

Datasafari

Citation for published version (APA):

Helms, R. W. (2015). *Datasafari: Exploreren om te innoveren*. Open Universiteit Nederland.

Document status and date:

Published: 13/02/2015

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 02 Jul. 2022

Open Universiteit
www.ou.nl



PROF. DR. IR. REMKO HELMS

Datasafari: exploreren om te innoveren



Open Universiteit
www.ou.nl



prof. dr. ir. Remko W. Helms

Datasafari: exploreren om te innoveren

Open Universiteit
www.ou.nl



© Copyright Remko W. Helms, 2015

All rights reserved. No part of this publication may reproduced, stored, in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Ontwerp omslag: Vivian Rompelberg, afdeling Visuele Communicatie, Open Universiteit
Opmaak binnenwerk: Maria Wienbröker-Kampermann, Open Universiteit
Bureauredactie: John Arkenbout

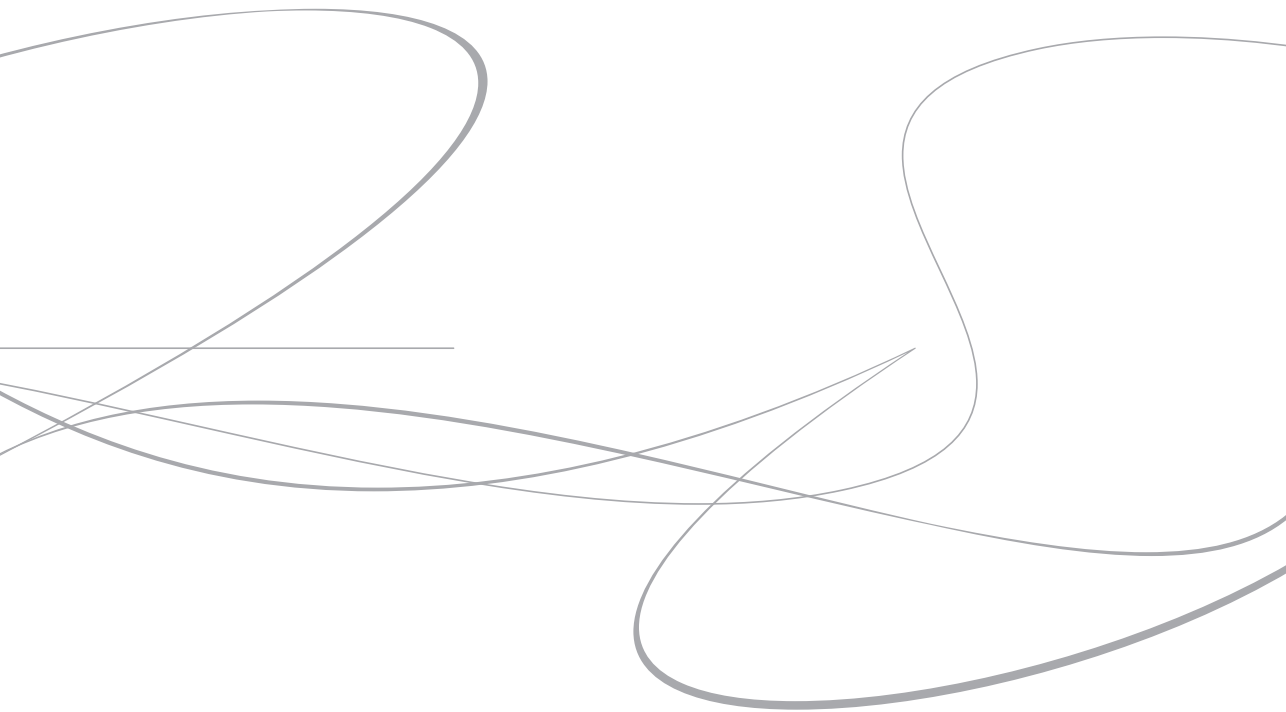
Printed in The Netherlands

Open Universiteit
www.ou.nl



Inhoud

- 1 Inleiding: data als grondstof voor de kenniseconomie 7
- 2 Big data: Volume, variëteit en snelheid 9
 - 2.1 Volume 9
 - 2.2 Variëteit 11
 - 2.3 Snelheid 14
- 3 Van data naar toegevoegde waarde 17
- 4 Technologie als aanjager 19
 - 4.1 Bigdata-infrastructuur 19
 - 4.2 Analytics 20
- 5 Nieuwe vaardigheden voor medewerkers 25
 - 5.1 Realisatie van bigdata-technologie 25
 - 5.2 Gebruik van bigdata-technologie 26
- 6 Organiseren van de datasafari 29
 - 6.1 Bigdata-strategie 29
 - 6.2 Datasafari als proces 30
 - 6.3 Organisatiestructuur 32
- 7 Conclusie 35
- Dankwoord 37
- Referenties 39
- Contactgegevens 43
- Bijlage: European e- Competence Framework 3.0 overview 45



Datasafari: exploreren om te innoveren

Rede

in verkorte vorm uitgesproken bij de openbare aanvaarding van het
ambt van hoogleraar Informatiekunde aan de Open Universiteit

op vrijdag 13 februari 2015

door prof. dr. ir. Remko W. Helms





Geachte rector, hooggeleerde en zeer geleerde collega's, dames en heren,

1 Inleiding: data als grondstof voor de kennis-economie

Nederland is voor zijn welvaart grotendeels afhankelijk van de kenniseconomie. Dit betekent dat het bedrijfsleven voor zijn concurrentiekracht afhankelijk is van de inzet van kennis (van Reijssen, Helms, Batenburg, & Foorthuis, 2014). Slimme inzet van kennis is noodzakelijk voor het genereren van innovaties, in zowel processen als producten, die leiden tot het verkrijgen van concurrentievoordeel. Mensen, of beter gezegd kenniswerkers, worden gezien als een belangrijke grondstof voor het komen tot deze innovaties. Zij beschikken namelijk over de kennis, ervaring en vaardigheden die voor deze innovaties nodig zijn. Maar er is een belangrijke grondstof voor innovatie bijgekomen en dat zijn data, of beter gezegd 'big data'. Zowel binnen als buiten de organisatie is een grote verscheidenheid aan databronnen aanwezig waarin mogelijk waardevolle kennis ligt verscholen. Ontginning van deze databronnen kan daarom leiden tot nieuwe innovaties. Een goed voorbeeld is UPS dat data verzamelt over de locatie en snelheid van zijn bezorgtrucks (Davenport, 2013). Het over langere tijd opslaan van deze data heeft UPS een geweldige schat aan data opgeleverd over gereden routes. Het geeft inzicht in hoe er gereden is en met welke snelheid. Door analyses uit te voeren op deze historische data heeft UPS de bezorgroutes kunnen optimaliseren. Een procesinnovatie die grote besparingen aan brandstof heeft opgeleverd voor het bedrijf.

Uit onderzoek van MIT blijkt dat bedrijven die big data toepassen 5 tot 6 procent beter scoren op productiviteit en winstgevendheid dan hun concurrenten (McAfee & Brynjolfsson, 2012). Het is daarom niet verwonderlijk dat big data veel aandacht krijgen van bedrijven, instellingen en de wetenschap. Elke maand zijn er wel congressen rondom deze thematiek en verschijnen er nieuwsberichten in de media met nieuwe toepassingen die de belofte van big data verder onderschrijven. Veel aandacht gaat niettemin uit naar de technologie die nodig is voor het opslaan, verwerken, analyseren en presenteren van big data. Maar hoe kun je als organisatie, overheid, multinational of midden- en kleinbedrijf nu op een zinvolle manier aan de slag met big data? Veelal lijkt de aanpak die bedrijven kiezen op een datasafari. Dat wil zeggen dat men gaat grasduinen in allerlei databronnen in de hoop daarin nuttige patronen en inzichten te ontdekken. Door de hoeveelheid en verscheidenheid aan data en databronnen is er grote kans dat je door de bomen het bos niet meer ziet en dat er kostbare tijd en geld wordt verspild aan een dergelijke datasafari.

De vraag rijst dan ook hoe je als organisatie op een verstandige manier met big data aan de slag gaat waarbij de datasafari uiteindelijk ook tot concrete innovaties leidt die het concurrentievermogen van de organisatie vergroten. Moet je daarvoor inderdaad, een leger specialisten aanstellen, zogenaamde 'data scientists', zoals vaak in artikelen wordt aangegeven (Davenport & Patil, 2012)? Of komt er meer kijken bij de toepassing van big data door organisaties? Deze typische informatiekundige vraagstelling, overeenkomstig mijn leerstoel aan de Open Universiteit, wil ik graag verder verkennen in deze lezing.

2 Big data: Volume, variëteit en snelheid

Hoewel de term big data inmiddels gemeengoed is geworden, wil ik toch eerst even stilstaan bij wat big data nu eigenlijk zijn. Er zijn drie kenmerken waardoor big data zich onderscheiden van gewone data, namelijk *volume*, *variëteit* en *snelheid*. Elk van deze kenmerken zal kort de revue passeren en ik zal met voorbeelden de potentie van elk kenmerk illustreren.

2.1 Volume

Het feit dat ze 'big' data worden genoemd geeft al aan dat het om de hoeveelheid, ofwel het volume, van de data gaat. In dat geval dringt zich de vraag op bij welke hoeveelheid we dan precies spreken van 'big' data. In sommige gevallen wordt er geprobeerd hiervoor een precies getal te geven. Maar elk cijfer dat er wordt opgeplakt zal binnen afzienbare tijd achterhaald zijn, gezien de snelle ontwikkelingen in de technologie (Watson, 2014). Om toch een indruk te geven van de hoeveelheid data waar het om gaat, zal ik een aantal voorbeelden geven ter illustratie. Maar voordat ik dat doe, is het handig om een referentiepunt te hebben waarbij we het dicht bij huis zoeken. De data die u privé opslaat, beslaan al gauw een aantal gigabytes of misschien zelfs een terabyte. Zoals in tabel 1 is te zien, zijn er voor een gigabyte aan data slechts 256 foto's nodig (van gemiddelde resolutie). Gezien het gemak waarmee we tegenwoordig foto's kunnen maken, hebt u al snel vele gigabytes voor uw foto's nodig. Als u dan ook nog eens enthousiast films digitaal opslaat dan komt zelfs de terabyte al in zicht.

