

# De Toekomst van Data

Peter Apers

## 0. Woord vooraf

Iedereen van harte welkom. Leden van het CvB: van harte welkom; Collega's van zowel binnen als buiten de universiteit: van harte welkom; vrienden en familieleden: van harte welkom. Iedereen welkom in mijn huiskamer. Als jullie een beetje inschikken past het wel.

Via LinkedIn heb ik mijn afscheid aangekondigd. Ik heb nog nooit zoveel "likes" gehad. Eerst dacht ik: leuk. Later bedacht ik me: misschien vinden ze het wel heel leuk dat ik weg ga.

We weten de betekenis van die likes dus niet echt en toch kennen we er een betekenis aan toe. Daar gaat Data Science over. Vanmiddag wil ik met jullie de ontwikkelingen van Data Science bespreken die natuurlijk nauw verbonden zijn met de ontwikkeling van het Internet, en de opkomende rol van Data.

## 1. Inleiding

Laat ik beginnen met het Internet, dat inmiddels volledig geïntegreerd is in ons dagelijks leven. Eerst was het een leuke manier om te mailen en foto's met elkaar te delen. Het was allemaal heel onschuldig.

- Nu horen we dagelijks over *nepnieuws* en *beïnvloeding* van onze mening op een manier waarop we niet bedacht waren. Aan commercials, die bedoeld zijn om ons koopgedrag te beïnvloeden, zijn we gewend; fake Facebook accounts om politieke ideeën te propageren en verkiezingen te beïnvloeden zagen we even niet aankomen;
- Nu horen we over dynamic pricing: 's ochtends zijn de prijzen anders dan 's middags, internet-kopen via een Apple apparaat is duurder dan via een Windows apparaat. Zijn we op weg naar *personalized pricing*: de prijs die je betaalt is afhankelijk van wat de verkoper van je weet: waar je werkt, welke auto je rijdt, waar je woont? Verzekeraars experimenteren hier al mee: uw geregistreerd rijgedrag beïnvloedt de premie van uw autoverzekering.
- Nu horen we over apps die ons helpen om ZELF in een vroegtijdig stadium ernstige *ziekten te detecteren*, denk aan huidkanker. Of je hoest tegen je smartphone of te kijken of je longontsteking hebt.

Eén ding is duidelijk, Internet doet onze beschaving op haar grondvesten schudden. Zeker als we weten dat nepnieuws de voorkeur krijgt bij social media. We zien de wereld rondom data zakelijker en professioneler worden. Hoe reageren we daarop?

Hoe die data-wereld zich zo snel heeft kunnen ontwikkelen en wat de trends zijn, daarover wil ik vandaag met jullie praten. Ik doe dat aan de hand van een paar tijdlijnen:

- ontwikkelingen van Data
- ontwikkelingen van ICT
- en mijn eigen activiteiten.

## 2. Data is van alle tijden

Dames en Heren, ik begon mijn betoog over het internet in ons dagelijks leven, nu wil ik de stap zetten naar mijn eigen vakgebied: Data. Data is van alle tijden. Het gaat altijd om het *vastleggen* en het *ordenen* van data om dan op basis van antwoorden op *vragen* over de data besluiten te kunnen nemen.

Eerst iets over **vastleggen**. Tijdens mijn oratie in 1986 vertelde ik jullie al iets over Quipu's. Quipu's werden door de Inca's gebruikt om de voorraden bij te houden, van bijvoorbeeld mais en bonen. Zoals je kunt zien, een quipu bestaat uit gekleurde koordjes met knopen. Lang is gedacht dat quipu's alleen gebruikt werden om numerieke gegevens vast te leggen. De antropologe Sabine Hyland is door de recente ontdekking van twee, lang verborgen gebleven quipu's op het idee gekomen dat quipu's eigenlijk verhalende brieven zijn. Het aantal kleur- en knoopcombinaties in quipu's komt min of meer overeen met het aantal symbolen van een logografisch schriftsysteem, vergelijkbaar met de Egyptische hiërogliefen en de Chinese karakters. Ontcijfering is echter nog steeds niet gelukt door het ontbreken van een Steen van Rosetta die een vertaling naar een bekende taal mogelijk zou maken.

