

DE DATAGEDREVEN TOEKOMST.NL



DENKWERK

Hoe we vormgeven
aan onze toekomst
in de datagedreven wereld



DenkWerk is een onafhankelijke denktank die met krachtige ideeën bij wil dragen aan een welvarend, inclusief en vooruitstrevend Nederland. Hiervoor brengt DenkWerk Nederlanders bij elkaar die hun rijke kennis, ervaring en creativiteit willen inzetten om richting te geven aan brede maatschappelijke vraagstukken die hen na aan het hart liggen.

Om dat te bereiken doet DenkWerk haar eigen onderzoek, gebruikmakend van een breed netwerk van experts. Hiermee willen we vraagstukken grondig onderzoeken, structuur brengen en inspiratie aandragen voor acties of verder onderzoek.

Om bij te dragen aan het maatschappelijk debat en verandering in gang te zetten, streven we ernaar de resultaten van ons werk in het publieke domein te delen.

Barbara Baarsma
Bernard ter Haar
Boudewijn Wijnands
Frans Blom
Hans Wijers
Jaap Winter
Marelle van Beerschoten
Rianne Letschert
Robert van der Veeke (roulerend)

Met ondersteuning van Anna Wijers

Eerdere publicaties:
Artificial Intelligence in Nederland (juli 2018)
Arbeid in Transitie (januari 2019)
Onrust in Voorspoed (juli 2019)
Klein Land, Grote Keuzes (januari 2020)
Publiek en Effectief, het kan! (juli 2020)
De Online Wereld.nl (februari 2021)

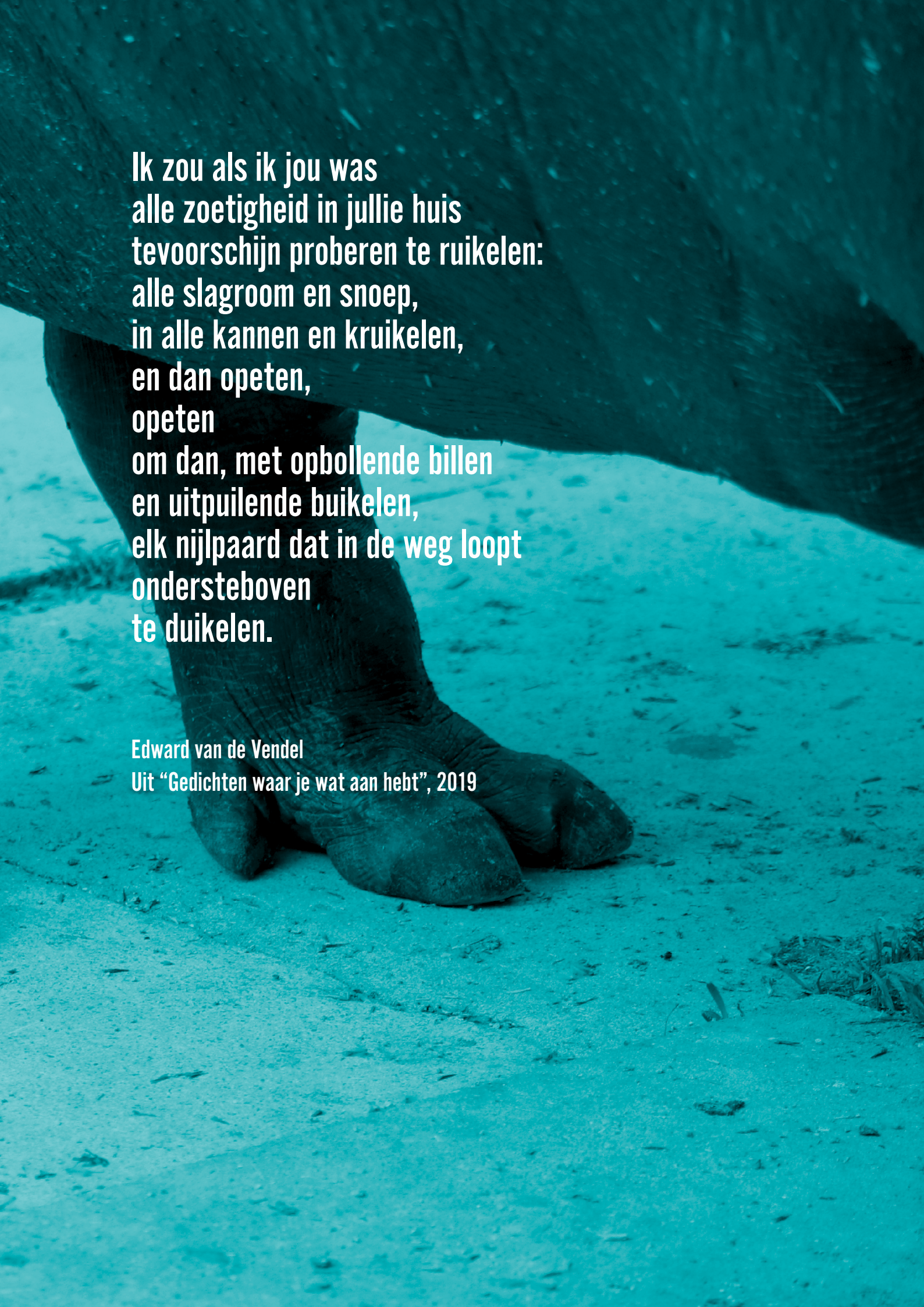
contact@denkwerk.online
www.denkwerk.online