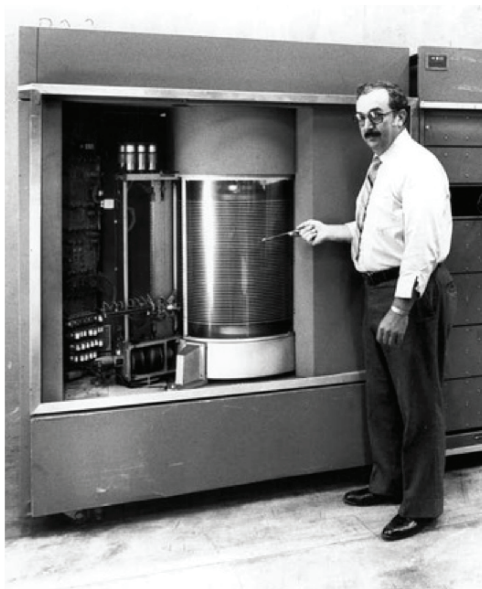
TABEL 1 Geheugenopslag en equivalent in pagina's tekst, foto's, DVD films en blu-ray films¹

<i>Eenheid geheugenopslag</i>	<i>Omrekenfactor</i>	<i>Genoeg voor opslag van</i>
1 megabyte (Mb)		873 pagina's tekst (van 1.200 karakters)
1 gigabyte (Gb)	1024 megabytes	894.784 pagina's tekst 256 foto's (van 4 Mb)
1 terabyte (Tb)	1024 gigabytes	916.259.689 pagina's tekst 262.144 foto's 233 DVD films (van 4,38 Gb) 40 blu-ray films (van 25 Gb)
1 petabyte (Pb)	1024 terabytes	938.249.922.368 pagina's tekst 268.435.456 foto's 239.400 DVD films 41.943 blu-ray films
1 exabyte (Eb)	1024 petabytes	960.767.920.505.705 pagina's tekst 274.877.906.944 foto's 245.146.535 DVD films 42.949.672 blu-ray films

¹ Getallen gebaseerd op <http://www.computerhope.com/issues/chspace.htm>



In de wetenschap en het bedrijfsleven gaat het, wanneer we spreken over big data, om een veelvoud hiervan. Neem bijvoorbeeld de deeltjesversneller in CERN waarmee deeltjes worden versneld en met elkaar in botsing worden gebracht. Met deze experimenten worden per jaar 13 petabytes aan data verzameld (Brumfiel, 2011). Bij het DNA Sequencing Project dat Barack Obama onlangs aankondigde gaat het zelfs om een veelvoud hiervan (Clarke & Begley, 2015). Het doel van dit project is om het DNA van 1 miljoen Amerikanen te verzamelen en hiermee doorbraken te realiseren in onder andere het onderzoek naar kanker. Om het genoom, dat bestaat uit 3 miljard baseparen, van één persoon op te slaan is 200 gigabytes nodig (ruwe data zoals het uit de DNA sequencer komt). Voor de database van 1 miljoen Amerikanen is dus grofweg 200 petabytes aan opslag nodig (Robinson, 2014). Maar ook in het bedrijfsleven neemt het datavolume toe. Neem bijvoorbeeld supermarktgigant Wal-Mart die met 2,2 miljoen werknemers wekelijks 250 miljoen klanten bedient. In 2008 traden ze toe tot de Petabyte Power Players club van leverancier Teradata. Op dat moment hadden ze 2,5 petabyte aan data betreffende aankopen van klanten en in 2014 is dit gestegen tot 30 petabyte (McMillon, 2014). Maar dit valt in het niet bij bedrijven waarbij data hun belangrijkste product zijn, bijvoorbeeld van Google of Facebook. Hoewel ze niet scheutig zijn met details over de omvang van hun datacenters zal het vermoedelijk in de orde van exabytes zijn (Wang, 2009). Dit staat in schril contrast met de opslagcapaciteit van de eerste harde schijf uit 1956 die 5 megabytes bedroeg. Google of Facebook zou 1012 (ofwel 1 biljoen) van deze harde schijven nodig hebben om hun data te kunnen opslaan. Gezien de omvang van deze schijven (zie figuur 1) zou je daar Duitsland en Frankrijk mee vol kunnen bouwen (bescheiden schatting van 1m2 per harde schijf).



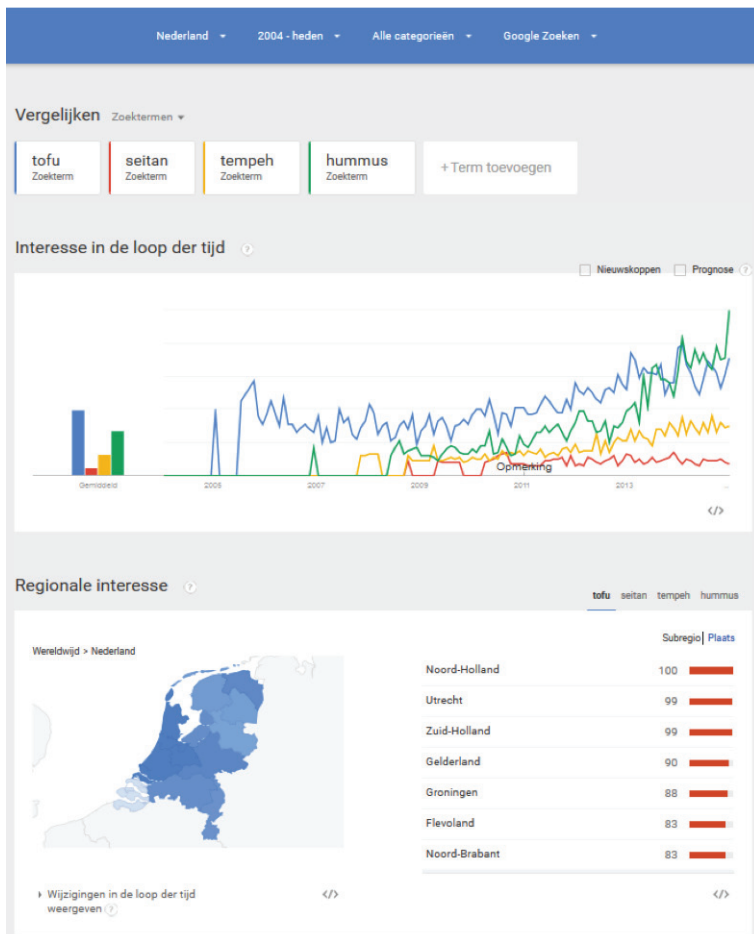
FIGUUR 1 IBM 305 RAMAC: harde schijf uit 1956 met opslagcapaciteit van 5 Mb

Het volume van de data, zoals die in de voorbeelden is gegeven, heeft drie belangrijke oorzaken. Allereerst is dat de schaal waarop data worden verzameld. De klanten van Wal-Mart zijn dusdanig talrijk dat dit meteen een enorme hoeveelheid aan data oplevert. Een tweede aspect is de nauwkeurigheid waarmee wordt gemeten. Dit is bijvoorbeeld het aantal verschillende aspecten dat één meting wordt vastgelegd bij het CERN. Ten slotte gaat het om de frequentie waarmee data worden vastgelegd. Zo kan UPS bijvoorbeeld elk half uur de locatie en snelheid meten maar ook iedere tiende seconde. Door het grote volume is het mogelijk om nieuwe analyses te doen, zoals bij het menselijk genoom, maar ook om nauwkeuriger analyses te doen, zoals de gevolgde route van een bezorgtruck. De potentiële analysemogelijkheden is wat het volume van de data zo interessant maakt.

2.2 Variëteit

Big data worden ook gekarakteriseerd door de variëteit aan data die beschikbaar zijn (Chen & Storey, 2012; Lycett, 2013; Watson, 2014). De variëteit betreft de verschillende soorten data waarover een bedrijf tegenwoordig de beschikking heeft (zie tabel 2 voor een overzicht). Laten we ter illustratie kijken naar de variëteit aan data die een keten van boekhandels uit de jaren '90 tot zijn beschikking had en deze vergelijken met de gegevens waarover een boekenketen zoals Bol.com en Amazon.com tegenwoordig beschikken. Een traditionele keten van boekhandels beschikte met name over gegevens uit een voorraad- en verkoopsysteem. Ze konden dus zien waar en hoeveel er werd besteld en tegen welke prijs, en welke titels in welke filialen goed werden verkocht. Wanneer de keten een klantenkaart hanteerde, dan wist men ook per klant wat die kocht en hoeveel. Online boekhandels beschikken daarentegen over een veel grotere variëteit aan data (McAfee & Brynjolfsson, 2012). Doordat ze de navigatiegeschiedenis van klanten op de site vastleggen, weten ze bijvoorbeeld welke andere boeken klanten hebben bekeken op de site en waar ze dus mogelijk in geïnteresseerd zijn. Ook weten ze wat klanten van boeken vinden door de boekbeoordelingen die op de site worden ingegeven en weer door anderen worden bekeken. Door deze twee te combineren kan inzichtelijk worden gemaakt in hoeverre klanten in hun koopgedrag worden beïnvloed door deze boekbeoordelingen. Daarnaast kunnen de gegevens over aankoopgeschiedenis en de navigatiegeschiedenis worden gebruikt om aanbiedingen te doen aan klanten welke vervolgens weer kunnen worden geanalyseerd of deze daadwerkelijk leiden tot een nieuwe aankoop. Naast gegevens die via de eigen informatiesystemen en website worden verzameld, kan Bol.com ook gebruik maken van andere wijzen om meer data over klanten te verzamelen. Via de Facebook-app Spin & Win die Bol.com in het verleden konden installeren, kon het toegang krijgen tot meer gedetailleerde informatie over zijn klanten. Zodra de app wordt geïnstalleerd verkrijgt Bol.com bijvoorbeeld toegang tot het Facebookprofiel en de vriendenlijst van de gebruiker.





FIGUUR 2 Trends in zoeken naar vegetarisch eten. Bron: <http://www.google.nl/trends>

Maar er zijn meer interessante data op Facebook beschikbaar voor een boekenverkoper. Welke boeken je hebt gelezen, welke je leuk vindt en welke je zou willen lezen bijvoorbeeld. Deze data kunnen worden gebruikt om gericht aanbiedingen te doen maar ook om uit te zoeken welke boeken via Bol.com worden aangeschaft en welke niet. Als de klant ook nog eens gebruik maakt van de mobiele Bol.com-app, kunnen ook de locatiegegevens van de klant worden uitgelezen. Al deze gegevens kunnen uiteindelijk worden gecombineerd zodat er een nauwkeurig profiel van klanten en hun transacties ontstaat, zoals wie, waar, wanneer en hoeveel.

Maar Bol.com zou nog verder kunnen gaan door ook gegevens van Google Trends te gebruiken². Google Trends laat zien wat favoriete zoektermen zijn, waar er deze op zoektermen wordt gezocht en wanneer (zie figuur 2). Hierdoor kunnen bepaalde trends worden gedetecteerd. Bijvoorbeeld naar welke buitenlandse auteurs in Nederland wordt gezocht, die hier nog niet zijn uitgebracht. En ook hier houdt de fantasie nog niet op want er zijn nog vele andere databronnen beschikbaar zoals historische gegevens over het weer en het verkeer. Door deze data te vergelijken met de eigen verkoopstatistieken kan worden nagegaan of het weer van invloed is op online verkopen en of het moment in de file wordt gebruikt om even snel via de mobiele Bol.com-app een boek aan te schaffen. Samenvattend, de variëteit zit in de verschillende kanalen en databronnen die worden aangeboord om data te vergaren over het proces, de klant en de markt. Deze combinatie maakt dat er nieuwe inzichten en verbanden kunnen worden blootgelegd die voorheen niet opgemerkt werden.

TABEL 2 Voorbeelden van databronnen

<i>Bedrijfsinformatiesystemen</i>	Data over producten, processen, transacties en klanten zoals worden vastgelegd in het productieplanningssysteem (ERP) en het klantbeheersysteem (CRM).
<i>Documentatie- en communicatiesystemen</i>	Data die liggen opgeslagen in tekstbestanden, zoals rapporten en e-mails, of html-bestanden zoals op het intranet en in wiki's. Naast de inhoud van deze berichten geeft het ook interessante data over wie met wie communiceert.
<i>Websites</i>	Data over interacties met de klant via een webshop of een online helpdesk. Zo is bijvoorbeeld bij te houden welke pagina's worden bekeken, in welke volgorde er wordt gebladerd en waarop wordt geklikt.
<i>Social media</i>	Data over individuen en wie hun vrienden zijn, wat ze leuk vinden en wat hen op dat moment bezig houdt (ofwel status updates).
<i>Mobile devices</i>	Data over de locatie en verplaatsingen van de gebruiker. Daarnaast kan er ook toegang worden gekregen tot andere data op deze devices zoals bijvoorbeeld de contactgegevens of het social-mediaprofiel.
<i>Smart sensors (Internet of Things)</i>	Data afkomstig van apparaten verbonden met het internet waarmee continu metingen kunnen worden gedaan aan een object of omgeving. Neem bijvoorbeeld de slimme thermostaten in huizen die met tijdintervallen het energieverbruik vastleggen.
<i>Open data</i>	Data die partijen gratis (of tegen een geringe vergoeding om de kosten te dekken) aan derden ter beschikking stellen. Met name overheden zijn actief bezig om gegevens van diverse registraties, bijvoorbeeld basisadministratie adressen en gebouwen, aan derden beschikbaar te stellen.

