven moeten worden door intuïtie en meer door beschikbare data en informatie. Echter, al die mogelijkheden zijn niet altijd zomaar voorhanden; om een dergelijke situatie te bereiken is er een nieuwe aanpak nodig, wat betekent dat er geïnvesteerd moet worden in spullen, mensen, skills, processen, management en governance.

De uitdagingen en complexiteit rond Big Data ontstaan door de samenkomst van drie aspecten – ook wel de 3 V's genoemd: de datahoeveelheid (volume), de verschillende datatypen (variety), en de analysesnelheid (velocity).

#### Volume

De afgelopen decennia is de data- en informatievloed exponentieel gestegen. Waar we in de floppy-tijd nog dachten in kilobytes, werden dat megabytes in het cd-rom tijdperk, gigabytes in het USB-stick tijdperk en terabytes in het tijdperk van de draagbare externe harde schijven. Volgens ICT specialist The International Data Corporation<sup>2</sup> groeit het digitale universum in 2013 met 50% tot 4 Zetabytes. Dat is een verviervoudiging van de wereldwijde datahoeveelheid in 2010. Voorspeld wordt nu dat we in 2020 wereldwijd over 40 Zetabytes (40 miljard Gigabytes) aan informatie beschikken – gelijk aan een stapel dvd's die van de Aarde tot halverwege Mars reikt, maar vanwege de snelle ontwikkeling van nieuwe technologieën, die we nu nog niet voorzien, onderschatten we de groei waarschijnlijk.

<sup>2</sup> IDC (2012). Predictions 2013: Competing on the 3rd Platform. Beschikbaar via: <http://www.idc.com/research/Predictions13/downloadable/238044.pdf>

#### Variety

Data wordt zowel binnen organisaties zelf geproduceerd (documenten, e-mail, afbeeldingen, productielogs en -rapportages, sensor meetgegevens) als extern (sociale media, externe logs en meetgegevens). Dankzij het Internet ontstaan steeds meer mogelijkheden om al die databronnen met elkaar te verbinden, maar met name voor ongestructureerde data – data die niet volgens een vooraf gedefinieerde opzet is gestructureerd, zoals sociaal media berichten, afbeeldingen, digitale rapporten – is dat nog een enorme uitdaging omdat traditionele dataverwerkingssoftware daar meestal nog niet op is toegespitst. De relaties tussen al die verschillende databronnen zitten dan vaak 'verborgen' en zijn op het eerste gezicht niet vanzelfsprekend.

#### Velocity

Gelet op de technologische ontwikkeling voorspelde Gordon Moore dat het aantal transistors op een chip elke 18 maanden zou verdubbelen tegen steeds lagere kosten. Echter, de datagroei overtreft deze Wet van Moore tegenwoordig, en dat wordt alleen maar meer in de toekomst. Om toch zinvol met die data om te gaan (de juiste data te vertalen naar de juiste informatie, en het op het juiste moment en op de juiste locatie zichtbaar te maken) vereist dus geavanceerde hardware en software die al die data met enorme snelheid kunnen verwerken.



### Van Big Data naar Big Information

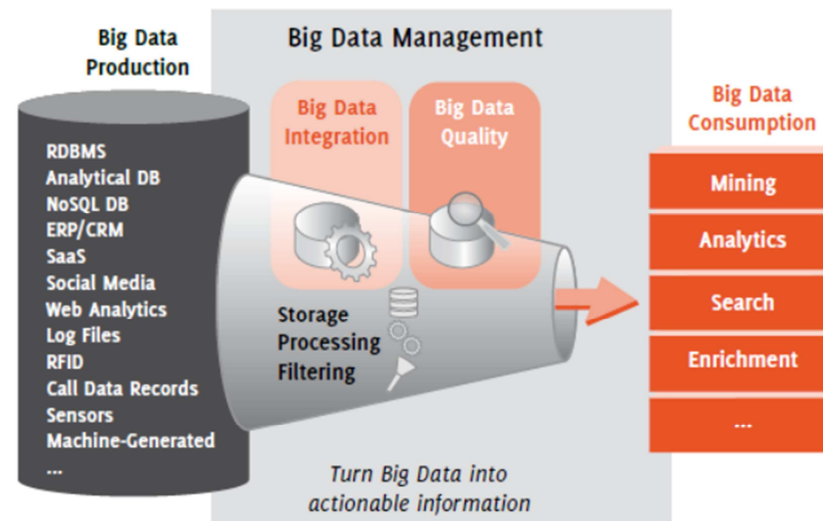
De waarde van Big Data zit niet zo zeer in de data zelf, maar in de relevante informatie die met de data gegenereerd kan worden. Verschillende stappen zijn te onderscheiden in het vertalen van Big Data naar 'Big Information' (figuur 2): dataverzameling ('productie'), data management ('integratie' en 'kwaliteit'), en vertaling naar toepasbare data ('consumptie').