Na vastleggen komt ordenen. Het **ordenen** van data heeft te maken met het vastleggen van verbanden tussen data. Bijvoorbeeld, een moeder-kind relatie, is een voorbeeld van een hiërarchische verband, oftewel een 1:N verband. Een moeder kan meerdere kinderen hebben. Echter een kind heeft maar één biologische moeder. Ik lijk Juf Ank wel uit de Luizenmoeder.

Een tekstdocument is natuurlijk ook data; daarin herkennen we een zekere ordening. Al in 1945 deed Vannevar Bush iets opmerkelijks. Hij wilde kunnen

verwijzen vanuit het ene document naar een deel van een ander document; hij deed dat met Memex en 'As we may think'. Het concept hypertext was toen geboren. Een concept dat we associëren met het World-Wide Web, het tegenwoordige Internet. Maar dat bestond nog lang niet. Dus het was echt een revolutionair idee.

Ik gaf net al aan dat het doel van het registreren en ordenen is om besluiten te kunnen nemen. Bijvoorbeeld, de herverdeling van maïsvoorraden. Dat kan alleen maar als je **vragen** kunt stellen over de data. Dat vereist een beschrijving van de data (we noemen dat een schema) en een vraagtaal.

### **3. Eind jaren 60 (1969-1970): het begin van opslag en communicatie van data**

Ik heb u inmiddels meegenomen naar mijn werkkamer. Bij de eerste databases leek het datamodel, waarin het schema beschreven werd, op de datastructuren zoals we die in programmeertalen kennen. Daar kwam al snel verandering in.

Eind jaren 60 werd het *Relationele datamodel* geïntroduceerd. De essentie was dat verbanden niet meer in datastructuren werden vastgelegd, maar via matching van attribuut-waarden van tuples. Zie het plaatje, het was een geweldig idee. Het betekende echter dat er op het niveau van de opslagstructuren verschillende verwerkingsmogelijkheden waren om een vraag te beantwoorden. De ene mogelijkheid kan echter veel meer tijd in beslag nemen dan de andere. Het doel is om juist die verwerking van de vraag te nemen die de minste tijd in beslag neemt. Het efficiënt verwerken van vragen is het wetenschapsgebied binnen Databases waar ik aangewerkt heb.

In diezelfde tijd als de introductie van het relationele datamodel, werden de eerste computers met elkaar verbonden (het ARPANET). Het betrof computers van UC Los Angeles en Santa Barbara, Stanford University en University of Utah. Later volgde er nog veel meer. Toen werd er ook het Internet Protocol overheen gelegd om computers met elkaar te laten communiceren. Dit is de start van ons huidige internet waar de data nu over vele miljoenen computers is verspreid. Realiseer je je dat dit nog geen 50 jaar geleden is.

**Eind jaren zestig is dus het begin van de opslag en communicatie van data.**

## **4. Eerste ICT-systemen (1974-1990): het begin van digitalisering**

Halverwege de jaren 70 kwamen de eerste ICT-systemen. Ingres, het eerste relationele database systeem, werd aan de Universiteit van Californië in Berkeley ontwikkeld door Mike Stonebraker. En min of meer parallel daaraan werd bij IBM gewerkt aan System R, ook een relationeel database systeem. Dit was ook de tijd dat grootschalige IR systemen gebouwd werden, zoals Lockheed Dialog System door Roger Summit.

Eind jaren 70 ben ik met mijn promotieonderzoek aan de VU begonnen: Optimalisatie van de verwerking van vragen en de gegevensopslag in gedistribueerde databases. Gegevens opslaan op verschillende met elkaar verbonden computers was nog iets voor de toekomst.

Eind jaren 80 werkten Nederlandse database onderzoekers samen met Philips aan een grote parallelle database machine. De PRISMA-machine: 100 nodes en alle data in main memory. Dat maakte internationaal pas indruk. Annita Wilschut schreef in die tijd het best geciteerde artikel uit mijn groep, het ging over een efficiënte implementatie van de hash join in een pipeline architectuur. Het voert te ver om dat nu verder uit te werken, echter neem van mij maar aan dat haar werk tot op de dag van vandaag terug te vinden is in de meeste database systemen.