De datagedreven toekomst.nl

Hoe we vormgeven aan onze toekomst in de datagedreven wereld

Februari 2021

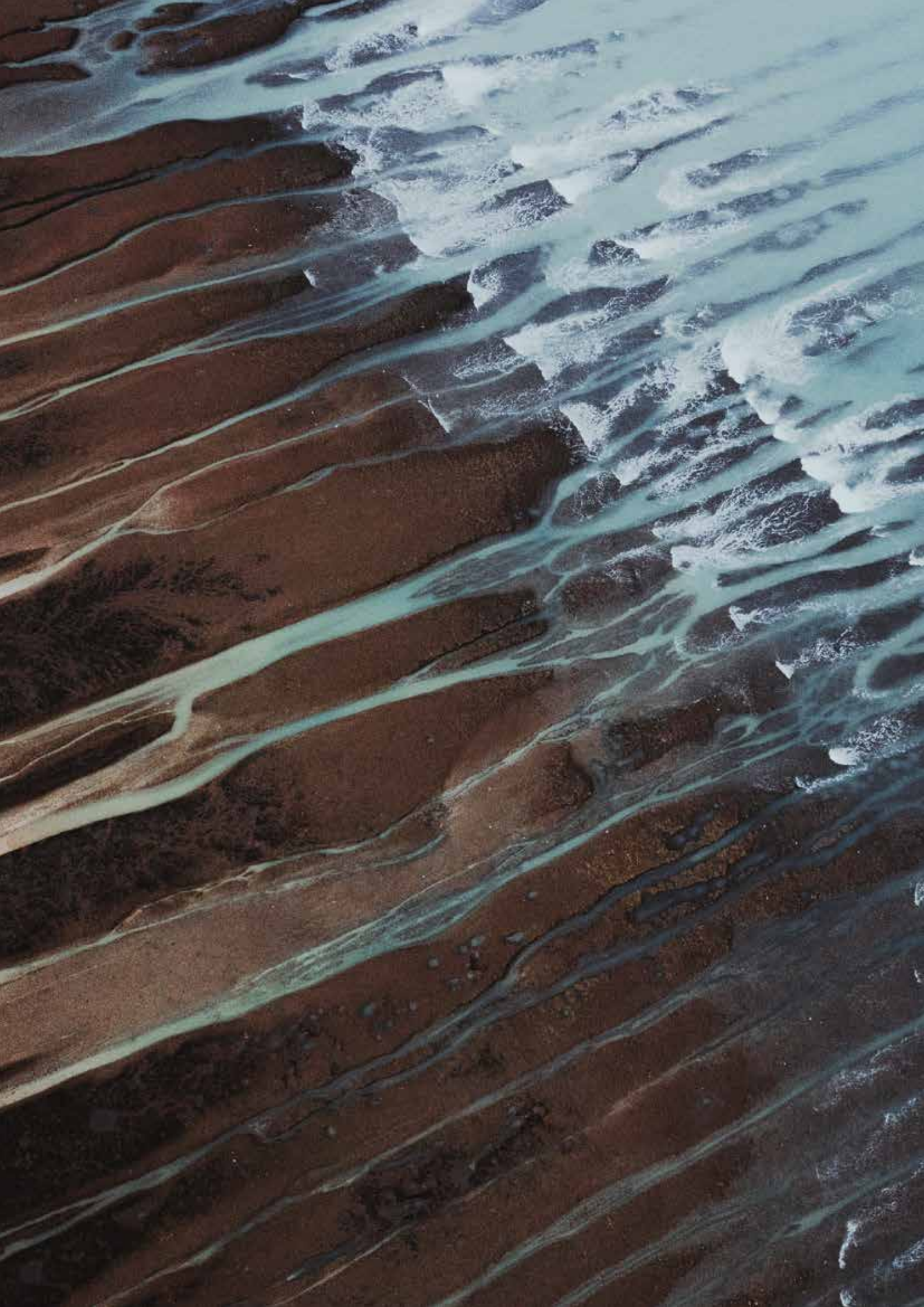




Ik zou als ik jou was
alle zoetheid in jullie huis
tevoorschijn proberen te ruikelen:
alle slagroom en snoep,
in alle kannen en kruikelen,
en dan opeten,
opeten
om dan, met opbollende billen
en uitpuilende buikelen,
elk nijlpaard dat in de weg loopt
ondersteboven
te duikelen.