2 <http://www.google.nl/trends>

2.3 Snelheid

De derde karakteristiek van big data is de snelheid waarmee de data worden gecreëerd en verwerkt. De huidige technologie maakt het mogelijk om zeer snel data te verzamelen via een grote verscheidenheid aan informatiesystemen, (slimme) sensoren en mobiele devices. Deze data kunnen worden opgeslagen en later geanalyseerd maar steeds vaker worden de data real-time geanalyseerd en direct gebruikt bij het nemen van beslissingen. De snelheid waarmee data worden gecreëerd en geanalyseerd, kan worden geïllustreerd met de America's Cup. Tijdens de laatste editie van deze zeilwedstrijd in 2013, speelde de toepassing van big data een belangrijke rol (Fisher, 2013). Het schip van Team USA was voor deze race uitgerust met 300 sensoren die elke tiende seconde 3.000 verschillende meetgegevens verzamelden over de actuele prestaties van de boot. Deze data werden direct doorgestuurd naar een centrale computer waar deze real-time werden gecombineerd met gegevens over het weer en data uit de trainingen. Zo kon worden bekeken of de boot op dat moment optimaal presteerde gegeven de omstandigheden. De zeilers en schipper kregen deze resultaten ook te zien op mobiele devices om hun pols (zie figuur 3). Zo konden ze direct ingrijpen om de prestaties van de boot te verhogen en dus sneller te varen dan hun uitdager team Nieuw Zeeland.



FIGUUR 3 America's Cup (onder: smart device op arm van de skipper)
Bron: Gilles Martin-Raget/acea en CNN.com

Een ander voorbeeld is de opwekking van windenergie (Dawson, 2013). Hoeveel energie er op ieder moment door windmolens wordt geleverd, hangt af van de weersomstandigheden. Om deze variëteit in de levering van energie op te vangen dient er dus altijd voldoende stroom op conventionele wijze te worden opgewekt. Als er daarentegen nauwkeuriger kan worden voorspeld hoeveel stroom een windmolen in de komende uren gaat leveren dan hoeft er minder conventionele energie als back-up te worden opgewekt. Hierdoor gaat minder energie verloren en kan het aandeel windenergie in de totale energievoorziening worden geoptimaliseerd. IBM heeft hiervoor een oplossing ontwikkeld die continu de prestaties en stroomopwekking van de windturbines meet en deze combineert met weergegevens. Door deze data continu te verzamelen en real-time te analyseren kan de bijdrage van een windmolenpark aan het energienet nauwkeurig worden voorspeld.

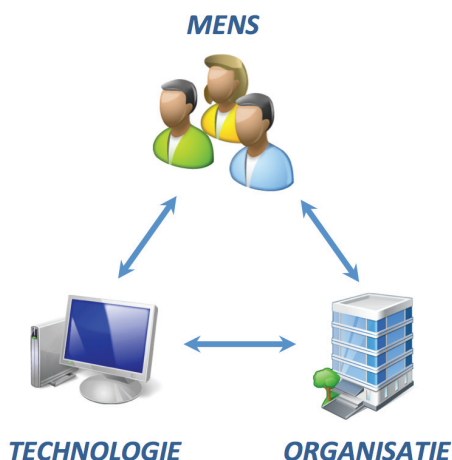
Bij de snelheid van big data draait het dus om de frequentie waarmee data worden verzameld en het feit dat ze real-time kunnen worden verwerkt tot stuurinformatie. Doordat er nagenoeg geen vertraging meer zit tussen het verzamelen van de data en het beschikbaar komen van analyseresultaten, kan er direct op actuele gebeurtenissen worden geanticipeerd en waar nodig worden bijgestuurd.





3 Van data naar toegevoegde waarde

De hiervoor genoemde voorbeelden maken duidelijk dat data een waardevolle grondstof zijn voor innovatie. Het mag ook duidelijk zijn dat slechts het bezitten van data geen garantie is voor succes. Of zoals collega René van Hezewijk het heel treffend in zijn afscheidsrede formuleerde: *data zijn geen kennis* (Hezewijk, 2015). Er is een bepaalde invalshoek, ofwel een theorie, nodig om een bepaalde hypothese te formuleren. Zonder deze invalshoek en hypothese is een analyse zinloos, want er ligt geen absolute kennis in data opgeslagen die met statistiek kan worden blootgelegd. Of de data nieuwe kennis opleveren die tot innovatie leidt, hangt dus af van met welke blik en welke vraag er naar de data wordt gekeken. Voor het omzetten van data naar kennis zijn dus mensen nodig om deze vertaalslag te kunnen maken maar er is ook technologie nodig om dit mogelijk te maken. Verder dient ook de organisatie de juiste randvoorwaarden hiervoor te scheppen.



FIGUUR 4 MOT-model

De samenhang tussen Mens, Organisatie en Technologie (MOT) komt ook tot uitdrukking in het zogenaamde MOT-model (zie figuur 4). Het is een veelgebruikt model binnen de Informatiekunde dat is gebaseerd op de systeembenadering. De oorsprong van het MOT-model is terug te voeren tot Leavitt's Diamond voor organisatieverandering dat uit de jaren '60 stamt (Leavitt, 1964). Later is dit model in allerlei verschillende varianten verschenen en wordt het in de Informatiekunde gebruikt om aan te geven dat technologie niet in isolatie moet worden beschouwd.

Om technologie succesvol toe te passen dient ook rekening te worden gehouden met de organisatie en de mensen die daarin werken. De organisatie verschaft namelijk de context waarbinnen de technologie zal worden gebruikt, terwijl de mensen de uiteindelijke gebruikers van de technologie zullen zijn. Indien de technologie de processen niet adequaat ondersteunt en de medewerkers niet met het systeem overweg kunnen, is de implementatie van de technologie gedoemd te mislukken. In het vervolg van mijn verhaal zal ik dit model gebruiken om aan te geven welke veranderingen er binnen technologie, mensen en organisatie nodig zijn om als organisatie succesvol data in innovaties te kunnen omzetten.

4 Technologie als aanjager

Technologie is een belangrijke aanjager voor big data. Er zijn daarbij twee belangrijke componenten te onderscheiden: de IT-infrastructuur voor de verwerking en opslag van data, en de tools die nodig zijn voor de analyse van data.

4.1 Bigdata-infrastructuur

De karakteristieken van big data stellen de technologie voor nieuwe uitdagingen, deze moet namelijk grotere hoeveelheden data aan kunnen. Voor een deel wordt deze uitdaging het hoofd geboden door de immer voortschrijdende ontwikkeling van processorkracht (wet van Moore) en geheugen dat steeds groter, sneller en goedkoper wordt. Maar er hebben ook revoluties plaatsgevonden in de wijze waarop data worden opgeslagen en verwerkt. Traditioneel worden data van bedrijven opgeslagen in datawarehouses die zijn gebaseerd op relationele databasemanagementsystemen (RDBMS). Hierin worden data op een gestructureerde, door het databaseschema vooraf gedefinieerde, wijze opgeslagen. De structuur kenmerkt zich door tabellen, bestaande uit rijen en kolommen, vergelijkbaar met de tabellen in een Excel-spreadsheet. Maar deze manier van data opslaan en verwerken loopt tegen zijn grenzen aan, in termen van schaalbaarheid en snelheid, wanneer het om big data gaat. Een eerste probleem is dat veel big data naast gestructureerde data ook ongestructureerde data omvatten. Voorbeelden van ongestructureerde data zijn tekst en foto's zoals die op Twitter, Facebook of Instagram worden geplaatst. Traditionele systemen zijn minder geschikt om ongestructureerde data op te slaan en dus zijn er nieuwe databasemanagementsystemen ontwikkeld zonder een vooraf gedefinieerd databaseschema, die beter met ongestructureerde data overweg kunnen. Een bekend voorbeeld daarvan is NoSQL.

Een tweede probleem is dat de hoeveelheid zo groot kan zijn dat het niet meer door een enkele server kan worden verwerkt of dat de snelheid te laag is. Met gespecialiseerde hardware kan aan deze beperkingen enigszins tegemoet worden gekomen, maar er zijn alternatieven ontwikkeld die goedkoper zijn. De aanjagers van deze ontwikkeling waren de grote zoekmachines (denk aan Google en Yahoo) en social media platforms (zoals Facebook en YouTube) die alternatieven zochten voor de opslag en verwerking van de grote hoeveelheden data die zij elke dag te verstouwen krijgen. Een van deze ontwikkelingen zijn systemen bestaande uit vele servers die data parallel kunnen verwerken. De meest bekende technologie, waar vele andere oplossingen op zijn gebaseerd, is Apache Hadoop.



Ook de analyse van grote databestanden vergt nieuwe ontwikkelingen. Voor de analyse is het namelijk nodig om gegevens uit de database op te halen en te bewerken. In traditionele oplossingen betekent dit dat er continu van en naar de harddisk moet worden geschreven. Samen met de hoeveelheid geheugen en rekenkracht bepaalt dit hoe snel de analyse kan worden uitgevoerd. Bij kleine databestanden zijn de reactietijden nauwelijks merkbaar maar bij bigdata-bestanden neemt deze reactietijd dermate toe dat er naar oplossingen is gezocht. Twee voorbeelden zijn in-database analytics en in-memory analytics. Bij in-database analytics wordt het verkeer tussen de applicatie en de database beperkt door de analyses als het ware in de database zelf uit te voeren. Bij in-memory analytics worden de data compleet in het geheugen geladen, ook op deze wijze wordt het relatief langzame schrijven van en naar een harddisk voorkomen.

Tot slot wil ik nog cloud computing noemen als belangrijke technologietrend in relatie tot big data. Cloud computing betekent dat de hardware en software die benodigd zijn voor big data, via het internet beschikbaar worden gesteld. Hierdoor hoeft een organisatie de benodigde technologische infrastructuur niet langer zelf in huis te hebben en te onderhouden. De organisatie betaalt een vast bedrag per maand of naar rato van gebruik, en hoeft dus niet zelf in mensen te investeren die de technologie kennen en onderhouden. Een ander voordeel is dat deze oplossingen schaalbaar zijn; wanneer er op een gegeven moment meer opslag- of reken capaciteit nodig is dan kan deze eenvoudig worden bijgeschakeld door de provider van de clouddienst. Op deze wijze is big data dus verworpen tot een dienst die je kan inkopen: Big-data-as-a-service. Dit is met name voor het midden- en kleinbedrijf interessant omdat die vaak beperkte middelen hebben om hierin te investeren. Naast het beschikbaar stellen van de infrastructuur zouden technologieaanbieders ook open data kunnen aanbieden in hun platform zodat bedrijven die data in hun analyses kunnen betrekken.

4.2 Analytics

Hoewel het beschikbaar hebben van een bigdata-infrastructuur belangrijk is, draait het uiteindelijk om de analyses die op deze data worden uitgevoerd. De analyses moeten leiden tot nieuwe inzichten die kunnen worden vertaald naar procesverbeteringen en productinnovaties. Het gebruiken van data om analyses uit te voeren is echter geen trend van de laatste tijd. Organisaties doen dit al vanaf het moment dat zij informatie opslaan over hun processen, producten en klanten. Ook toen deze dataopslag nog voornamelijk op papier plaatsvond. Met de intrede van computertechnologie heeft dit niettemin een steeds grotere vlucht genomen. Zo werden in de jaren '80 de zogenaamde managementinformatiesystemen en beslissingsondersteunende systemen geïntroduceerd. Deze systemen waren geschikt voor het creëren van managementrapportages, het uitvoeren van drill-downanalyses en het ondersteunen van groepsprocessen voor besluitvorming. In de jaren '90 ontwikkelde zich dit verder

in Business Intelligence gebaseerd op datawarehousingtechnologie en deed data mining zijn intrede. De huidige generatie technieken en tools ter ondersteuning van analyse en besluitvorming worden aangeduid met Analytics (Sharda, Delen, & Turban, 2013; Watson, 2014). De belangrijkste ontwikkeling daarbij is dat de analysetechnieken in de loop der tijd steeds geavanceerder zijn geworden waarbij er gebruik gemaakt wordt van statistiek, algoritmen, wiskundige modellen en artificiële intelligentie.