Het Internet stond nog in de kinderschoenen: computers waren net met elkaar verbonden, email kon verstuurd worden (je wil alleen niet weten hoe) en de eerste indexprogramma's kwamen beschikbaar. Ik kan me nog herinneren dat aan het eind van mijn eerste decanaat ik een soort van sabbatical bij het CWI had. Ik zat als een klein kind de structuur van het Internet te ontdekken. Zoeken op Internet was toen nog best ingewikkeld. Google bestond nog niet.

Ik noemde u al de geboorte van hypertext in 1945. In deze periode werd het werkelijkheid. Je kon in een document een verwijzing opnemen naar een ander document op een computer aan de andere kant van de wereld. Alle ingewikkelde zaken: hoe je andere computers kunt vinden en hoe computers met elkaar communiceren etc werden onzichtbaar gemaakt. Dit heeft enorm bijgedragen aan de populariteit van het Internet bij het grote publiek.

Het zit database-onderzoekers even tegen. De populariteit van het Internet zorgde in de database wereld voor een behoorlijke identiteitscrisis. Want het Internet hield zich bezig met data, maar stopte die data niet in een database. Overboord met het

schema, de closed world assumption, overboord met fatsoenlijke vraagstukken. Overboord dus met alle goodies uit de database wereld. En het Internet werd steeds populairder en daardoor groeide de belangstelling en waardering voor IR, de basis van de Googles van deze wereld.

Een illustratie van die identiteitscrisis. Ik werd uitgenodigd voor een workshop in Como, met alleen toponderzoekers. Dus een hele eer. We gingen het hebben over uitdagende onderwerpen: Databases en het Internet. Ik voelde me behoorlijk uitgedaagd om voor zo'n gezelschap met iets echt uitdagends te komen. Met mijn AIOs overlegd. Het was de tijd dat je een file vanaf je PC naar de printer op de gang kon sturen. En wij hadden bedacht dat data niet op je PC maar in het Internet opgeslagen zou worden.

Mike Stonebraker, de database pionier die ik al eerder noemde, was genadeloos in zijn kritiek: "complete onzin, geen bedrijf zal zijn data ooit in het Internet opslaan". Hij kreeg veel bijval. Tijdens de pauze, in de toilet, kreeg ik van een paar collega's - die medelijden met me hadden - te horen dat sommige ideeën iets te vroeg komen. Het begrip "data in the cloud" bestond nog niet. Nu kunnen we ons niet meer voorstellen dat er een tijd bestond zonder "cloud".

**In deze periode van eerste ICT-systemen is, nu terugkijkend, de eerste stap naar digitalisering gezet.**

## **5. Millenniumwisseling (1990-2010)**

Dames en heren, u herinnert zich waarschijnlijk de eerste internet bubble die in het jaar 2000 knapte. Misschien heeft u dat indertijd veel geld gekost. Maar er speelde nog iets met de millenniumwisseling, weet u nog? De beruchte millenniumbug. Om geheugen te besparen werden van een jaartal alleen de laatste twee cijfers opgeslagen. Van 1999 dus alleen 99. En dus van 2000 alleen 00. Het vervelende is dat 00 voor 99 komt terwijl 2000 na 1999 komt. Totale chaos werd voorspeld. Iedereen ging hamsteren. Gelukkig hebben nogal wat ICT-bedrijven er goed aan verdiend en konden we weer rustig verder leven.

Weet u nog dat [Booking.com](http://Booking.com) in 1996 startte? Google in 1997 en [Bol.com](http://Bol.com) 1999. In niet meer dan twintig jaar tijd zijn ze een grote rol in ons leven gaan spelen, elk op hun eigen manier. Na 2000 gingen we natuurlijk gewoon door: Wikipedia 2001, Facebook 2004 en de eerste iPhone 2007.

Als je dat op een rijtje zet zie je de impact die ICT als technologie op onze samenleving gehad heeft en nog steeds heeft. Nooit eerder heeft een technologie zo snel zo'n grote impact gehad.