Edward van de Vendel

Uit "Gedichten waar je wat aan hebt", 2019

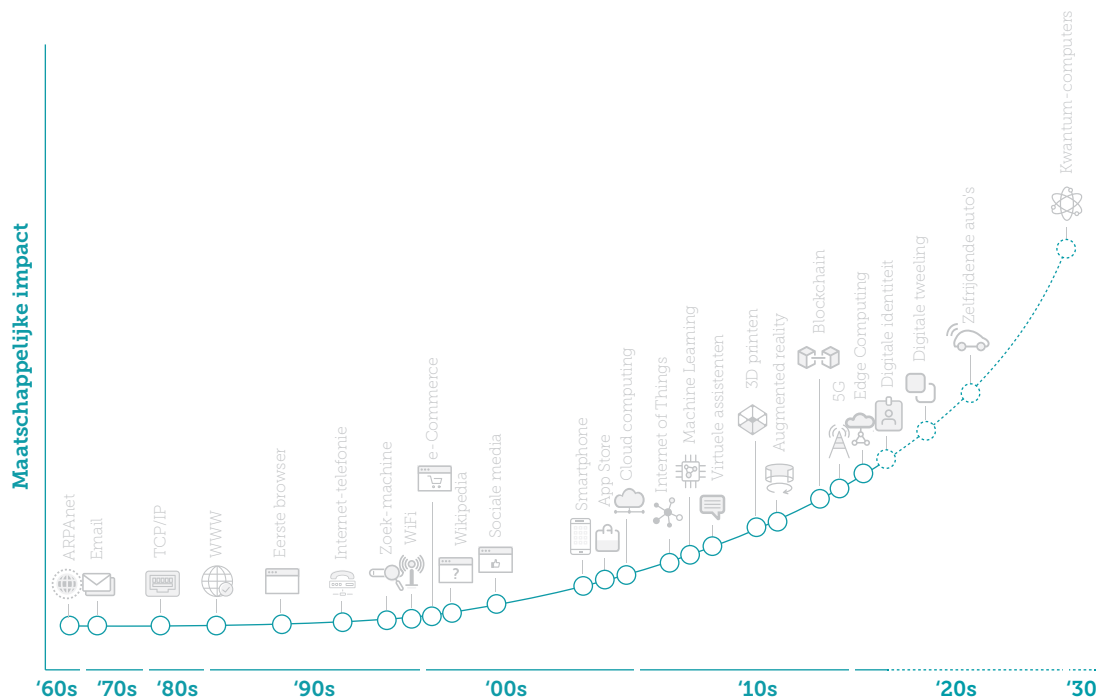


1. INLEIDING

De online wereld is de afgelopen tien jaar enorm gegroeid. In 2007 introduceerde Apple de eerste smartphone en in 2020 waren er wereldwijd 3,5 miljard smartphones in gebruik¹. De term 'cloud computing' werd in 2007 voor het eerst in een krantenartikel genoemd, om in sneltreinvaart uit te groeien tot een markt van meer dan 200 miljard euro in 2020². Platforms als Spotify, Uber en Netflix werden in Nederland geïntroduceerd in respectievelijk 2010, 2012 en 2013, en zijn vandaag onmisbaar voor miljoenen Nederlanders.

Ondanks de razendsnelle groei die de online wereld de afgelopen tien jaar heeft doorgemaakt, is de verwachting dat de groei in de komende tien jaar verder zal accelereren. Het internet evolueert van een netwerk van verbonden mensen naar een netwerk van verbonden objecten, ook wel 'internet der dingen' genoemd. Deze verbonden objecten kunnen data uitwisselen om daarmee 'intelligente' autonome beslissingen te nemen.

Figuur 1. De maatschappelijke waarde van het internet groeit nog altijd hard door



Noot: De meerderheid van de getoonde innovaties is niet in één jaar opgekomen maar over langere tijd gegroeid.
Bron: DenkWerk-analyse. De grafiek is een stijlvolle weergave.

