

VELDWIJK

# Over de datarevolutie

Recentelijk las ik een verhaal van de SF-schrijver Isaac Asimov dat sterk van toepassing lijkt op de revolutie op gegevensgebied die om ons heen woedt. Het verhaal, *Het Dode Verleden*, gaat over een hoogleraar geschiedenis die geen toestemming krijgt om zijn onderzoeksobject, het antieke Carthago, te onderzoeken met een chronoscoop, een apparaat waarmee het verre verleden direct kan worden waargenomen. Gefrustreerd door alle tegenwerking krijgt hij een natuurkundige zover om zelf een chronoscoop te maken. Deze ontdekt dat het theoretisch onmogelijk is om een apparaat te maken dat ver terugkijkt in het verleden, maar dat het eenvoudig is om het recente verleden waar te nemen. De overheid had besloten om iedereen over de aard van de chronoscoop te misleiden, omdat men zich realiseerde dat een maatschappij niet kan functioneren als iedereen elkaars doen en laten kan bekijken. Het verhaal eindigt met het uitlekken van het geheim en het einde van de maatschappij die uiteindelijk is gegrondvest op vergaande privacy voor bijna iedereen.

Dit verhaal lijkt mij relevant voor wat zich nu lijkt af te spelen op gegevensgebied. Er bestaat geen chronoscoop waarmee we alles en iedereen kunnen waarnemen, maar ons doen en laten in de echte wereld slaat wel op grote schaal neer in de wereld van de gegevens. En die wereld wordt in rap tempo ontsloten voor toegang door eenieder die interesse heeft. Ik heb het natuurlijk over de enorme hoeveelheid gegevens die benaderbaar is via het internet en die kan worden 'opgegoogeld' via zoekmachines. Veel van wat er in de echte wereld gebeurt – van de uitslagen van de plaatselijke korfbalclub tot foto's van kleine Amalia en grote Lynndie – is direct toegankelijk via het nieuwe grote oog dat internet heet. Natuurlijk beschrijven alle Petabytes aan gegevens maar een fractie van ons leven, maar die fractie groeit snel. Ook is het soms niet mogelijk om in te zoomen op een bepaalde persoon met een teveel voorkomende naam. Jan Jansen is bijvoorbeeld veel beter beschermd tegen spiedende blikken dan, zeg, Hans Lamboo. Daar staat echter tegenover dat rondkijken via een zoekmachine soms allerlei hoogst bruikbare *Aha-Erlebnisse* oplevert. Zo keek ik laatst naar informatie over de buurt van een huis dat mijn interesse heeft, maar vond ik informatie over allerlei BV's die daar failliet waren gegaan. Handig voor de onderhandelingen met de verkopende partij natuurlijk.

Sluipend maar onmiskenbaar is het informatiegedrag van mensen aan het veranderen. Moderne info-consumenten gebruiken zoekmachines routinematig om informatie te verwerven over zaken en mensen die op hun weg komen. En omdat internet en

de zoekmachines zelf van alles archiveren, is die zoekmachine zich meer en meer aan het ontwikkelen tot de chronoscoop uit het verhaal van Asimov. Wee straks eenieder met een ongebruikelijke naam. Echt eng wordt het overigens pas wanneer professionals de beschikking krijgen over de I-chronoscoop. We hoeven dan niet eens te denken aan zaken als Echelon en de CIA. Een vriend vertelde mij geïmponeerd hoeveel informatie er uit de screening komt die men bij zijn werkgever routinematig uitvoert voor mensen die op gevoelige plaatsen werken. De combinatie van gestructureerde informatie uit reguliere administratieve systemen waartoe zijn werkgever toegang heeft en ongestructureerde informatie uit het internet-domein, is in handen van geschoold personeel uitermate krachtig en dat fenomeen is wat ik in deze column met u wil delen – het privacy-probleem is alleen maar ter illustratie.

Wat mij zorgen baart is hoe snel de wereld van gestructureerde gegevens naar de marge wordt gedrongen door de ontwikkelingen in het domein van ongestructureerde gegevens. Google komt in milliseconden met een antwoord op een zoekvraag in een berg van ongestructureerde informatie, terwijl een eenvoudige query op een database met gestructureerde gegevens vaak vele seconden kost. Netjes gestructureerde gegevens in administratieve systemen die vaak met veel inspanning zijn verworven en vastgelegd, komen nauwelijks tot hun recht doordat die systemen effectief niet koppelbaar zijn – al het denken over corporate gegevensmodellen, datawarehouses en Business Intelligence ten spijt. Een vreselijke gedachte: als bedrijven hun gestructureerde gegevens in HTML-vorm op hun intranet zouden dumpen en zoekmachines in *local mode* zouden draaien, dan zou de waardevolle bedrijfsinformatie veelal veel beter worden benut dan nu het geval is. Weer een markt voor Google maar een moreel faillissement voor ons administratieve ICT'ers. De boodschap van deze column is niet dat administratieve automatisering en uw en mijn werk wordt bedreigd. Er zullen altijd vele processen blijven die staan of vallen bij de beschikbaarheid van netjes gestructureerde gegevens. De boodschap is wel dat er iets wezenlijks mis is met de manier waarop wij aankijken tegen de ruimere wereld van data en dat we met zijn allen bezig zijn om de datarevolutie te missen die zich om ons heen afspeelt. Reden tot zorg en mogelijkheden tot vernieuwing. *Et ceterum censeo Cartago delendam esse!*

**Dr. R.J. Veldwijk** ([rene.veldwijk@faapartners.com](mailto:rene.veldwijk@faapartners.com)) is werkzaam bij FAA Partners.