Vroeg een verslaggever in 1998 mensen op straat of ze een mobiele telefoon wilden hebben, dan keken ze hem verbaasd aan. Altijd bereikbaar zijn? Wat een onzin. Vandaag de dag krijgen we serieuze ontwenningverschijnselen als we onze smartphone vergeten zijn.

Ik zie deze innovatiegolf als de eerste tsunami van ICT, nu staan we aan de vooravond van een tweede tsunami, namelijk Big Data. De eerste golf was een technologische, de tweede is een meer content gedreven golf. De derde golf is moeilijk te voorspellen. Ik verwacht dat het verzamelen van data op tal van terreinen voor doorbraken zal zorgen en dat we de fase in gaan van "personalized solutions" bijv in de gezondheidszorg. ICT-bedrijven zullen een dominante rol in de zorg gaan spelen.

In 2003 mocht ik de Diesrede verzorgen. Om inzichtelijk te maken wat de ICT ons allemaal zou brengen hebben we toen wat filmpjes gemaakt. Hier ziet u een verkorte versie van een van de filmpjes. Het leuke is dat alle zaken die toen voorspeld werden nu volledig geïntegreerd zijn in ons dagelijks handelen. Nog geen 15 jaar geleden.

Het werd intussen hoog tijd dat de informatica en ICT-onderzoekers zich beter zouden gaan organiseren, o.a. voor erkenning als wetenschapsgebied. In 1997 werd de eerste NOAG-I (Nationale OnderzoeksAgenda Informatica) door het toenmalige SION-bestuur, waar ik toen leiding aan mocht geven, vast gesteld. Later volgde de NOAG-ICT in 2005. Het was de eerste keer dat het veld zich achter een beperkt aantal wat grotere thema's schaarde. Het concept van de tweetrapsraket is toen geboren. Tijdens de eerste trap kwamen deze thema's op de wetenschapsbrede onderzoeksagenda, bij de tweede trap konden onderzoekers in competitie projecten voor deze ICT-thema's indienen. De taart werd groter en daardoor ook het aantal taartpunten voor onderzoekers.

Even een vooruitblik. De volgende initiatieven hebben ook substantieel bijgedragen aan het verkrijgen van een hogere organisatiegraad van het ICT-werkveld:

- 3TU.NIRICT
- ICTRegie
- COMMIT
- EIT Digital, toen nog ICT Labs,

- ICT Topteam
- COMMIT2Data

Bijna iedereen die hierbij betrokken is of is geweest zit in de zaal, heel erg veel dank voor jullie bijdragen.

Door de grote ICT-gerelateerde veranderingen in de samenleving begon de overheid zich te realiseren dat we ook in Nederland ICT-onderzoek moesten financieren, denk maar aan de vele FES en Bsik projecten, naast natuurlijk de vele projecten in de EU-kaderprogramma's. **Nu, in 2018, nu alle ICT FES-projecten zijn afgerond en de financiering is opgedroogd, neemt de vraag om onderzoek naar digitalisering weer toe. Er is dus nog hoop. Nu alleen de financiering nog.**

## 6. Op weg naar Data Science

Graag wil ik met jullie het opkomen van Data Science bespreken. Toen ik in 1976 als onderzoeker begon kon ik niet vermoeden dat de hoeveelheid data zo zou exploderen. We realiseren ons nu dat data iets heel belangrijks bevat, namelijk kennis. Deze explosie van data zal de samenleving, het bedrijfsleven en ook de wetenschap de komende jaren behoorlijk door elkaar schudden.

Jim Gray was een topwetenschapper op Database gebied en heeft de grondslagen gelegd voor data-intensieve wetenschap. Hij sprak over *The fourth paradigm of science*, na de eerste fase waarin experimenteren centraal stond, kwam de tweede fase van theorievorming, daarna de derde fase van simulaties, nu gevolgd door data-intensieve wetenschap, waarin grote hoeveelheden data gebruikt worden om mbv tools tot nieuwe inzichten en doorbraken te komen. Denk bijvoorbeeld aan de radio-astronomie.