Zo bewegen we richting een datagedreven wereld. Objecten produceren data met sensoren, voeden deze data aan het internet, en krijgen andere data terug om intelligente beslissingen mee te nemen. Hierdoor ontstaan 'intelligente huizen', met koelkasten die automatisch melk en eieren bestellen, verlichting die personen door de kamers in huis volgt en een koffiezetapparaat dat de sterkte van de koffie kan afstellen op de kwaliteit van iemands

Zeggen dat je “op het internet gaat” zal in de toekomst net zo vreemd klinken als vandaag zeggen dat je “op het elektriciteitsnet gaat” – zowel mens als apparaat is immers altijd verbonden

nachtrust. Er ontstaan 'intelligente steden', waar nutsvoorzieningen, infrastructuur, publieke diensten en de openbare ruimte gedataficeerd en verbonden worden ten behoeve van bijvoorbeeld het minimaliseren van files, het afstemmen van elektriciteitsverbruik op actueel aanbod en het bestrijden van criminaliteit. Ook biologische en fysiologische processen worden gedataficeerd en intelligent: sensoren bewaken de luchtkwaliteit, akkers reguleren zelf de chemische samenstelling van de grond en via chips in kleding worden ons hartritme, bloeddruk, spijsvertering en lichaamstemperatuur gemeten. Het internet wordt op deze manier het zenuwstelsel van de datagedreven wereld. Zeggen dat je “op het internet gaat” zal in de toekomst net zo vreemd klinken als vandaag zeggen dat je “op het elektriciteitsnet gaat” – zowel mens als apparaat is immers altijd verbonden.