“We don’t have better algorithms. We just have more data.”

Peter Norvig (director at Google)

Over het algemeen wordt er onderscheid gemaakt naar drie verschillende soorten analytics, te weten: descriptive, predictive en prescriptive analytics (Chen & Storey, 2012; Wixom et al., 2011)³. Bij descriptive analytics gaat het om eenvoudige overzichten en mogelijkheden om selecties te maken. Denk bijvoorbeeld aan het aantal verkochte vluchten per route en per maand. Daar is makkelijk uit af te leiden wat populaire routes zijn en hoe de afzet per seizoen fluctueert. Daarbij gaat het vooral om historische data, dus er wordt vooral gekeken naar wat heeft plaatsgevonden en waarom dat heeft plaatsgevonden. Dit soort analyses waren al mogelijk met de managementinformatiesystemen uit de jaren '80.

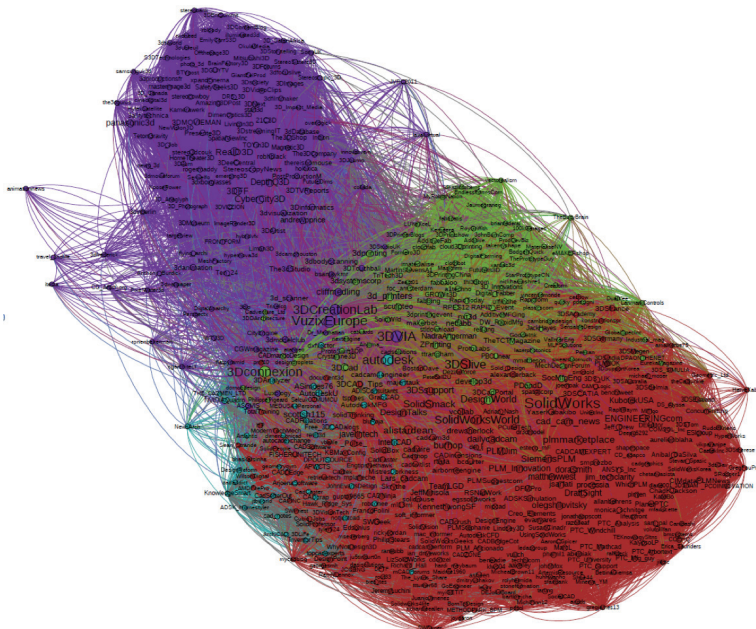
Maar we willen niet alleen in onze achteruitkijkspiegel zien wat er is gebeurd, we willen ook inzicht in wat ons in de nabije toekomst te wachten staat. De technieken en tools hiervoor worden aangeduid met predictive analytics en voorbeelden zijn regressieanalyse, machine learning en neurale netwerken (Watson, 2014). Het doel van deze analysetechniek is om modellen te ontwikkelen waarmee voorspellingen kunnen worden gedaan. Zijn er bijvoorbeeld patronen te ontdekken in de persoonlijke kenmerken en het gedrag van de klant waarmee valt te voorspellen wanneer deze een volgende vlucht gaat boeken? Of in welke mate heeft de vliegtuigmaatschappij last van een negatief sentiment in sociale media als het gaat om het aantal verkochte vliegtuigstoelen in de komende weken?

De laatste categorie analytics betreft prescriptive analytics en geeft suggesties over hoe optimale prestaties kunnen worden gehaald. Hiervoor worden technieken voor wiskundige optimalisatie toegepast, een voorbeeld is lineair programmeren. Een bekend voorbeeld waarbij optimalisatie wordt toegepast is het bepalen van de verkoopprijs voor een vliegstoel of een hotelkamer. Zoals u vast al eens hebt meegemaakt, heeft de persoon naast u in het vliegtuig een geheel andere prijs betaald. Aan de hand van een aantal parameters wordt namelijk een inschatting gemaakt hoeveel u zou willen betalen. Zo wordt onder andere naar uw historie van boekingen gekeken maar ook bijvoorbeeld naar uw surfgedrag. Surf dus anoniem als u de analytics te slim af wil zijn en een goede deal wilt voor uw vliegticket.

3 Een ander aardig overzicht dat uitgaat van 8 niveaus van analyse is te vinden op:
http://www.sas.com/news/sascom/analytics_levels.pdf

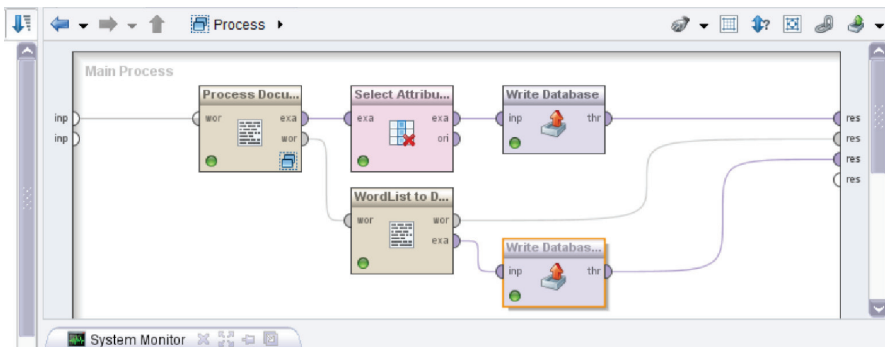


De analysetechnieken die worden toegepast, kunnen ook verschillen per databron. Er wordt dan ook wel gesproken van data analytics, text analytics, web analytics, network analytics en mobile analytics om het gebruik van analysetechnieken voor deze verschillende databronnen aan te duiden (Chen & Storey, 2012). Network analytics richt zich bijvoorbeeld op de data die op sociale-mediaplatforms voorhanden zijn en via API's, vaak in beperkte mate of tegen betaling, toegankelijk zijn voor iedereen met enige programmeervaardigheden. Een uniek aspect van de beschikbare data is de relaties tussen personen die op het social media platform actief zijn. Zo kan een bedrijf zien wie de beïnvloeders zijn binnen het netwerk en of er een bepaalde clustering is van volgers die duidt op een segmentatie in klantengroepen (Helms & Werder, 2013). In deze tak van analytics wordt dan ook veel gebruikgemaakt van socialenetworkanalyse om de structuur van de netwerk relaties in kaart te brengen. Hierop ben ik zelf ook actief geweest met afstudeerders en mijn promovendus Jurriaan van Reijssen. In een van deze onderzoeken zijn we bijvoorbeeld meer dan 200 miljoen profielen op Twitter langsgeslagen om data te verzamelen over de volgers van de grootste softwarebedrijven van Europa. In figuur 5 staat een voorbeeld van het netwerk van Twittervolgers van een van deze bedrijven. Om het overzichtelijk te houden zijn de volgers die zelf minder dan 100 volgers hebben daarin weggelaten. Een dergelijke datavisualisatie illustreert ook waarom data op zichzelf geen kennis zijn. Zonder verder achtergrondinformatie is het moeilijk om te doorgronden wat het plaatje laat zien en wat de door de kleuren in het plaatje gesuggereerde clustering inhoudt. Omdat er met analysetools eenvoudig allerlei plaatjes kunnen worden gegenereerd ligt het gevaar van verkeerde interpretaties op de loer. Er is dus kennis en kunde vereist bij de toepassing van dergelijke tools.



FIGUUR 5 Twitter community Dassault (visualisatie d.m.v. Gephi)

Er zijn veel verschillende tools beschikbaar die de verschillende soorten analytics ondersteunen. Ook de bekende statistische software zoals SPSS en R kan tot deze tools worden gerekend. De tools die beschikbaar zijn verschillen enorm in de analysetechnieken die worden ondersteund, maar ook in de mogelijkheden voor bijvoorbeeld datavisualisatie, data-import en -export, en de ondersteunde databases. Een andere onderscheidende factor is of de tools ook daadwerkelijk grote datasets, ofwel big data, aan kunnen. Hoewel de analysetechnieken kennis van zaken vragen, maken veel leveranciers het gebruik van de tools zo eenvoudig mogelijk. Daarmee komen analysetools ook voor minder technologisch onderlegde gebruikers binnen bereik. Twee voorbeelden van gebruikersvriendelijke tools zijn RapidMiner (rapidminer.com) en Orange (orange.biolab.si). Met behulp van drag-en-dropfunctionaliteit in deze tools is het mogelijk om standaard analyse-elementen te kiezen en zo je eigen analyseworkflow te bouwen. Zie figuur 6 voor een voorbeeld van zo'n workflow dat elementen bevat waarmee text mining kan worden gedaan op een IMAP e-mailaccount. De koppeling tussen de elementen geeft aan in welke volgorde ze zullen worden uitgevoerd. Uiteraard ben je met dit soort tools niet zo flexibel als het R waarmee de gebruiker zelf scripts kan schrijven om analyses op maat te maken.



FIGUUR 6 Text mining workflow in RapidMiner.

Bron: <http://www.completebusinessanalytics.com>



5 Nieuwe vaardigheden voor medewerkers

Uit het voorgaande hoofdstuk is duidelijk geworden dat de technologie een belangrijke aanjager is voor big data. In dit hoofdstuk richten we ons op de mensen die nodig zijn voor een succesvolle realisatie en gebruik van bigdata-technologie.

“Think of big data as an epic wave gathering now, starting to crest. If you want to catch it, you need people who can surf.” (Davenport & Patil, 2012)

5.1 Realisatie van bigdata-technologie

De realisatie betreft de bouw van de bigdata-infrastructuur die nodig is voor big data analytics. Binnen organisaties is de IT-afdeling hiervoor verantwoordelijk. Zij zullen de mankracht en kennis beschikbaar moeten hebben om de, veelal ingekochte, technologie in de organisatie te implementeren. Dit betekent dat medewerkers moeten worden getraind of nieuwe medewerkers aangetrokken die over de benodigde vaardigheden beschikken. Eventueel kan de IT-afdeling nog worden ondersteund door leveranciers of kan er voor een cloudplossing worden gekozen waardoor de IT-infrastructuur grotendeels is uitbesteed.

De introductie van bigdata-technologie heeft geresulteerd in een aantal nieuwe functies binnen de IT-afdeling. Hieronder staan deze nieuwe functies weergegeven met tussen haakjes het gemiddelde jaarsalaris (Hein, 2014). Hoewel het nieuwe functies betreft, gaat het volgens de branche niet om radicaal nieuwe functies. In de laatste versie van het European e-Competence Framework (versie 3.0)⁴, en de 23 IT-profielen die hierop zijn gebaseerd, zijn namelijk maar heel beperkt wijzigingen aangebracht in relatie tot big data. Het zijn eerder de skills van bestaande functies die de veranderingen in de technologie volgen.

- ETL developers (\$110-130k)
- Hadoop developers (\$150-175k)
- Visualization developers (\$150-175k)
- OLAP developers (\$98-116k)
- Data warehouse appliance specialist (\$98-124k)
- Information Architects (\$114-136k)

Door de relatief snelle opkomst van big data is er een tekort aan deze functies, dit komt ook tot uitdrukking in de jaarsalarissen voor deze functies (ter vergelijking het gemiddelde inkomen in de US ligt op \$43.041 voor 2013⁵). Dit biedt dus goede perspectieven voor studenten die een technische IT-opleiding volgen. Een baan lijkt haast gegarandeerd bij afstuderen, iets wat zeker niet het geval is bij minder technische opleidingen.