#### Dataverzameling

'Meten is weten', zo klinkt het gezegde. Dat is niet het complete verhaal, maar weten kan wel beginnen met meten. Op allerlei manieren kan data automatisch worden verzameld; via sensors (omgevingsdata), via zogenaamde API's – of Application Programming Interfaces (social media data), logbestanden (machinedata), etc. Wat dataverzamelmethode gelijk hebben is dat ze gedrag registreren, van mensen, machines, of van de natuurlijke omgeving. Data kan uit zowel interne (zelf meten) als externe bronnen (aankopen) worden gehaald.

#### Datamanagement

Als meten – het genereren van data – de eerste stap is, dan is het managen van data de tweede stap. Dat gaat over drie dingen: opslag van data, verwerking van data, en filteren van data. Voor grootschalige gedistribueerde gegevensverwerking zijn traditionele databases ongeschikt en zijn nieuwe oplossingen



Figuur 2.

Het vertalen van Big Data naar Big information gebeurt in een opeenvolging van dataverzameling ('productie'), datamanagement, en datatoepassing ('consumptie').

(Bron: VINT, 2013)

nodig, zoals bijvoorbeeld het Hadoop Distributed File System. Daarnaast wordt het steeds meer de norm om servers te huren als virtuele machines in een cloudomgeving en daar data op te slaan. Met name als de data uiteindelijk geïntegreerd gaat worden met andere (externe) data bronnen – kan dat een uitkomst zijn. Na opslag en integratie is het van belang om de data te filteren, zodat het 'schoon' en bruikbaar wordt gemaakt voor de gewenste toepassingen.

#### Datatoepassing

Verzamelde, opgeslagen, geïntegreerde, en gefilterde data is klaar om vertaald te worden naar informatie voor een specifiek doel. Big Data doelstellingen worden doorgaans onderverdeeld in de drietrup: understand – predict – act. Ofwel: begrip kweken van wat er gebeurt (herkennen van patronen, analyseren van causale verbanden), het inzetten van patroonherkenning voor relevante voorspellingen van wat er in de toekomst zal gebeuren (aan de hand van bijv. simulatiemodellen), en het ontwikkelen van actieprotocollen om naar die voorziene toekomst te gaan handelen (i.p.v. het verleden als handelingsbasis te nemen).



### Big Data ontwikkelingen

De Big Data onderzoeksnotities van VINT signaleren de volgende negen Big Data ontwikkelingen:

#### Best practices zijn nog 'under construction'

Veel organisaties voeren op dit moment een zoektocht naar best practices rond Big Data, maar het veld is op dit moment nog zo erg in ontwikkeling dat men hooguit van '*emerging next practices*' kan spreken.

#### Big Data bezetenheid

Men wil 'iets' met Big Data, en wat precies dat maakt eigenlijk niet zo uit. Met een hoop Big Data kom je ongetwijfeld tot een hoop hele interessante informatie die de bedrijfsvoering ten gunste komt, maar vergeet niet dat een groot deel van die inzichten wellicht ook op andere goedkopere manieren te verkrijgen zijn. De Big Data bezetenheid ligt op de loer!

#### Social Analytics als Big Data-troetelkind

'Social analytics' – ofwel: het analyseren van sociaal gedrag van mensen – is op dit moment de Big Data toepassing die de grootste aanhang vindt binnen het bedrijfsleven. Met name als het gaat over het voorspellen van klantgedrag lopen een hoop organisaties warm voor Big Data.

#### De technologie breekt nu door

Technologisch gezien is er rond Big Data oplossingen van alles in beweging. Geavanceerde software die snel en effectief grote volumes data kunnen doorzoeken komen steeds meer binnen handbereik.