Ik heb kort met hem mogen samenwerken toen ik Editor-in-Chief was van de VLDB Journal. Jim werkte toen voor Microsoft R&D, die een één-mans locatie in San Francisco had. Daar woonde Jim namelijk en daar lag zijn zeilboot. Voor zijn werk heeft hij de Turing Award gekregen.

Ik spreek over Jim in de verleden tijd. Van zijn zeiltocht om de as van zijn moeder voor de kust van San Francisco uit te strooien - is hij nooit terug gekomen. Binnen enkele dagen na dit voorval, zaten veel mensen satellietbeelden van de Stille Oceaan voor de kust voor San Francisco te bestuderen om te kijken wat er gebeurd kon zijn. In al zijn treurigheid is dit wel een mooi voorbeeld van data-intensief onderzoek: een zeilboot van 40-voet in duizenden satellietbeelden terug

vinden, is als een speld in heel veel hooibergen zoeken. Dat is de uitdaging van Data Science: kennis vinden in heel veel data.

Voor deze uitdaging worden Machine Learning technieken gebruikt.

In [dit filmpje](#) zien we een eenvoudig neuraal netwerk in de leerfase. Bijvoorbeeld: een auto leren remmen bij bepaalde waarnemingen. Onderaan ziet u de drie waarnemingsbolletjes: elk of een nul of een één. Bovenaan staat het output bolletje, met links de gewenste output (remmen of niet), ook een nul of een één. De bolletjes op de middelste rij worden telkens aangepast om de gewenste output te benaderen. Dat heet leren. Bovenaan ziet u dat tijdens het leerproces de kans op een fout besluit afneemt. Aan het eind zit de kennis, om het eenvoudig te houden, in de middelste bolletjes van het neurale netwerk. Nu kan de auto zelf beslissen om te remmen of niet. Dit is een extreem simpel voorbeeld van Machine Learning om een auto veilig rijgedrag bij te brengen.

Het aantal toepassingen dat kennis uit data wil halen en dan toe te passen is groot. We zien dan ook een groeiende vraag naar Data Scientists. Emile Aarts heeft laten zien dat er een mooie uitdaging voor de HBO's en de universiteiten ligt. Ook dank aan Willem Jonker die mbv EIT Digital op Europees niveau niet alleen deze uitdaging oppakt, echter, ook de business development kant meeneemt.

Laten we schouder aan schouder gezamenlijk hoogwaardige Data Science opleidingen verzorgen om de enorme vraag op te kunnen vangen. Laten we dat doen voordat onze data puur voor commerciële belangen mishandeld wordt. **Er zijn trouwens nog legio uitdagingen die nader onderzoek vereisen. Ik noem er een paar: eigenaarschap, transparantie, ethiek, rechtsgeldigheid, schaalbaarheid.**

## 7. Trends in Digitalisering

Zo, het wordt nu tijd om naar de toekomst te kijken. De data-explosie is de drijvende kracht achter de digitalisering van de samenleving. Dit wordt weer mogelijk gemaakt door de steeds verdere *decentralisatie* van het Internet. Deze decentralisatie maakt van ons *producenten* van data. De digitalisering van de samenleving werpt pas zijn vruchten af als ook het systeem van de samenleving mee *innoveert*. Over deze drie trends - *decentralisatie*, *producent* en *systeeminnovatie* - wil ik het nu met jullie hebben.

### 7.1 Decentralisatie van IT

We beginnen met de decentralisatie trend. In 1943 zei Thomas Watson: "I think there is a world market for maybe five computers". Het werden er toch wat meer



dan eerst gedacht. Toen ik 12 of 13 was mocht ik met mijn vader naar het computercentrum van de KLM om de nieuwe IBM 360, een main frame, te bewonderen. Gerrit Blaauw, emeritus hoogleraar van deze universiteit was mede-ontwerper van de 360-serie.

Na de grote mainframes kwamen de PCs. We hebben toen ook nog een tijd gehad van de nauwelijks te tillen portable PCs. In 2007 kwam de eerste iPhone. Na de smartphones komen nu de wearables. De markt richt zich nu mn op sport en fitness. Echter het moment dat de wearables geïntegreerd worden in uw kleding laat niet lang op zich wachten. De fashion wereld staat te popelen van ongeduld om deze nieuwe markt te openen.