⁴ <http://www.ecompetences.eu/>

⁵ Volgens Social Security bureau: <http://www.ssa.gov/oact/cola/central.html>



5.2 Gebruik van bigdata-technologie

Naast de genoemde technische functies is ook veel belangstelling voor een andere nieuwe functie, namelijk die van Data Scientist. Data scientists richten zich op de toepassing van bigdata-technologie en het accent ligt bij deze functie daarom op het uitvoeren van analyses. De functie van data scientist krijgt veel aandacht in de media en de literatuur, en is al de meest sexy job van de 21e eeuw genoemd (Davenport & Patil, 2012).

Het McKinsey Global Institute schat het tekort aan data scientists voor 2018 op 140 à 190 duizend, en dat betreft dan alleen de Verenigde Staten (Manyika et al., 2011). Om aan deze vraag te voldoen zijn er al veel nieuwe opleidingen opgezet om deze data scientists op te leiden (Wixom et al., 2011). Ook bij de Open Universiteit wordt op dit moment aan een aantal (deeltijd)opleidingen gewerkt met partners binnen het kenniscentrum Business Intelligence & Smart Services dat gekoppeld is aan de Smart Services Hub in de regio Heerlen. De eerste deeltijdopleiding, gericht op Business Intelligence en Smart Services, dient per 1 januari 2016 van start te gaan. Vanwege de schaarste aan data scientists zijn de salarisvooruitzichten voor hen voorlopig goed. Maar hierdoor ligt ook het gevaar van functie-inflatie op de loer. Vanwege de goede salarisvooruitzichten noemen veel mensen, die maar enigszins data-analysevaardigheden hebben, zichzelf data scientist. In LinkedIn geeft een zoekopdracht op de term data scientist in het profiel van een gebruiker inmiddels 244 duizend resultaten⁶.

De data scientist is direct verantwoordelijk voor het vertalen van big data in concurrentievoordeel voor de organisatie. Hiervoor dient een data scientist over een aantal specifieke vaardigheden te bezitten die onder andere zijn beschreven door Hammerbacher (2009) en Davenport & Patil (2012):

– *Nieuwsgierig en creatief*: Om nieuwe ontdekkingen te kunnen doen dienen er continu nieuwe vragen te worden opgeworpen. Hierbij is het de kunst om verbanden te zien waar door anderen nog niet aan is gedacht. Deze vragen dienen vervolgens te worden getoetst aan de beschikbare data dan wel nieuwe databronnen te worden aangeboden om toetsing mogelijk te maken.

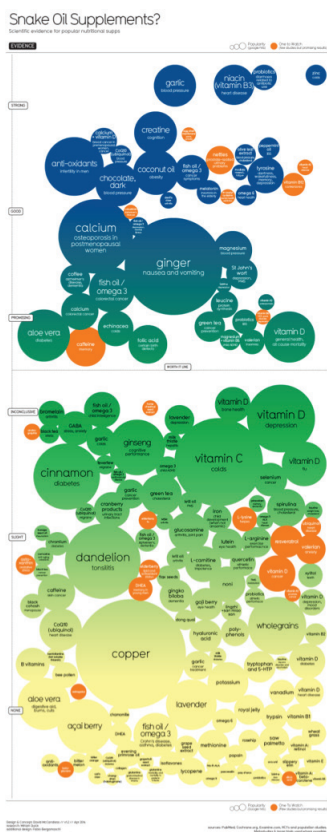
– *Domeinkennis*: Kennis van het toepassingsdomein is nodig om in te kunnen schatten welke inzichten wel van waarde zijn en welke niet. Zo zit de markt van energielevering anders in elkaar dan die van diepvriesmaaltijden; hoe je concurrentievoordeel kan behalen is dus verschillend in die markten. Daarnaast zijn er bedrijfsfuncties, zoals verkoop, marketing, financiën, productie, inkoop en ontwikkeling, die elk hun eigen domeinkennis vereisen.

– *Programmeervaardigheden*: Vaak zijn de data niet binnen één systeem beschikbaar en dienen deze vanuit verschillende systemen te worden gecombineerd. Bijvoorbeeld verkoopcijfers uit een verkoopsysteem en tekstberichten van sociale media. De data scientist moet dan code kunnen schrijven die nodig is om de data binnen te halen maar ook om ze vervolgens naar het juiste formaat over te zetten en te combineren.

⁶ Zoekopdracht uitgevoerd op 10 januari 2015 met als filter "People".

- *Analyticsvaardigheden*: Wanneer de data beschikbaar zijn, kunnen er analyses op worden uitgevoerd om verbanden en patronen te ontdekken. Op basis hiervan kunnen dan bijvoorbeeld geavanceerde voorspellingsmodellen worden ontwikkeld. Dit vereist kennis van statistiek, wiskunde en algoritmen.
- *Sociale en datavisualisatievaardigheden*: De verworven inzichten moeten kunnen worden vertaald in een advies aan de organisatie. Allereerst moeten daarvoor de getallen uit de analyse naar begrijpelijke taal worden omgezet. Hierbij kan datavisualisatie een krachtig hulpmiddel zijn (zie voorbeeld in figuur 7). Vervolgens dient het management hiermee te worden overtuigd wat sociale vaardigheden vereist.

Een persoon die de genoemde vaardigheden combineert klinkt als het schaap met de vijf poten. Het is een combinatie van studenten wiskunde, informatica, informatiekunde, economie en bedrijfskunde. Bovendien wordt alle verantwoordelijkheid om met innovaties te komen, bij één specifieke functie binnen de organisatie belegd. Dit vertoont overeenkomsten met specialisten bij banken die complexe financiële producten en systemen ontwikkelen die alleen zij nog begrijpen.

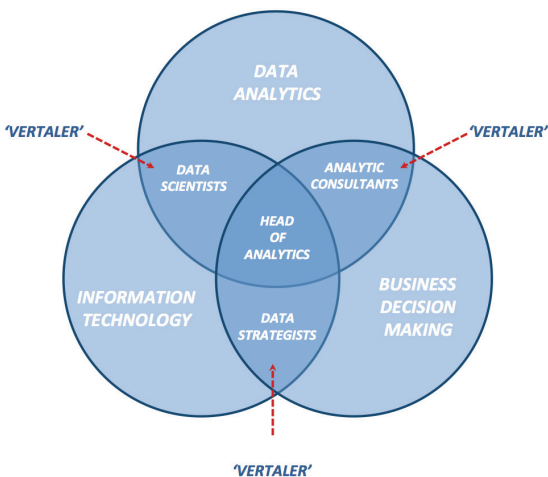


FIGUUR 7 Wetenschappelijk bewijs voor populaire voedingssupplementen.
Bron: <http://www.informationisbeautiful.net>

Dit is in mijn ogen geen wenselijke situatie en daarnaast is het de vraag of alle vaardigheden wel in één persoon verenigd kunnen worden. Het ligt meer voor de hand dat de data scientist samenwerkt met managers die van nature meer kennis hebben van de business en daardoor beter de link kunnen leggen met de bedrijfsdoelstellingen. Deze managers zouden data savvy moeten zijn (Manyika et al., 2011), zodat ze begrijpen wat de mogelijkheden zijn van analytics, en de datasafari de juiste richting kunnen geven door het stellen van de juiste vragen. Dit betekent dat er ook bij de alfa- en gamma-opleidingen aandacht moet komen voor data analytics, en dan vooral op de toepassing ervan, zodat ze gesprekspartner kunnen zijn van de data scientist.

Anderzijds zouden data scientists moeten samenwerken met de IT-afdeling omdat zij kennis hebben van het samenbrengen en toegankelijk maken van data. De data scientist zou zich dan vooral kunnen focussen op het uitvoeren van de data-analyse. In lijn met deze gedachte is er geopperd dat er 'vertalers' nodig zijn die een brug kunnen slaan tussen de gebieden: IT, Analytics en Business (Ariker, Breuer, & McGuire, 2014). Dus niet de data scientist als alleskunner maar een samenwerking tussen verschillende disciplines met daartussen vertalers om de disciplines nader tot elkaar te brengen (zie figuur 8). De vertaalrollen zijn informatiekundigen op het lijf geschreven. Het is dan ook belangrijk dat het vak data analytics zijn intrede doet in informatiekunde-opleidingen in Nederland.

Bij het volwassen worden van big data en analytics binnen bedrijven zullen de rollen en hun verantwoordelijkheden zich verder uitkristalliseren. Mogelijk dat dit leidt tot verdergaande aanpassingen in een volgende versie het European e-Competence Framework. Onderzoek van Nederland ICT, uitgevoerd door Dialogic, naar de Nederlandse ICT-markt ziet namelijk al een trend van functies die niet goed onder te brengen zijn binnen het framework (Gillebaard, Jager, Velde, Steur, & Vankan, 2014). Het onderwijsveld dient deze ontwikkelingen te volgen en indien nodig in lijn hiermee de inhoud van de curricula bij te stellen.



FIGUUR 8 Verschillende rollen voor data analytics. Bron: (Ariker, Breuer, & McGuire, 2014)

6 Organiseren van de datasafari

Het derde en laatste element uit het model is: organisatie. Hier draait het om het *richten* en *inrichten* van de organisatie zodat met data toegevoegde waarde kan worden gecreëerd. De belangrijkste aspecten die hier zullen worden behandeld zijn de strategie, het proces en de organisatiestructuur.

6.1 Bigdata-strategie

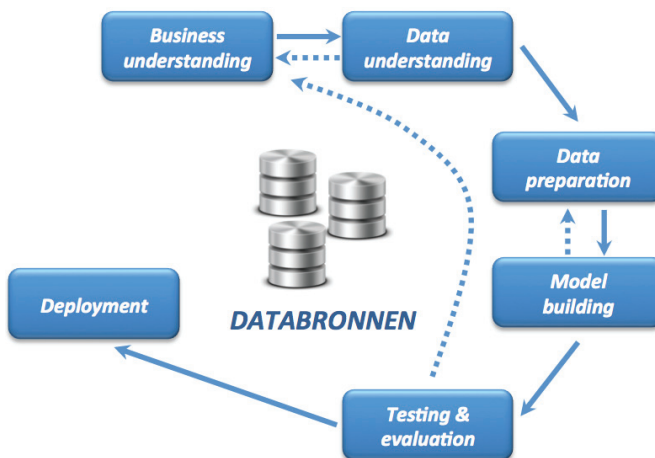
In mijn colleges Strategisch Management van Organisaties en ICT vertel ik studenten hoe organisaties concurrentievoordeel kunnen behalen met IT. De eerste voorbeelden daarvan stammen al uit de jaren '70 (Ward & Peppard, 2002). Een veel genoemd voorbeeld, dat ik ook in mijn colleges aanhaal, is het SABRE-reserveringssysteem van American Airlines voor de reservering van vliegtuigstoelen. Vanwege de impact van IT is het een steeds belangrijkere rol gaan vervullen in het strategisch planningsproces van organisaties. Strategische afstemming tussen de IT- en organisatiestrategie, ofwel business-IT alignment (Henderson & Venkatraman, 1993), staat dan ook hoog op de IT-agenda van veel organisaties. Dat big data en analytics tot concurrentievoordeel kunnen leiden is wat dat betreft niets nieuws. Sterker nog, onderzoek heeft al aangetoond dat het concurrentievoordeel kan opleveren (Lavalley, Lesser, Schockley, Hopkins, & Kruschwitz, 2011; McAfee & Brynjolfsson, 2012). Big data en analytics zijn daarom opties die meegenomen dienen te worden in het strategisch planningsproces. Daarbij is de technologie vooral een gegeven en dienen organisaties vooral met creatieve oplossingen te komen hoe ze deze kan toepassen.