#### Big Data zal de belofte van hypertargeting inlossen

Hypertargeting - het verschaffen van informatie op basis van persoonlijke behoeften - behoort tot de mogelijkheden van Big Data toepassingen. De advertenties van Facebook zijn hier een voorbeeld van, maar ook de 'wellicht-vind-je-dit-ook-interessant' systematiek van bijvoorbeeld Bol.com.

#### Big Data als IT-feestje

Worden Big Data toepassingen echt binnenkort betaalbaar en ook praktisch implementeerbaar? Hier is veel meer voor nodig dan IT alleen. De mensen moeten de systemen ook willen en kunnen gebruiken.

#### Big Data-ReturnOn Investment tekent zich af

Twee jaar na publicatie van het vaak geciteerde McKinsey rapport over Big Data<sup>3</sup> blijken de ontwikkelingen niet zo hard te zijn gaan als verwacht. Dankzij aan aantal pioniers met *emerging next practices* begint de Big Data-ROI zich nu af te tekenen.

<sup>3</sup> McKinsey Global Institute (2011): Big Data: The Next Frontier for Innovation, Competition, and Productivity.

#### Toekomstbeelden

De huidige Big Data praktijken en voorziene ontwikkelingen in ogenschouw genomen schetsen de VINT trendonderzoekers een negental (elkaar niet uitsluitende) potentiële toekomstbeelden die binnen een aantal jaar werkelijkheid zouden kunnen zijn.

#### Zelfdenkende datasystemen

Allerlei verbanden zitten verborgen in Big Data: patronen waar nog nooit iemand bij heeft stilgestaan. Het is niet ondenkbaar dat systemen in de toekomst zelf op zoek gaan naar correlaties en ze ons proactief voorschotelen.

#### Levensechte persona's creëren

Persona's – concrete persoonsbeschrijvingen die kenmerkend zijn voor typische klanten – worden in de toekomst steeds meer gemaakt op basis van echte klantdata. Hierdoor worden voorspellingen over klantbehoeften beter en wordt het ontwikkelen van relevante nieuwe producten of diensten gemakkelijker.

#### Real-time voorspellen

De huidige datatoevloed maakt het mogelijk om real-time data te matchen met patronen uit het verleden. Voorspellingen worden dan niet meer gebaseerd op fysisch-deterministische modellen, maar op 'black boxes' die patronen herkennen. Als we op die manier kijken naar het 'Big Data vergezicht', wordt al



gesproken over een wereld waarin theorievorming en de wetenschappelijke methode overbodig wordt.

### Naar super slimme organisaties

Alles in een organisatie is tegenwoordig digitaal: e-mail, financiën, licht- en klimaatsystemen, presentaties, urenregistratie, noem maar op. Met behulp van de (geïntegreerde) data die daar uit voortkomt kunnen bedrijfsprocessen worden geoptimaliseerd op basis van de daadwerkelijke gang van zaken. Een systeem dat op die manier de 'digitale organisatie' in kaart brengt kan problemen en kansen signaleren, sturen op risico's en ongewenst gedrag, en wellicht zelfs voorstellen doen voor sociale verbeteringen.

### Validatie en zelfregulatie door variatie

Systemen zullen zelf 'testen' met beschikbare data kunnen uitvoeren om tot een meest optimale interventie te komen. Door (binnen bepaalde marges) op een intelligente manier te bepalen welke interventies mogelijk positieve uitwerkingen hebben (voorkomen van normoverschrijding, aanpassen verbruikspatronen, etc.) kunnen uitkomsten worden getest en vertaald worden naar optimale oplossingen.

### Het einde van onvoorspelbaarheid?

We verzamelen meer en meer data; maar tot hoe ver gaat dat? Kunnen we over zó veel data beschikken dat we tot een situatie komen waarin we de bewegingen op

de markten, in koersen en innovaties kunnen zien aankomen? Hoe zal onze economie er uit gaan zien als gedrag kan worden voorzien en er dus geen sprake meer is van risico's nemen in ruil voor geld?

### De transparante mens

Data over (vaak onbewust) gedrag van mensen – ook wel 'Big Social' genoemd – kan een mooie rol spelen om mensen beter inzicht te geven in wie ze werkelijk zijn en welk gedrag ze vertonen: wie ben je, hoe reageer je op allerlei prikkels, in wat voor werk komen je kwaliteiten het best tot zijn recht, en welke relaties hebben de grootste kans van slagen? Data kan dan worden ingezet om de 'conditio humana' te kraken.

