

# Big Data Bla Bla

**Author(s)**

van Bussel, G.J.

**Publication date**

2015

**Document Version**

Final published version

**Published in**

IP: Vakblad voor Informatieprofessionals

**License**

Unspecified

[Link to publication](#)

**Citation for published version (APA):**

van Bussel, G. J. (2015). Big Data Bla Bla. *IP: Vakblad voor Informatieprofessionals*, (4), 25.

<https://informatieprofessional.nl/2015/05/column-big-data-bla-bla/>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact/questions>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Geert-Jan van Bussel

# COLUMN

## Big Data Bla Bla

In 2008 schreef Chris Anderson, redacteur bij *Wired*, dat de zondvloed aan data (later *big data* genoemd) en nieuwe analysetools de wetenschappelijke methode onnodig maakten en een nieuwe manier waren om de wereld te begrijpen. 'Correlation supercedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all'.

Big data-blabla'ers (ook evangelisten genoemd) hebben deze idee met vele anekdotes ondersteund en verkondigd. Volgens hen breekt er een nieuwe tijd aan: de dataficatie van onze maatschappij maakt alles meetbaar en doet alle bestaande problemen verdwijnen. Gemakshalve wordt eraan voorbijgegaan dat (zoals onderzoeksbedrijf Gartner stelt) 55 procent van alle big data-projecten mislukt, vooral omdat ze niet de optimistische resultaten leveren die vooraf zijn ingecalculleerd. Zélfs het paradepaardje van de big data-evangelisten niet: het in 2008 gestarte Google Flu Trends (GFT). Het geprofeteerde succes bleek rafelrandjes te kennen. In 2014 werd in *Science* aangetoond dat de resultaten (nog steeds) niet overeenkwamen met de 'echte' feiten. Het voorspellen van griep trends lukte vele malen beter met drie weken oude, op traditionele wijze verzamelde en geanalyseerde gegevens van griepcentra.

De 500 miljoen zoekvragen per dag van Google leveren gevisualiseerd dus nog geen accuraat beeld van wat er werkelijk gaande is. Dat kan niet als (zoals bij GFT) de gebruikte gegevens onbe-



**'Gemakshalve wordt eraan voorbijgegaan dat 55 procent van alle big data-projecten mislukt'**

Geert-Jan van Bussel is ondernemer en lector Digital Archiving & Compliance bij de Hogeschool van Amsterdam

trouwbaar en onjuist zijn en/of de verkeerde gegevens zijn verzameld of geselecteerd. In *Science* werd aangetoond dat betrouwbare voorspellingen voor de verspreiding van griep mogelijk zijn als de analyses uit big data gecombineerd worden met traditionele, wetenschappelijke vormen van gegevensverzameling en -analyse. Een van de problemen van GFT (en andere analyses uit big data) is dat de resultaten veelal niet gerepliceerd kunnen worden. Dit komt vooral doordat algoritmes,

formules, zoektermen en datasetselecties geheim zijn en eigendom van private bedrijven.

Ik waag te betwijfelen of we wel alleen op data kunnen vertrouwen als indicator voor beslissingen, welvaart of zingeving. Jay Liebowitz propageert in zijn boek *Bursting the Big Data Bubble* om naast de resultaten van data-analyse in beslissingsprocessen (met dan ook nog een voorkeur voor 'small data') ruimte te laten voor ervaring en intuïtie.

Het is zeker dat door ervaring en intuïtie fouten gemaakt (en voorkomen!) worden. Het is zeker dat data beslissingen kunnen verbeteren (als ze objectief worden gebruikt!). Het is ook aangetoond dat we gegevens negeren als ze niet overeenkomen met wat we subjectief denken (zelfs als de objectiviteit van die gegevens vaststaat!).

Big data-evangelisten verkondigen dat 'you can only manage what you measure'. Blabla. Want de financiële crisis heeft aangetoond dat we slecht zijn in het managen van wat we meten. Mislukte fusies en productlanceringen, veelvuldige imago-problemen en social media-escapades geven aan dat we vooral beter moeten worden in het managen van datgene wat we niet kunnen meten.

Met of zonder big data.

<