Voor organisaties die nog geen gebruik maken van big data en analytics, is het interessant om te weten hoe andere organisaties dit toepassen. Dit kan worden bepaald door het bedrijfsmodel van deze organisaties te analyseren (Chesbrough & Rosenbloom, 2002). Voor het analyseren van een bedrijfsmodel wordt een raamwerk gebruikt dat bestaat uit verschillende bouwstenen, bijvoorbeeld bediende klantsegmenten, partners en kostenstructuur (Osterwalder, Pigneur, Smith, & Movement, 2010). Door middel van het raamwerk kan het bedrijfsmodel voor een specifieke organisatie in kaart worden gebracht. Bedrijfsmodellen die zich toespitsen op big data en analytics worden datagedreven bedrijfsmodellen genoemd. Het raamwerk voor deze modellen bevat bouwstenen zoals de gebruikte databronnen en de uitgevoerde data-activiteiten (Hartmann, Zaki, Feldmann, & Neely, 2014). Aan de hand van het raamwerk kunnen organisaties hun eigen bigdata-strategie vormgeven. Met verder onderzoek naar datagedreven bedrijfsmodellen zou ik generieke bedrijfsmodellen willen opstellen. Deze kunnen dienen als voorbeeld en inspiratie voor andere organisaties, bijvoorbeeld voor het midden- en kleinbedrijf die de stap naar big data en analytics nog grotendeels moeten maken.



6.2 Datasafari als proces

Of de gekozen strategie ook wordt waargemaakt wordt bepaald door de uitvoering ervan: het datasafariproces. In de literatuur zijn verschillende procesmodellen hiervoor voorgesteld, maar de mate van detail van deze modellen varieert sterk. Zo beschrijft het Business Analytics Succes Model (BASM) vooral op hoofdlijnen de stappen die nodig zijn, te weten: analyse, inzicht, besluit en actie (Seddon, Dod, & Constantinidis, 2012). Een meer gedetailleerd model is het Business Analytics Innovation Model (BAIP) dat beschrijft welke vijf stappen nodig zijn om tot innovatie te komen met behulp van analytics (Shanks & Bekmamedova, 2013). De stappen zijn rudimentair uitgewerkt en geven aan welke activiteiten worden uitgevoerd en wie er bij betrokken zijn. Het CRISP-DM-model, dat al in de jaren '90 voor data mining is ontwikkeld, behoort tot de meer gedetailleerde modellen (*IBM SPSS Modeler CRISP-DM Guide*, 2011). Het beschrijft in detail de stappen die worden uitgevoerd plus de activiteiten die daarbij horen. Tevens is er een tool op de markt beschikbaar die de verschillende stappen ondersteunt. In figuur 9 zijn de hoofdstappen uit het CRISP-DM-model weergegeven.



FIGUUR 9 CRISP-DM-model. Bron: Sharda, Delen, & Turban (2013)

De belangrijkste stappen hierin zijn in mijn ogen het inzicht krijgen in de organisatie-context (business understanding) en de beschikbare databronnen (data understanding). In deze stappen dienen namelijk de vragen te worden gesteld waarop de data-analyse antwoord moet gaan geven. Alleen wanneer de behoefte van de organisatie en de mogelijkheden van de data goed worden begrepen, zullen de juiste vragen kunnen worden gesteld.

“He who finds diamonds must grapple in mud and mire because diamonds are not found in polished stones. They are made.” ~ Henry B. Wils

De procesmodellen suggereren enigszins dat de stap van data naar innovatie deductief kan worden afgeleid. In de praktijk is het echter in belangrijke mate een inductief proces waarbij creativiteit en analytisch vermogen een belangrijke rol spelen. Om het creatieve proces te stimuleren kunnen innovatieworkshops worden toegepast. Hiervoor kan bijvoorbeeld de Voort Innovatiemethode worden toegepast waar collega's bij managementwetenschappen onderzoek naar doen⁷. Dergelijke methodes hebben als doel mensen aan te zetten tot alternatieve manieren van denken. Een aantal leveranciers van bigdata-technologie leveren dergelijke workshops als dienst aan (potentiële) klanten om inzicht te geven in de mogelijkheden voor hun organisatie.

De datasafari is niet louter een interne aangelegenheid voor de organisatie. Uit de innovatieliteratuur is bekend dat organisaties beter presteren wanneer ze ook de kennis uit hun omgeving weten te benutten door middel van crowdsourcing (Blohm, Leimeister, & Krcmar, 2013; Chesbrough, 2003; Helms & Booij, 2012). Organisaties dienen daarom open te staan voor het betrekken van derden in het exploratieproces. Uiteraard spelen concurrentiegevoeligheid en privacy van de data een belangrijke rol bij deze overweging. Een goed voorbeeld van crowdsourcing in het kader van de datasafari zijn zogenaamde hackathons (Burnham, 2012; Raatikainen, Komssi, Bianco, Kindstom, & Jarvinen, 2013). Dit houdt in dat een organisatie zijn data beschikbaar stelt aan ontwikkelaars en programmeurs die in de data gaan zoeken naar interessante inzichten en toepassingen. Aan het einde van de hackathon worden deze gepresenteerd aan de organisator. De samenwerking met de omgeving kan ook een meer structureel karakter worden gegeven door middel van het opzetten van een (online) community. Een community heeft als doel het uitwisselen van kennis en het ontwikkelen van best practices, binnen een groep personen met dezelfde belangstelling en interesses. Zelf heb ik veel onderzoek gedaan naar (online) communities en hoe deze het beste kunnen worden opgezet (Grootveld & Helms, 2008; Helms, Vrij, & Voogd, 2006; Iancu, Helms, Grahlmann, & Harmsen, 2012; van Reijssen et al., 2014). In de context van het datasafari-proces zou er in plaats van community ook kunnen worden gesproken van een datasafari-ecosysteem, waarin bedrijven zowel onderling als met kennisinstellingen samenwerken om hun data te exploreren. Naar dergelijke datasafari-ecosystemen is nog nauwelijks onderzoek gedaan.

Tot slot is het succes van de datasafari afhankelijk van de beschikbare data. Zoals de term big data suggereert is het idee vaak 'hoe meer, hoe beter'. Hierdoor kan de neiging ontstaan om zoveel mogelijk data uit zoveel mogelijk verschillende bronnen te verzamelen. Deze datahonger kan echter een bedreiging vormen voor de datasafari. Het accent blijft te veel liggen op de dataverzameling en men komt niet tot data-analyse omdat men eerst genoeg data wil hebben. Daarnaast kan deze datahonger leiden tot data overload, of misschien is het beter om naar analogie van surfen te spreken van een data-wipeout. Men wordt overspoeld met data waardoor men door de bomen het bos niet meer ziet en er onvoldoende capaciteit beschikbaar is om alle data te analyseren. Enig pragmatisme gedurende de datasafari is daarom gewenst om zo datahonger en data-wipeout te voorkomen.

⁷ <http://www.voort-innovatie.nl/voort-stappen/>



6.3 Organisatiestructuur

Als data een belangrijke grondstof zijn voor innovatie dan is het van belang de verzameling, opslag en analyse ervan goed te verankeren in de organisatie. Dit roept de vraag op of er voor big data en analytics ook een aparte afdeling nodig is? Voor het beheer van andere belangrijke grondstoffen, zoals financiële, personele en technologische middelen, zien we dat er in grote organisaties aparte stafafdelingen aanwezig zijn. Deze afdelingen zijn er voor verantwoordelijk dat de middelen op het juiste tijdstip en in de juiste hoeveelheid beschikbaar worden gesteld aan het primaire proces van de organisatie. Elk van deze afdelingen beschikt daartoe over zijn eigen expertise en processen.

Of er voor big data en analytics een eigen afdeling moet worden ingericht hangt er dus vanaf of het eigen expertise en processen vereist. In dit geval betreft de expertise enerzijds bigdata-technologie en anderzijds data analytics. In het voorgaande is al aangegeven dat de technologische expertise goed kan worden ondergebracht bij de IT-afdeling. Voor expertise op het gebied van analytics is het moeilijker een bestaande afdeling aan te wijzen. Twee mogelijke kandidaten zijn de kwaliteitsafdeling en de R&D-afdeling. Zij houden zich immers bezig met respectievelijk procesverbetering en productinnovatie, maar zij beschikken over het algemeen niet de benodigde data-analyticsvaardigheden. Het ligt dus voor de hand om een Analytics-afdeling in het leven te roepen zoals bijvoorbeeld in vakpublicaties, zoals die van Manyika et al. (2011), wordt voorgesteld.

De afdeling heeft ook zijn eigen proces, namelijk het datasafari-proces. In de voorbereidende fase van een datasafari werkt de analytics-afdeling samen met de IT-afdeling om de infrastructuur gereed te maken. Daarnaast is het hun taak om ervoor te zorgen dat de juiste databronnen beschikbaar zijn. Hiervoor dienen continu nieuwe interne en externe databronnen te worden geïdentificeerd en in de bigdata-infrastructuur te worden ondergebracht. In het uitvoeren van de datasafari neemt de analytics-afdeling werkt ze samen met andere afdelingen en externe partijen. In multidisciplinaire teams wordt er op projectbasis samengewerkt aan proces- en productinnovaties. Hiervoor is het ook noodzakelijk dat er aan opleiding wordt gedaan. De teamleden van andere afdelingen, die meedoen in het project, dienen namelijk voldoende data savvy te zijn. In deze opleidingen zouden ze kennis moeten maken met data analytics en zo inzicht verkrijgen in de mogelijkheden en implicaties ervan. Tot slot, zouden analyticsmedewerkers ruimte moeten krijgen voor eigen projecten, naar het bekende voorbeeld van bedrijven zoals Google en 3M (Crisuolo, Salter, & Wal, 2013). Dit wil zeggen dat medewerkers een bepaald deel van hun tijd mogen spenderen aan het naar eigen inzicht exploreren van de data.

Hoewel er nu goede redenen zijn aan te wijzen om een aparte analytics-afdeling in te stellen, is het ook voorstelbaar dat deze op niet al te lange termijn weer zal verdwijnen. Nu steeds meer opleidingen een vak als data analytics in het curriculum opnemen zal de medewerker van de toekomst steeds vaker over deze vaardigheden beschikken, terwijl dat nu nog een zeldzaamheid is. Ook zullen de tools ter ondersteuning van data-analyse zich verder ontwikkelen waardoor ze beter toegankelijker worden, net zoals dat in het verleden is gebeurd met tools voor managementrapportages. Product- en procesinnovatie op basis van data worden dan business as usual en kunnen door de organisatie zelf worden opgepakt. De belangrijkste vaardigheid die dan nog overblijft, is die voor het coördineren van het datasafari-proces. De tijd zal leren of dit scenario werkelijkheid wordt, in mijn onderzoek zal ik daarom de ontwikkeling van analytics binnen de organisatie nauw blijven volgen.