### Het is niet alles goud wat er blinkt

Frankwatching schetst in één van hun gepubliceerde online artikelen over Big Data<sup>4</sup> een overzicht van de risico's en gevaren die Big Data met zich meebrengen. Die schuilen in de combinatie van verschillende factoren, als volgt samengevat:

<sup>4</sup> Frankwatching (2013). Ober? Er zit big data in mijn gehaktbal (Door: Jop Esmijer, TNO). Beschikbaar via: <http://www.frankwatching.com/archive/2013/04/10/ober-er-zit-big-data-in-mijn-gehaktbal/>

### De ultieme hackers droom

Geen enkel systeem is ooit compleet veilig. De technologie rondom spam en cyberaanvallen heeft laten zien dat mensen met minder idealistische bedoelingen ICT ook inzetten voor hun eigen gewin. Wat kunnen hackers aanrichten op het moment dat ze toegang verkrijgen tot digitale beslissingssystemen?

### Gebrekkig besef van data context

Het gevaar bestaat dat beschikbare data wordt behandeld als een afspiegeling van 'de waarheid', zonder de specifieke context en vooral ook de limitatie in het oog te houden. Welke data hebben we niet tot onze beschikking en wat betekent dat voor de conclusies die we uit de wel beschikbare data kunnen trekken? En welke aannames gebruiken we in een model? Die aannames geven de marges aan van het gebied waarover conclusies kunnen worden getrokken.

### Schijnzekerheid van grote getallen

Hoe meer datasets worden geïntegreerd, hoe complexer de correlaties en hoe lastiger het is om het echte signaal te onderscheiden van 'ruis'. Hoe groter de data variatie, hoe groter het risico dat allerlei niet-bestaande causale verbanden worden gelegd. Intuïtief denken we dat veel data veel zekerheid verschaft. Echter, als men de kwaliteit van de data, analyse en interpretatie uit het oog verliest is dit slechts een schijnzekerheid.



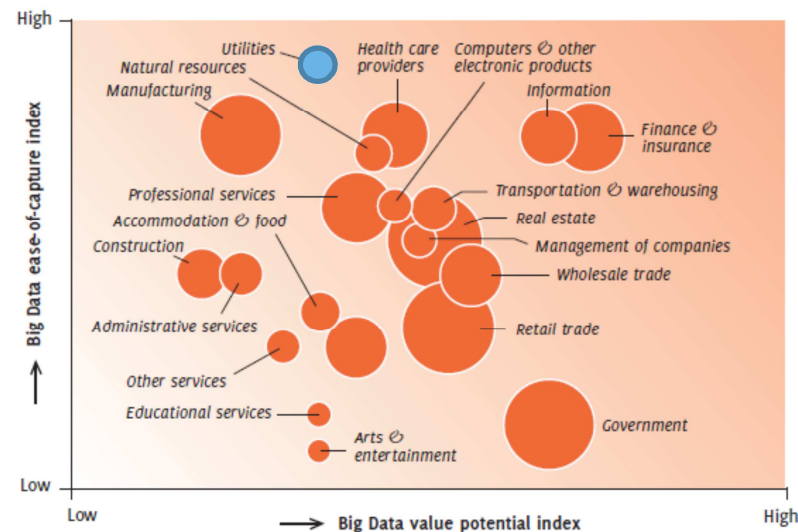
## Relevantie

Het Big Data veld is in ontwikkeling; best practices zijn nog schaars – zeker specifiek in de watersector – en op dit moment kan hooguit inspiratie worden opgedaan uit *'emerging next practices'*. Big Data technologie breekt door en experimenteren met Big Data gaat daarmee tot de mogelijkheden behoren. Verwacht wordt dat besparingen en efficiencylagen in het Big Data vooruitzicht liggen en steeds meer organisaties in de watersector krijgen het gevoel 'iets te moeten' met de Big Data hype. Echter, Big Data toepassingen zijn een middel en niet een doel op zich. Bij onoplettendheid kan de Big Data bezetenheid resulteren in oogkleppen die maken dat alles wat niet in digitale data gevat kan worden buiten de aandacht raakt, zinloze nietszeggende statistische relaties worden uitgewezen, en onterecht wordt gegeneraliseerd door de schijnzekerheid die enorme datavolumes met zich meebrengen.