Je smart phone is samen met wearables voor apps bezig om allerlei data te verzamelen: aantal stappen, hoeveel verdiepingen, GPS locatie, of je staat of niet, je hartslag, of je aan het sporten bent, hardlopen, fietsen, auto rijden etc. Je smart phone is dus eigenlijk een heel krachtige sensor. Je denkt misschien nog dat het een telefoon is. Ja, daar kunt je hem ook voor gebruiken (dan ben je trouwens wel heel erg ouderwets). En hij is miljoenen keren krachtiger dan de IBM 360 en ook heel veel kleiner (auto vs iets wat je in je hand kan houden). **Conclusie van de decentralisatie trend: bijna alle rekenkracht zit nu in de haarvaten van het Internet waar continue data opgehaald wordt. Decentralisatie ten top.**

Dit is een goed moment om even stil te staan bij wat mn de Elektrotechniek in 50 jaar aan miniaturisering heeft weten te bereiken. Als we de iPhone 5S met transistoren uit de jaren 60 zouden bouwen dan zou hij zo groot zijn als de Eiffeltoren en had je twee kerncentrales nodig om hem te laten werken (Bram Nauta, bedankt voor deze vergelijking). Bedenk je dat de transistor in 1947 is uitgevonden. Eén transistor in het lab. Vandaag draag je honderden miljoenen of zelfs miljarden transistoren bij je, in jouw smartphone, je wearables. Vaak nemen we deze innovatie als van zelf sprekend aan en verwachten we dat deze nog jaren lang doorgaat. Het is belangrijk te realiseren dat deze innovatie het resultaat is van een enorme academische en industriële inspanning. **Decentralisatie is alleen maar mogelijk geweest door deze miniaturisering.**

## 7.2 De consument wordt producent

De tweede trend is die van de consument die - bewust of onbewust - mbv smartphones en wearables data-producent wordt.

De vraag is hoe we hiermee om willen gaan. Hier twee perspectieven. Je zou kunnen zeggen, de data is van ons en we maken zelf wel uit wat we willen delen.

Of: laten we gewoon alles delen zodat we wetenschap en bedrijfsleven ruim baan bieden om de kwaliteit van leven verder te verbeteren.

In het eerste perspectief richten we ons op systemen die onze data beschermen. Volgens het tweede perspectief richten we ons op waardecreatie. Een fundamentele keus, die door de ethische en economische belangen, niet altijd eenvoudig is. Bijvoorbeeld, hoe gaan we om met eigenaarschap? Is de maker of de eigenaar van een wearable de eigenaar van de verzamelde data?

Steeds meer ICT-bedrijven verzamelen medische gegevens. En wellicht weten die ICT-bedrijven al eerder dat het niet goed met u gaat, nog voordat uzelf of uw huisarts het door heeft. Google kan van een patiënt al vrij nauwkeurig voorspellen of hij binnen 48 uur overlijdt. Interessant! Is er een meldingsplicht? En willen we dat wel weten? Steeds vaker zullen ICT-bedrijven voor medische doorbraken zorgen.

Nu we ons realiseren dat onze smart phones en wearables hoofdzakelijk data verzamelen, komt de typisch Nederlandse vraag op of ik wel moet betalen voor een apparaat dat mijn kostbare data verzamelt. Wellicht levert de data in een business model voor een health app wel meer op dan wat het kost om het apparaat te maken. Oftewel, waar verdient Apple over 5 jaar zijn geld mee, aan de winst op ICT-producten of aan de winst op de verzamelde data. Of, nog een stap verder, is over 5 jaar Apple nog steeds een ICT-bedrijf of is het een health bedrijf geworden met een ICT-divisie?

**Conclusie van de data produceren trend: hoe controleren we de data stromen en wat krijgen we als individu of als samenleving er voor terug?**

### **7.3 Systeeminnovatie**

De derde trend betreft systeeminnovatie. De decentralisatie in de ICT-wereld zien we ook terug in bijv de energiesector. We zien daar een hele transformatie van grote centrales naar zonnecollectoren op uw dak. Dus van centrale energie-opwekking naar decentrale. Dit is natuurlijk onderdeel van de energietransitie om de overgang van fossiele naar duurzame energiebronnen te realiseren. De traditionele consumenten worden ook hier producenten (vergelijkbaar als met de data).