7 Conclusie

In mijn lezing heb ik getracht u te overtuigen dat data een cruciale grondstof zijn voor bedrijven waarmee competitief voordeel kan worden behaald. Aan de hand van een aantal voorbeelden heb ik laten zien dat big data, gekenmerkt door volume, variëteit en snelheid, nieuwe kansen biedt voor organisaties. Vervolgens heb ik het MOT-model (mensen, organisatie en technologie) gebruikt om aan te geven wat het voor een organisatie betekent om met big data aan de slag te gaan. Zo dienen organisaties al de bij ontwikkeling van de bedrijfs- en IT-strategie met big data rekening te houden. Mogelijk zal als gevolg daarvan ook het bedrijfsmodel moeten worden herzien. Vervolgens zal het datasafari-proces in lijn hiermee moeten worden ingericht. Een creatief proces dat op basis van data-analyse tot voorstellen voor proces- en productinnovatie moet leiden. Tijdens dit proces dient er over de eigen organisatiegrens heen te worden gekeken en derden door middel van crowdsourcing te worden betrokken. Als ondersteuning van dit proces dient een goede bigdata-infrastructuur beschikbaar te zijn. De rol van de technologie is daarbij die van aanjager en geen doel op zichzelf. Het is vooral de mens die een centrale rol vervult binnen het datasafari-proces. Goede analyticsvaardigheden zijn daarbij onmisbaar maar deze zijn momenteel maar zeer beperkt verkrijgbaar op de arbeidsmarkt. Doordat veel opleidingen momenteel analytics componenten in hun opleidingen opnemen zal dit tekort in de toekomst naar verwachting afnemen. Hiermee zal een datasafari in de toekomst business as usual worden voor organisaties.

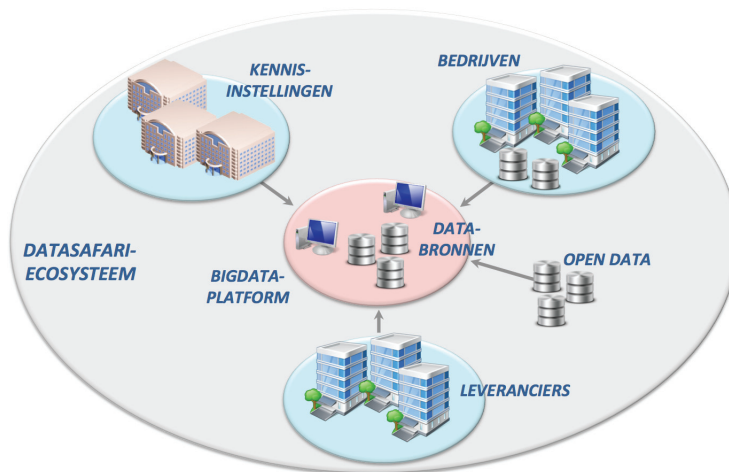
In het kader van mijn leerstoel aan de Open Universiteit wil ik onderzoek doen naar de datasafari's die organisaties ondernemen. Daarmee onderscheid ik me van het onderzoek dat al op het gebied van big data wordt uitgevoerd en zich voornamelijk richt op de ontwikkeling van technologie, tools en algoritmen (Pospiech & Felden, 2012). Een focus die je ook ziet terugkomen in de financiering van onderzoek naar big data door bijvoorbeeld NWO. Hiermee wil ik niet suggereren dat dit onbelangrijk is, technologie is zoals aangegeven immers een belangrijke aanjager. Maar nu de technologie op een niveau is gekomen dat hij door een breed aantal organisaties kan worden toegepast, zou er ook meer aandacht moeten zijn voor hoe organisaties hiervan optimaal kunnen profiteren. Hierbij gaat het niet alleen om grote organisaties maar zeker ook het midden- en kleinbedrijf dat wordt gezien als de motor van onze economie.

In mijn onderzoek wil ik me specifiek richten op de volgende onderwerpen:

- Datagedreven bedrijfsmodellen als hulpmiddel bij het formuleren van bigdata-strategieën voor organisaties.
- Datasafari-proces, inclusief tools en technieken, waarmee effectief proces- en productverbeteringen kunnen worden gegenereerd.
- Datasafari-ecosysteem waarin data, technologie, bedrijven en kennisinstellingen worden samengebracht om samen te werken aan innovatie.



Aan deze onderwerpen zal ik binnen het expertisecentrum Business Intelligence en Smart Services met plezier samenwerken met collega's van de andere betrokken kennisinstellingen, zoals Zuyd Hogeschool en de Universiteit Maastricht. In het onderzoek zullen we nadrukkelijk samenwerking zoeken met de organisaties die zich hebben aangesloten bij de Smart Services Hub. Met deze organisaties zou ik een datasafari-ecosysteem willen inrichten dat fungeert als een zogenaamd living lab (zie figuur 10). De twee belangrijkste componenten van het ecosysteem zijn een community van bedrijven, technologieaanbieders en kennisinstellingen, en een bigdata-infrastructuur waarin zowel data van organisaties als open data kunnen worden samengebracht. Binnen het ecosysteem kunnen bedrijven en kennisinstellingen samen experimenteren met big data. Voor studenten biedt dit een prachtige leeromgeving waarin zij hun opgedane vaardigheden en inzichten direct kunnen toepassen. Daarnaast kan er aan het datasafari-ecosysteem ook onderzoek worden gedaan om zo het datasafariproces verder te verbeteren en zo een verdere stimulans te geven aan innovatie binnen de bedrijven in de regio. Studenten en promovendi spelen een belangrijke rol in dit geheel. Zij zullen participeren in het onderzoek maar zij zijn ook diegenen die na afronding van hun opleiding de datasafari in organisaties moeten gaan uitvoeren. Het is dan ook van belang dat er gekeken wordt waar curricula moeten worden aangepast en indien nodig nieuwe curricula worden ontwikkeld. De eerste stappen zijn hierin al genomen en we hopen in de nabije toekomst een nieuwe generatie studenten alsook het bedrijfsleven hiermee te bedienen.



FIGUUR 10 Datasafari-ecosysteem

Dankwoord

Tot slot wil ik nog een aantal woorden van dank uitspreken aan diegenen die er aan hebben bijgedragen dat ik deze lezing vandaag heb kunnen geven. Maar om u niet langer dan noodzakelijk van de borrel af te houden, zal ik me beperken tot de hoofdrolspelers waarbij ik onvermijdelijk velen tekort doe.

Allereerst wil ik Anja Oskamp, rector magnificus van de Open Universiteit, Gerard Mertens, decaan van de faculteit Management, Science & Technology, en Lex Bijlsma, emeritus hoogleraar en decaan van de voormalige faculteit Informatica, bedanken dat zij mij het vertrouwen hebben gegeven en mij hebben aangesteld als hoogleraar Informatiekunde. Hierdoor ben ik sinds september 2013 aan de OU werkzaam bij de groep Informatiekunde & Bedrijfsprocessen, welke gedreven wordt geleid door Rob Kusters. Ik wil de collega's van de groep, en ook die van Informatica waar ik veel mee samenwerk, bedanken voor de collegiale samenwerking. Van hen wil ik Frank Wester nog speciaal bedanken voor het mij wegwijs maken in de OU-organisatie en de geboden ondersteuning bij mijn programmaleiderschap voor de opleiding Informatiekunde.

De stap naar de OU heb ik kunnen maken door de ervaring die ik heb kunnen opdoen bij het departement Informatica aan de Universiteit Utrecht. Daar ben ik in 2003 als universitair docent bij de opleiding Informatiekunde aangesteld door Jörgen van de Berg, grondlegger van Informatiekunde in Utrecht, en Lex Bijlsma, op dat moment onderwijsdirecteur bij het departement Informatica in Utrecht. Binnen de Organisatie & Informatie-groep van Sjaak Brinkkemper heb ik me als wetenschapper kunnen ontwikkelen en mogen werken aan mijn onderzoek op het gebied van kennismanagement en sociale media. Met veel plezier heb ik in het onderzoek samengewerkt, en werk ik nog steeds samen, met mijn promovendi: Jurriaan van Reijssen, Jurriaan Souer, Arjan de Kok, Gerard Wagenaar en Ilja Heitlager. Ook wil ik de vele afstudeerders bedanken voor hun bijdragen aan mijn onderzoek. Dit alles heb ik kunnen doen in een inspirerende omgeving en ik wil daarvoor graag mijn directe collega's bedanken. Sjaak Brinkkemper wil ik speciaal bedanken voor zijn directe betrokkenheid bij mijn carrière, die er zo toe heeft bijgedragen dat ik de stap naar hoogleraar heb kunnen maken.

Mijn academische scholing heb ik aan de Technische Universiteit Eindhoven doorlopen. Zowel de opleiding Technische Bedrijfskunde als mijn promotieonderzoek heb ik in Lampegat volbracht. Na mijn studie ben ik als organisatieadviseur in dienst getreden bij M.I.S Organisatie-ingenieurs. Maar dankzij de directeurs, Dick Mandemaker en Arjan de Kok, heb ik de stap terug naar de wetenschap kunnen maken omdat ze me de mogelijkheid boden in deeltijd een promotieonderzoek te doen. Dat deze stap de opmaat was voor vandaag heb ik toen nooit kunnen bevroeden.

Voor het voltooien van mijn promotieonderzoek blijf ik voor altijd schatplichtig aan Hans Wortmann, eerste promotor, Aarnout Brombacher, tweede promotor, en Henk Jan Pels, copromotor. Ik denk nog steeds met plezier terug aan de vele discussies die ik met Henk Jan heb gehad over het vrijgeven van voorlopige informatie.

Verder was vandaag niet mogelijk zonder Chrisja Muris en Iris Jacobs; zij hebben de organisatie van vandaag op zich genomen. Dat een en ander zo vloeiend is verlopen, is geheel hun verdienste. Dan is er John Arkenbout die ik wil bedanken voor het coördineren van de totstandkoming van mijn oratieboekje, Maria Kampermann voor de lay-out en vormgeving en de dienst Visuele Communicatie voor het omslag. En voor feedback op een ruwe versie van mijn oratie ben ik dank verschuldigd aan Sjaak Brinkkemper en Lex Bijlsma.

Natuurlijk ben ik ook mijn ouders Juus en Wim veel dank verschuldigd. Ik heb in een fijn gezin kunnen opgroeien samen met mijn twee broers Joost en Casper, en ben actief gestimuleerd om te gaan studeren. Dat ik de richting van de informatietechnologie ben opgegaan is de verdienste van mijn vader. Hij zag al vroeg de mogelijkheden hiervan en zo kwam er thuis al snel een Commodore64 en kreeg ik later bijvoorbeeld ook een PLC van zijn werk in mijn handen gedrukt om eens te proberen. Dat heeft ertoe bijgedragen dat ik destijds in Eindhoven Technische Natuurkunde ben gaan studeren, alhoewel in de keuze voor Eindhoven ook het basketbal een belangrijke rol speelde.

Ten slotte, ben ik de meeste dank verschuldigd aan Pien. Dank dat je er voor me bent en me altijd de ruimte hebt gegeven voor mijn wetenschappelijke loopbaan. Samen hebben we drie geweldige kinderen, Allard, een pitcher, Thijs, een catcher, en Isa, de grootste fan van haar honkbalbroers. Dit thuisfront geeft zowel de rust als de energie die nodig is om mijn werk met plezier te kunnen doen.

Ik heb gezegd.