Ook binnen de watersector wordt meer en meer data verzameld. Voornamelijk is dat gestructureerde data (zoals waterkwaliteit, -kwantiteit, -druk, -temperatuur, -verbruik, etc.), maar ook aan ongestructureerde data wordt steeds meer aandacht besteed (social media berichten, visuele inspectiedata, etc.). Echter, wat achterblijft is het vertalen van data naar voor de watersector nuttige informatie, kennis en wijsheden.

## Waarde

In 2011 presenteerde McKinsey Global Institute het toonaangevende rapport 'Big data: The next frontier for innovation, competition, and productivity'<sup>5</sup> Zij deden onder andere onderzoek naar de waarde van Big Data in verschillende sectoren van de Amerikaanse economie. Figuur 3 geeft voor de onderzochte sectoren de relatieve omvang van het Big Data potentieel (horizontale as) tegenover het relatieve gemak van dataverzameling (verticale as).



Figuur 3.

De potentie van Big Data in verschillende sectoren van de Amerikaanse economie tegenover het gemak van dataverzameling in de betreffende sectoren. De grootte van de bollen geeft de relatieve bijdrage van de sector aan het BBP aan.

(Bron: VINT, 2013 – Naar: McKinsey Global Institute, 2011)

<sup>5</sup> McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition and productivity. Beschikbaar via: [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)



Voor de nutssector (waar ook de drinkwatervoorziening binnen valt) wordt het verzamelgemak zeer hoog ingeschat. De beschikbaarheid van procesdata is al relatief hoog dankzij het veelvuldige gebruik van sensoren in het productieproces, maar ook rond waterverbruik (o.a. slimme meters), asset beheer (o.a. visuele inspectietechnieken), klantervaringen (o.a. social media) wordt steeds meer data verzameld. Water professionals hebben vaak een technische opleiding gehad, resulterend in relatief hoog analytisch vermogen, en de neiging naar een data-driven mindset. Waterbedrijven willen 'iets' met Big Data en staan op het punt om de IT-intensiteit om die reden te verhogen.

De potentie van Big Data in de watersector concentreert zich rond procesoptimalisatie, asset management, en klantinteractie. De volgende Big Data voordelen zijn daarin te onderscheiden:

- M.b.t. winning en zuivering: detecteren van verontreinigingsbronnen, betere monitoring en control, efficiëntere afstemming tussen winning, zuivering, opslag, distributie en watervraag, optimalisering van productieproces.
- M.b.t. het asset management: effectievere monitoring van het distributienet, opsporen en voorspellen van leidingbreuken en lekkages, preventief (maar niet te vroeg) onderhoud plegen.

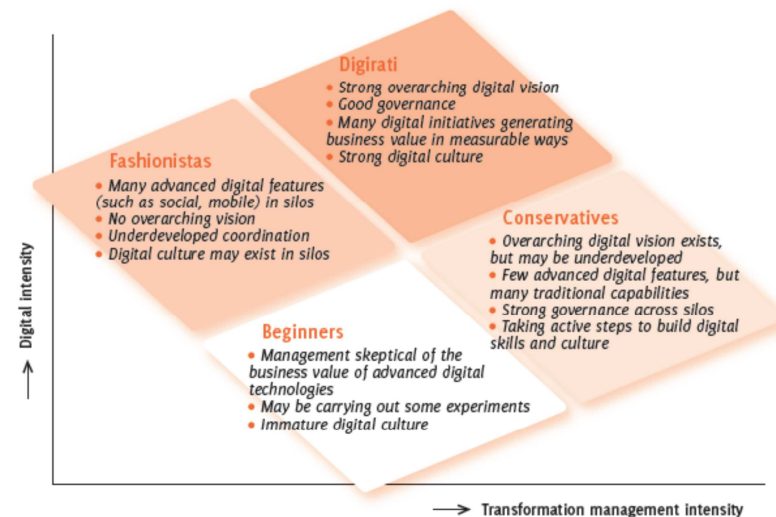
- M.b.t. klant & markt: monitoren en inspelen op wensen van de klant, monitoren van publieke opinie, imago-management, herkennen en benutten van actieve burgerinitiatieven.