Een anekdote. Er zijn veel kleine energiecoöperaties. 's Avonds in de bar zegt de directeur van het lokale zwembad dat hij het zwembad moet sluiten. Hij is niet meer in staat om de energierekening te betalen. Een boer die naast hem zit, zegt dat hij geen vergoeding krijgt voor de windenergie die hij opwekt. Door samen te

werken lossen ze elkaars probleem op. Dit is natuurlijk een bedacht verhaal. Echter, het roept wel wat vragen op: mag je elektriciteit onbeperkt terug het netwerk insturen, tegen welke vergoeding en wie bepaalt die vergoeding. Het laat wel zien dat er een update van onze wet- en regelgeving nodig is om decentrale energieopwekking mogelijk te maken.

Ook in de gezondheidszorg, net als in de energiesector, maken ICT-innovaties verbeteringen mogelijk die soms niet goed passen in ons toch enigszins verouderde en dicht getimmerde systeem van wet- en regelgeving. Om profijt te hebben van de digitalisering is vaak een systeeminnovatie noodzakelijk. Landen die geen regulering op micro-schaal hebben, zullen veel eerder profijt hebben van de digitalisering. Dus de ultieme vraag is waar willen we met Nederland over 5 jaar staan?

**Conclusie van de systeeminnovatie trend: laat de politiek leiderschap tonen door dmv systeeminnovaties ruimte te bieden voor ICT-innovaties die de kwaliteit van leven verbeteren.**

Kijkend naar alle drie de trends dan zien we dat door de decentralisatie het verzamelen van data mogelijk geworden is. Op basis van de kennis vergaard uit die verzamelde data worden business modellen aangepast waarbij vaak niet de creativiteit de bottleneck is maar de beperkingen in ons systeem van wet- en regelgeving dat veel langzamer verandert. Op alle drie vlakken is actie nodig om de kwaliteit van leven te verbeteren.

## 8. Boodschappen

Ik heb inmiddels mijn werkkamer ontruimd zodat Joost Kok, mijn opvolger, aan de slag kan, dus ik neem jullie weer mee terug naar mijn huiskamer.

Ik wil jullie nog een paar overwegingen meegeven. Als je van grote afstand naar de ontwikkelingen van het Internet kijkt zie je de volgende ontwikkelingen:

1. eerst een technologische ontwikkeling en we wisten niet precies wat we er mee konden doen. De eerste files werden uitgewisseld en de eerste schuchtere emails werden verstuurd;
2. daarna kwamen de zoekfuncties en werden de eerste producten via Internet verkocht;
3. gevolgd door de fase van de data-platformen (lees de Googles, social media);
4. nu zitten we in de fase van data verzamelen van individuen;

5. de volgende fase is om nieuwe diensten en producten op te zetten op basis van de nieuwe kennis vergaard uit de verzamelde data.

We staan aan de vooravond van de Big Data tsunami. Alle ingrediënten liggen op tafel. Wat doen we er mee? Wie gaan hier voordeel van hebben? In mijn ogen, degenen die begrijpen dat er kennis in data zit én hun business model aanpassen in een land waar de overheid dat faciliteert.

Nog drie overwegingen voor bij de borrel.

### **Overweging 1: data verzamelen**

Aan de basis van data verzamelen ligt de hele keten die begint met sensoren, sensing en sensornetwerken, gekoppeld via Internet of Things. De toepassingen liggen voor het oprapen: robotica, autonoom rijden, smart industry, personalized health care, duurzaamheid. De Universiteit Twente is sterk in de hele keten, dankzij samenwerking tussen nanotechnologie, ICT, bedrijfskunde en gedragswetenschappen en een scala aan toepassingsgebieden. Grijp je kans om dit geheel in een partnership met het bedrijfsleven in te brengen.