Referenties

- Ariker, M., Breuer, P., & McGuire, T. (2014). How to get the most from big data. *McKinsey Insights*. Retrieved January 10, 2015, from http://www.mckinsey.com/insights/business_technology/how_to_get_the_most_from_big_data
- Blohm, I., Leimeister, J. M., & Krcmar, H. (2013). Crowdsourcing: How to benefit from (too) many great ideas. *MIS Quarterly Executive*, 12(4), 199–211.
- Burnham, K. (2012). Inside Facebook's Hackathons: 5 Tips For Hosting Your Own | CIO. *CIO.com*.
- Chen, H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11, 529–555.
- Chesbrough, H. W. (2003). The era of open innovation. *MIT Sloan Management Review*, 44(3), 35–41.
- Clarke, T., & Begley, S. (2015). U.S. proposes effort to analyze DNA from 1 million people | Reuters. Reuters. Retrieved February 01, 2015, from <http://www.reuters.com/article/2015/01/30/us-usa-obama-precisionmedicine-idUSKBN0L313R20150130>
- Criscuolo, P., Salter, A., & Wal, A. T. Ter. (2013). Going underground: bootlegging and individual innovative performance. *Organization Science*, 25(5), 1–19.
- Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, 91(12), 65–72.
- Davenport, T. H., & Patil, D. J. (2012). Data Scientist: the sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70–77.
- Dawson, F. (2013, August 12). IBM News room - 2013-08-12 Made in IBM Labs: IBM Drives the Future of Renewable Energy with New Wind and Solar Forecasting System - United States. *IBM News*. Retrieved January 04, 2015, from <http://www-03.ibm.com/press/us/en/pressrelease/41310.wss>
- Fisher, D. (2013). Winging It: How Larry Ellison Harnessed Big Data To Win The America's Cup - Forbes. *Forbes*. Retrieved January 04, 2015, from <http://www.forbes.com/sites/danielfisher/2013/12/23/winging-it-how-larry-ellison-harnessed-big-data-to-win-the-americas-cup/>



- Gillebaard, H., Jager, C.-J., Velde, R. te, Steur, J., & Vankan, A. (2014). *Dé ICT'er bestaat niet: analyse van vraag en aanbod op de Nederlandse ICT-arbeidsmarkt* (p. 60). Utrecht. Retrieved from [http://www.nederlandict.nl/Files/TER/Dialogic%282014%29 Onderzoeksrapport ICT-arbeidsmarkt.pdf](http://www.nederlandict.nl/Files/TER/Dialogic%282014%29%20Onderzoeksrapport%20ICT-arbeidsmarkt.pdf)
- Grootveld, A., & Helms, R. W. (2008). Development and Application of a Factor Framework to Diagnose Possible Failure in Communities of Practice. In *The 9th European Conference on Knowledge Management* (pp. 249–256). Southhampton, UK: Academic Conferences.
- Hammerbacher, J. (2009). Information platforms and the rise of the data scientist. In T. Segaran & J. Hammerbacher (Eds.), *Beautiful Data* (pp. 73–84).
- Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). *Big Data for Big Business?. Cambridge Service Alliance Blog* (pp. 1–29).
- Hein, R. (2014). The 8 Most In-Demand Big Data Roles. *CIO.com*. Retrieved January 28, 2015, from <http://www.cio.com/article/2369410/it-organization/135970-The-8-Most-In-Demand-Big-Data-Roles.html>
- Helms, R. W., & Booi, E. (2012). Reaching Out: Involving users in Innovation Tasks through Social Media. In *Proceedings of the 20th European Conference on Information Systems* (p. paper 193). Barcelona, Spain.
- Helms, R. W., Vrij, N. de, & Voogd, P. (2006). Application of a Community of Practice to improve knowledge sharing in offshoring relations. In R. S. Bliof (Ed.), *IEEE Proceedings of the of the 7th International Workshop on Theory and Applications of Knowledge Management* (pp. 161–165). Krakow, Poland: IEEE Computer Society.
- Helms, R. W., & Werder, K. (2013). Who Reads Corporate Tweets ? Network Analysis of Follower Communities. In *Proceedings of the 19th Americas Conference on Information Systems* (pp. 1–11). Chicago, IL: AIS Library.
- Henderson, J. C., & Venkatraman, H. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 32(1), 472–484.
- Hezewijk, R. van. (2015). *Data zijn geen kennis* (p. 60). Heerlen: Open University.
- Iancu, E., Helms, R. W., Grahlmann, K. R., & Harmsen, F. (2012). Improving Knowledge Sharing: Leading Practices Applied to a Distributed Program Management Community. In *Proceedings of the 13th European Conference on Knowledge Management*. Cartagena, Spain.

- IBM SPSS Modeler CRISP-DM Guide*. (2011) (p. 53). Retrieved from ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/CRISP_DM.pdf
- Lavalle, S., Lesser, E., Schockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21–31.
- Leavitt, H. J. (1964). Applied Organization Change in Industry: structural, technical and human approaches. In W. Cooper, H. J. Leavitt, & M. W. I. Shelly (Eds.), *New Perspectives in Organization Research* (pp. 55–71). New York, NY: John Wiley.
- Lycett, M. (2013). “Datafication”: Making sense of (big) data in a complex world. *European Journal of Information Systems*, 22(4), 381–386.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: the management revolution. *Harvard Business Review*, 90(10), 61–68.
- McMillon, D. (2014). Picking up the Pace of Change for the Customer. *WalMart News*. Retrieved January 04, 2015, from <http://news.walmart.com/executive-viewpoints/picking-up-the-pace-of-change-for-the-customer>
- Osterwalder, A., Pigneur, Y., Smith, A., & Movement, T. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. *Journal of Business* (1st ed., Vol. 5, p. 288). John Wiley and Sons.
- Pospiech, M., & Felden, C. (2012). Big data - A State-of-the-Art. In *18th Americas Conference on Information Systems 2012, AMCIS 2012* (Vol. 5, pp. 3918–3928).
- Raatikainen, M., Komssi, M., Bianco, V. dal, Kindstom, K., & Jarvinen, J. (2013). Industrial Experiences of Organizing a Hackathon to Assess a Device-centric Cloud Ecosystem. In *2013 IEEE 37th Annual Computer Software and Applications Conference* (pp. 790–799). IEEE.
- Robinson, R. J. (2014). How big is the human genome? *Precision medicine*. Retrieved February 01, 2015, from <https://medium.com/precision-medicine/how-big-is-the-human-genome-e90caa3409b0>



- Seddon, P. B., Dod, H., & Constantinidis, D. (2012). How does business analytics contribute to business value? In *International Conference on Information Systems, ICIS 2012* (Vol. 4, pp. 3380–3396).
- Shanks, G., & Bekmamedova, N. (2013). Understanding business analytics innovations using dynamic capabilities: The BAIP model. In *ECIS 2013 - Proceedings of the 21st European Conference on Information Systems*.
- Sharda, R., Delen, D., & Turban, E. (2013). *Business Intelligence: A Managerial Perspective on Analytics*. Harlow, Essex, UK: Pearson Education Limited.
- Van Reijssen, J., Helms, R., Batenburg, R., & Foorhuis, R. (2014). The impact of knowledge management and social capital on dynamic capability in organizations. *Knowledge Management Research & Practice*, (Advanced online publication January 20), 1–17.
- Wang, B. (2009). Next Big Future: Google is Planning for 10 Million Servers and an Exabyte of Information. *nextBIG future*. Retrieved January 04, 2015, from <http://nextbigfuture.com/2009/10/google-is-planning-for-10-million.html>
- Ward, J., & Peppard, J. (2002). *Strategic planning for information systems. Wiley series in information systems* (3rd ed., p. 624). Chichester: John Wiley & Sons, Ltd.
- Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(1), 1247–1268.
- Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U., & Phillips-Wren, G. (2011). The current state of Business Intelligence in academia. *Communications of the Association for Information Systems*, 29(1), 299–312.

Contactgegevens

Prof. dr. ir. Remko W. Helms
Hoogleraar Informatiekunde

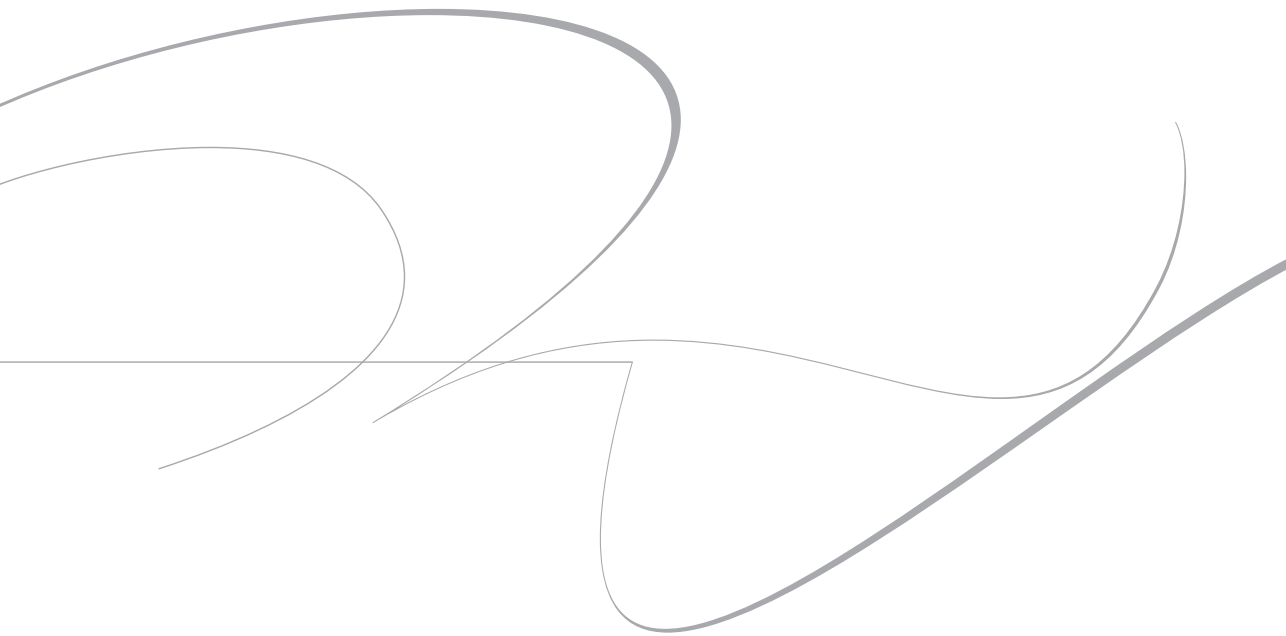
Postadres:
Open Universiteit
Faculteit Management, Science & Technology
Postbus 2960
6401 DL Heerlen

Bezoekadres:
Milton Keynes Gebouw, 3e verdieping
Valkenburgerweg 177
6419 AT Heerlen

Telefoon: 045-576 2828 (secretariaat)
E-mail: remko.helms@ou.nl
Website: <http://remhelms.wordpress.com>
Twitter: [@remhelms](https://twitter.com/remhelms)⁸

⁸ Twopcharts: Remko has been on Twitter for 64.9% of its existence. Remko has been on Twitter longer than 98.609% of all other Twitter users.





Bijlage: European e- Competence Framework 3.0 overview

Dimension 1 5 e-CF areas (A – E)	Dimension 2 40 e-Competences identified	Dimension 3 e-Competence proficiency levels e-1 to e-5, related to EQF levels 3–8				
		e-1	e-2	e-3	e-4	e-5
A. PLAN	A.1. IS and Business Strategy Alignment					
	A.2. Service Level Management					
	A.3. Business Plan Development					
	A.4. Product/Service Planning					
	A.5. Architecture Design					
	A.6. Application Design					
	A.7. Technology Trend Monitoring					
	A.8. Sustainable Development					
	A.9. Innovating					
B. BUILD	B.1. Application Development					
	B.2. Component Integration					
	B.3. Testing					
	B.4. Solution Deployment					
	B.5. Documentation Production					
	B.6. Systems Engineering					
C. RUN	C.1. User Support					
	C.2. Change Support					
	C.3. Service Delivery					
	C.4. Problem Management					
D. ENABLE	D.1. Information Security Strategy Development					
	D.2. ICT Quality Strategy Development					
	D.3. Education and Training Provision					
	D.4. Purchasing					
	D.5. Sales Proposal Development					
	D.6. Channel Management					
	D.7. Sales Management					
	D.8. Contract Management					
	D.9. Personnel Development					
	D.10. Information and Knowledge Management					
	D.11. Needs Identification					
	D.12. Digital Marketing					
E. MANAGE	E.1. Forecast Development					
	E.2. Project and Portfolio Management					
	E.3. Risk Management					
	E.4. Relationship Management					
	E.5. Process Improvement					
	E.6. ICT Quality Management					
	E.7. Business Change Management					
	E.8. Information Security Management					
	E.9. IS Governance					

Bron: <http://www.ecompetences.eu>