### Houding

Gelet op de voorziene Big Data voordelen, welke houding en positie neem je in als waterbedrijf? Zet je in op digitalisering? Start je een verandertraject om meer data-gedreven te gaan werken? Gezamenlijk onderzoek van Capgemini en MIT<sup>6</sup> onderscheidt vier archetypen van Big Data

gebruikers, afhankelijk van de mate van digitalisering binnen een organisatie (beschikbare faciliteiten) en de mate van transformatief management binnen een organisatie (veranderingsgezindheid, 'meegaand met de tijd'). Deze archetypen gaan niet over wat goed of slecht is, maar is slechts een typering van hoe organisaties om kunnen gaan met de Big Data uitdagingen (figuur 4).

*Figuur 4.*  
*Archtypering van Big Data gebruikers, op basis van de mate van investering in digitale technologie (verticaal) en de mate van aanpasbaarheid op die digitale vernieuwingen (horizontaal).*



<sup>6</sup> Capgemini Consulting & MIT Center for Digital Business (2012). The Digital Advantage: How Digital Leaders Outperform Their Peers in Every Industry. Beschikbaar via: <http://ebooks.capgeminiconsulting.com/The-Digital-Advantage>





Beginners geloven niet erg in de kracht van digitale technologie en experimenteren hooguit wat. Echter, de echte oorzaak van het achterblijven op Big data gebied is dat hun ambitie om te veranderen ook niet heel hoog is. Conservatieven hebben die transformatieve ambitie en capaciteiten wel, maar hen visie op digitalisering is weinig enthousiast en de ICT toepassingen binnen een dergelijke organisatie zijn wel goed georganiseerd, maar niet heel geavanceerd. Digirati hebben zowel een sterke visie op digitale zaken en een zeer transformatieve vooruitstrevende cultuur, dit type organisatie heeft een helder doel voor ogen en biedt plaats voor veel digitale initiatieven die ook meetbaar waarde toevoegen. Fashionista's worden ook wel omschreven als meelopers. Dit type organisatie lijkt een hoop te ondernemen op ICT gebied. Echter, de initiatieven blijven te gefragmenteerd en dienen geen overkoepelend doel; hip, maar minder effectief.

Hoe gaat de drinkwatersector op dit moment met Big Data om? Wordt er geïnvesteerd in digitalisering voor dataverzameling en dataverwerking? Is er een heldere overkoepelende visie waarin Big Data toepassingen worden gekoppeld aan de organisatiedoelen? Welke houding wil je uiteindelijk aannemen? Blijf je als beginner experimenteren met de mogelijkheden zonder duidelijke richting? Of ga je mee met de hype en investeer je in allerlei moderne systemen?

### Samenwerking

Het Big Data-denken vraagt om samenwerking. Samenwerking met ICT bedrijven en softwareleveranciers om toegang te krijgen tot de technologie. Maar: Big Data gaat ook over het integreren van datasets. Daarmee is ook samenwerking nodig met leveranciers van externe data die wellicht relevant is ('correleert') met waterdata, zoals waterschappen (watersysteem), het KNMI (klimaatdata), en energiebedrijven (huishoudelijk energieverbruik). Hoe ga je als waterbedrijf zelf om met data- en informatie eigendom? Ben je bereid om eigen data 'open' te gooien zodat de koek gezamenlijk zo groot mogelijk gemaakt kan worden voordat hij verdeeld wordt?

### Voorbeelden

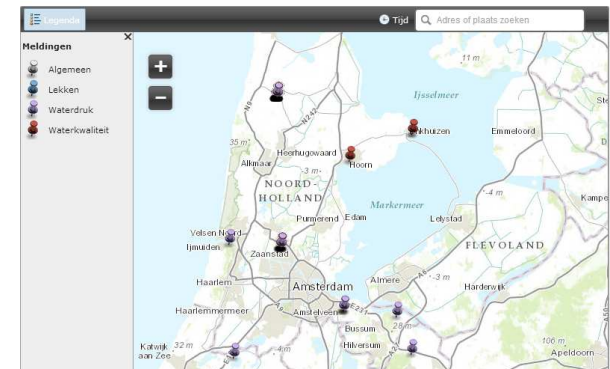
Voor de nodige inspiratie volgen hier drie voorbeelden van *emerging next practices* als het gaat over Big Water Data:

