

# Waanzin

**Author(s)**

van Bussel, G.J.

**Publication date**

2015

**Document Version**

Final published version

**Published in**

IP: Vakblad voor Informatieprofessionals

**License**

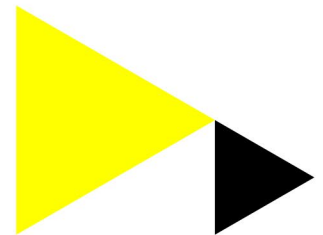
Unspecified

[Link to publication](#)

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

van Bussel, G. J. (2015). Waanzin. *IP: Vakblad voor Informatieprofessionals*, (3), 29.

<https://informatieprofessional.nl/2015/04/column-waanzin/>

**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

## Waanzin

Deze week werd ik weer eens geconfronteerd met de uitspraak dat opslag van data goedkoop is en dat alles makkelijk kan worden bewaard voor een big data-strategie. Ik kon mijn lachen niet inhouden. In 2000 kostte een gigabyte opslag gemiddeld 8,31 euro, in 2010 7 eurocent. De toegenomen hoeveelheid data en het effectievere gebruik van opslaghardware

in diezelfde periode zorgden voor die substantiële verlaging van de opslagprijzen. De springvloed in Thailand verhinderde een verdere verlaging van de opslagprijzen na 2011. Veertig procent van de productiecapaciteit van opslagschijven werd toen vernietigd, waardoor de prijzen voor opslagschijven verdubbelden. Die prijzen zijn in het vervolg weer omlaaggegaan, maar ze hebben het niveau van 2011 niet meer bereikt.

Wat in gigabyte-opslagprijzen nooit is meegenomen is het gebruik en de terugvindbaarheid van data. Van 2002 tot 2010 zijn de softwarekosten daarvoor gestegen van 4,8 miljard tot 10,9 miljard euro. De reden voor die kostenstijging is het feit dat data betrouwbaar en duurzaam moeten zijn. Ze worden immers opgeslagen om te worden gebruikt.

Na 2011 neemt de hoeveelheid data en databestanden explosief toe (naar 'multiple zettabytes' in 2020, volgens onderzoeksbureau IDC). Hoewel een groot deel van die stijging te wijten valt aan 'streaming' video en audio, vergen de hoeveelheden daadwerkelijk opgeslagen data (ongeveer twintig procent van die 'multiple zettabytes') steeds meer capaciteit.

Het information governance-regime, nodig voor vertrouwelijkheid, privacy, compliance en erfgoed, vergt met de stijging



**'Het verdwijnen van data, waar Vint Cerf onlangs voor waarschuwde, is zo erg nog niet'**

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

van de hoeveelheid data meer investeringen in software.

Maar hoe hoog zijn die kosten dan? Twee voorbeelden.

Het Internet Archive heeft de goedkoopste oplossing in gebruik, volledig geautomatiseerd, met weinig management en simpele procedures. In 2011 had het twee petabytes aan webarchief en groeide het met een kwart petabyte per jaar. De kosten bedroegen tussen 9 en 13 miljoen euro per jaar, gemiddeld 40 eurocent per gigabyte. Portico verzamelt academische

literatuur in een edepot, met nadruk op management en kwaliteitsprocedures. In 2011 had het vijftig terabytes in beheer en groeide het met vijf terabytes per jaar. De kosten bedroegen 5 tot 6,5 miljoen euro per jaar, gemiddeld 9 euro per gigabyte. Na 2011 hebben ze (gezien de groei in de hoeveelheid te bewaren data) hun opslagcapaciteit uitgebreid. De gemiddelde opslagkosten per gigabyte zijn ongeveer hetzelfde gebleven, maar er is een stijgende lijn te constateren.

Stel dat we een exabyte wereldwijd bewaren (een miljard gigabytes). Dat is een half promille van de wereldwijd geproduceerde twee zettabytes (tweeduizend exabytes) aan data in 2011. We slaan ongeveer twintig procent (vierhonderd exabytes) daarvan op. Het bewaren van die ene exabyte kost volgens het Internet Archive 400 miljoen euro en volgens Portico 9 miljard euro per jaar.

In 2020 produceren we misschien vijftien of twintig zettabytes!

Het verdwijnen van data, waar Vint Cerf onlangs voor waarschuwde, is zo erg nog niet.

Waarderen, selecteren en vernietigen van data wordt enorm belangrijk. Bibliothecarissen en archivarissen zullen er hun handen vol aan hebben.

Alles bewaren is totale waanzin. <