### **Overweging 2: data delen**

De samenleving wordt volledig overspoeld met data-platformen uit Amerika. Deze data-platformen worden gebruikt om data te verzamelen en te delen. Inmiddels weet Facebook meer over de burgers van bijvoorbeeld Enschede dan de ambtenaren hier doen. Het wordt echt tijd dat we in Europa investeren in Europese data-platformen, al is het maar om meer over onze eigen burgers te weten en om daarmee de kwaliteit van onze samenleving te waarborgen. Het heeft geen zin om een soort Europees Facebook te bouwen, maar grijp de kans om specifieke en goed afgeschermd platformen op het gebied van health te ontwikkelen. Als jullie het niet doen, doet bijv Amazon het wel.

### **Overweging 3: business model aanpassen**

In Frankrijk gaat de postbode tijdens de reguliere postbezorging -natuurlijk tegen een vergoeding - bij senioren langs om te kijken of het nog goed met hen gaat. Dát is begrijpen dat de asset van een postbedrijf het feit is dat ze dagelijks bij elke burger langs gaan en dát is ook waardecreatie. Door op een andere manier tegen de logistieke data aan te kijken ontstaat er een win-win situatie voor bedrijf en samenleving.

Mag ik u uitnodigen om tijdens de receptie voor uw eigen organisatie minstens één nieuwe dienst te bedenken? Ter inspiratie: is Apple over 5 jaar meer dan een ICT-

bedrijf, is de NS over 5 jaar meer dan een spoorwegbedrijf, is een universiteit over 5 jaar meer dan een opleidingsinstituut?

We onderschatten in Nederland nog steeds de impact die Digitalisering op de samenleving gaat hebben. Als het over ICT gaat, lijken we in een behoudende reflex te schieten. We moeten nu écht in actie komen, anders zijn wij straks de enige die niet begrijpen wat de kleurrijke Quipu's, waarmee ik mijn verhaal begon, ons te vertellen hebben.

Ter illustratie. Boudewijn Haverkort citeerde, tijdens mijn laatste vergadering met de informatica disciplineraad, uit mijn oratie. Hij concludeerde dat de oproep die ik daarin deed nog even actueel was.

Ik was gerustgesteld dat ik toen kennelijk geen onzin uitgekraamd heb, en tegelijkertijd teleurgesteld in het weinige dat we bereikt hebben. Het lijkt soms als een rimpeling in een woelige zee.

## **9. Dankwoord**

Tot slot wil ik jullie bedanken.

Op de eerste plaats wil ik een ieder bedanken die mij geïnspireerd en geholpen heeft met de voorbereiding van deze afscheidsrede. Ook wil ik de sprekers van het symposium bedanken. Zij hebben op een perfecte wijze aan gegeven dat er met Data Science nog een mooie uitdaging voor ons ligt.

Op de tweede plaats wil ik de universiteit met zijn eco-systeem bedanken voor de inspirerende omgeving die zij mij 33 jaar lang gegeven heeft. Een continue stroom van nieuwe studenten, nieuwe AIOs, nieuwe medewerkers, nieuwe projecten, nieuwe toepassingen, nieuwe rectorates en nieuwe structuren. Altijd heel uitdagend en heel inspirerend. Ik heb me nooit verveeld.

Op de derde plaats wil ik mijn collega's bedanken. Collega's, nationaal en internationaal, op mijn vakgebied en op aanpalende gebieden. Ik ben dankbaar voor de erkenning die ik, samen met mijn promovendi, als wetenschapper heb gekregen én ik ben blij met waardering voor de faciliterende rol die ik binnen en buiten de universiteit voor andere wetenschappers ver buiten mijn vakgebied heb mogen spelen.

Op de vierde plaats wil ik Arnold Smeulders bedanken. Het voelt goed om een kompaan te hebben zonder eigen agenda en volledig in dienst van de ICT community. Het was een plezier om met je samengewerkt te hebben. Ik heb veel

van je geleerd. Op deze plaats wil ik ook Amandus Lundqvist bedanken. Zonder hem waren we nooit zover gekomen.

Tot slot wil ik Berta en de kinderen bedanken voor het feit dat ze me een beter mens probeerden te maken, terwijl ik daar eigenlijk geen tijd voor had.

Ik heb gezegd.