#### Human Sensor (PWN, TU Delft, Royal Haskoning DHV)

Zo snel mogelijk in kaart brengen waar lekken en verontreinigingen zich bevinden in je drinkwater-distributienet; dat is het doel van The Human Sensor. Daarnaast beoogt de toepassing de betrokkenheid van de drinkwaterklant te vergroten. Hiertoe worden

publiek toegankelijke data van Twitter geïntegreerd met bedrijfsdata rond klachten, onderhoud en storingen. Gezamenlijk wordt dit gekoppeld aan een stromingsmodel van het distributienet, dat vervolgens mogelijke locaties van het lek of de vervuilingbron berekent. Deze informatie wordt getoond op een plattegrond, welke beschikbaar wordt gesteld voor klanten en call-center medewerkers om op de hoogte te raken van wat er aan de hand is, maar ook aan onderhoudsmedewerkers die direct weten waar ze naartoe moeten om de calamiteit te verhelpen. Een mooi voorbeeld van een *emerging next practice* rond sociale ongestructureerde data.

Meer informatie: <http://www.pwn.nl/thehumansensor>



Figuur 5. Interface van The Human Sensor.  
(Bron: Website PWN)

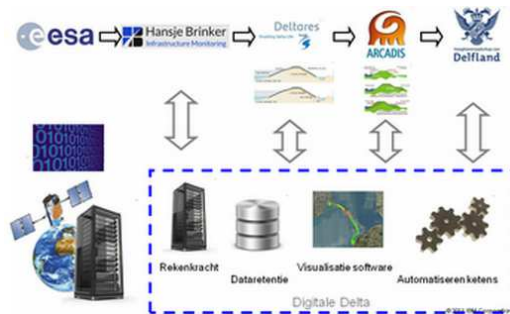


## Digitale Delta

(Rijkswaterstaat, HH Delfland, Deltares, TU Delft)

Onderzoeksprogramma Digitale Delta, gelanceerd in juni 2013, is gericht op het ontwikkelen van een open platform dat zoveel mogelijk data die relevant is voor waterbeheer (in de ruimste zin des woords) bevat en ontsluit. Het platform moet de diverse actoren in de watersector in staat stellen efficiënter te werken bij het uitoefenen van hun primaire taken. Onderzocht wordt hoe de grote variëteit aan water-gerelateerde data geïntegreerd kan worden in één gestandaardiseerd open data systeem, dat als basis kan dienen voor allerlei toepassingen die de data (voor zowel professionals als leken) snel en efficiënt kan omzetten naar relevante informatie.

Meer informatie: <http://www.digitaledelta.nl>



Figuur 6. De Digitale Delta combineert diverse databases.  
(Bron: Website Digitale Delta)

## Dijk-Data Service Centrum

(IJKdijk, Nelen & Schuurmans, Fugro)

Het Dijk Data Service Centrum is een platform voor de opslag van landelijke data die worden gemeten in en rond dijken en waterkeringen. Voorbeelden van data die kunnen worden opgeslagen zijn: hoogtemetingen, zettingen (in x,y,z-richting), (grond)waterstanden, bodemvocht, temperatuur, infrarood- en radarscans. Door de koppeling van de gegevens van meerdere waterschappen, bestaat de mogelijkheid om data van soortgelijke dijken in de tijd te vergelijken, en daarmee betere voorspellingen te kunnen maken in onderhoudsstappen die moeten worden genomen om calamiteiten af te wenden. Zo kan het systeem bijvoorbeeld een alarm afgeven als vooraf ingestelde waarden worden overschreden.

Meer informatie: <http://www.ddsc.nl>



Figuur 7. Het DDSC bevat en relateert landelijke dijkdata.  
(Bron: Website DDSC)

## Meer informatie

- VINT Sogeti Trendlab, Big Data onderzoeksnotities: 1) Helderheid creëren met Big Data (2012); 2) Big Social – Gedrag voorspellen met Big Social (2012); 3) Privacy, Technologie en de Wet (2013); 4) Uw Big Data potentie (2013). Beschikbaar via: <http://vint.sogeti.com/downloads/>
- VINT Sogeti Trendlab (2013). THINGS – Internet van businesskansen. Onderzoeksnotitie 1 van 4. Beschikbaar via: <http://vint.sogeti.com/downloads/>
- VINT Sogeti Trendlab (2013). Video: Big Data – The Future Recorded. Beschikbaar via: <http://vint.sogeti.com/videos/>