Voor de totstandkoming van deze datagedreven wereld is het van essentieel belang dat data onbelemmerd kan stromen van plaatsen waar ze wordt geproduceerd naar plaatsen waar ze kan worden toegepast. Dit is niet vanzelfsprekend, met name voor toepassingen van data buiten de organisatie of zelfs de sector waar deze is gegenereerd. Organisaties gebruiken namelijk vaak hun eigen technische formats en taxonomieën, waardoor ze elkaars data niet kunnen gebruiken. Ook kan er een gebrek aan vertrouwen tussen partijen zijn om data uit te wisselen. Een bedrijf moet er immers op kunnen rekenen dat de wederpartij zijn data niet doorverkoopt aan concurrenten, of onzorgvuldig omgaat met privacygevoelige informatie van zijn klanten.

Big Tech bedrijven in de Verenigde Staten en China creëren vrije doorstroom van data door een breed spectrum aan datagedreven activiteiten te combineren binnen hun eigen ecosysteem^a. Op deze manier kunnen de 'dataconglomeraten' proposities ontwikkelen die zijn gebaseerd op vernieuwende toepassingen en combinaties van data. In een datagedreven wereld is dat een overweldigend concurrentievoordeel.

Om tegenwicht te bieden aan een dergelijk groot concurrentievoordeel, zullen Nederlandse en Europese bedrijven zelf ook toegang moeten organiseren tot veel en diverse data; niet door deze te kopen van dataconglomeraten maar door onderling data te delen. Datadelen is daarom een absoluut essentiële ontwikkeling die bepalend gaat zijn voor de concurrentiepositie van Nederland. Zonder datadelen kunnen Nederlandse bedrijven onvoldoende innoveren en raakt onze concurrentiepositie achterop vergeleken met de VS en China. Bovendien verliezen we sturing op de algoritmen die vormgeven aan onze leefwereld – denk bijvoorbeeld aan de risico-afwegingen die een zelfrijdende auto maakt. We behouden weliswaar ons wetgevende instrument om deze leefwereld te reguleren, maar de dagelijkse beslissingen binnen het wettelijk kader worden gemaakt door algoritmen van buitenlandse programmeurs in plaats van Nederlandse burgers en bedrijven.

^a Zie 'De online wereld.nl' voor een overzicht van de ecosystemen van dataconglomeraten.

Hoeveel regie Nederland kan voeren op de transitie naar een datagedreven wereld zal bepalend zijn voor de economische positie en de vormgeving van Nederland in de 21e eeuw.

De grootste belemmering voor het realiseren van datadeling op schaal is het gebrek aan urgentie dat wordt gevoeld door bedrijven en politiek. De overheid heeft een belangrijke verantwoordelijkheid in het aanjagen en regisseren van initiatieven ter bevordering van datadelen. Maar de kabinetsvisie uit 2018 getuigt met name van reactiviteit^b. Ook bedrijven hebben een belangrijke verantwoordelijkheid voor datadelen, want zij bezitten de meeste data en zijn het best gepositioneerd om innovatieve toepassingen te ontwikkelen. Maar bedrijven zijn vaak te afwachtend en zijn liever volger dan initiator. Wij herkennen hierin dezelfde afwachtende houding die overheid en bedrijfsleven twintig jaar geleden aannamen bij de vormgeving van de digitale infrastructuur, en tien jaar geleden bij de vormgeving van de platformsamenleving – wat geleid heeft tot de diverse problemen zoals beschreven in het rapport 'De online wereld.nl'. Deze afwachtende houding is nu extra gevaarlijk aangezien in de datagedreven wereld de wet van de groeiende voorsprong geldt.

Tegelijkertijd heeft Nederland een goede uitgangspositie om datadelen op schaal te realiseren. Er zijn enkele voorlopers met veel kennis, zoals de Data Sharing Coalition en de Nederlandse AI Coalitie. En Nederland is traditioneel goed in maatschappelijke samenwerking. Maar politiek en bedrijfsleven moeten wel met spoed in actie komen. Dataconglomeraten uit de VS en China bouwen hard door aan de datagedreven wereld, en wanneer zij eenmaal een voorsprong hebben zijn ze moeilijk in te halen. Een gezond gevoel van paniek is dus op zijn plaats.

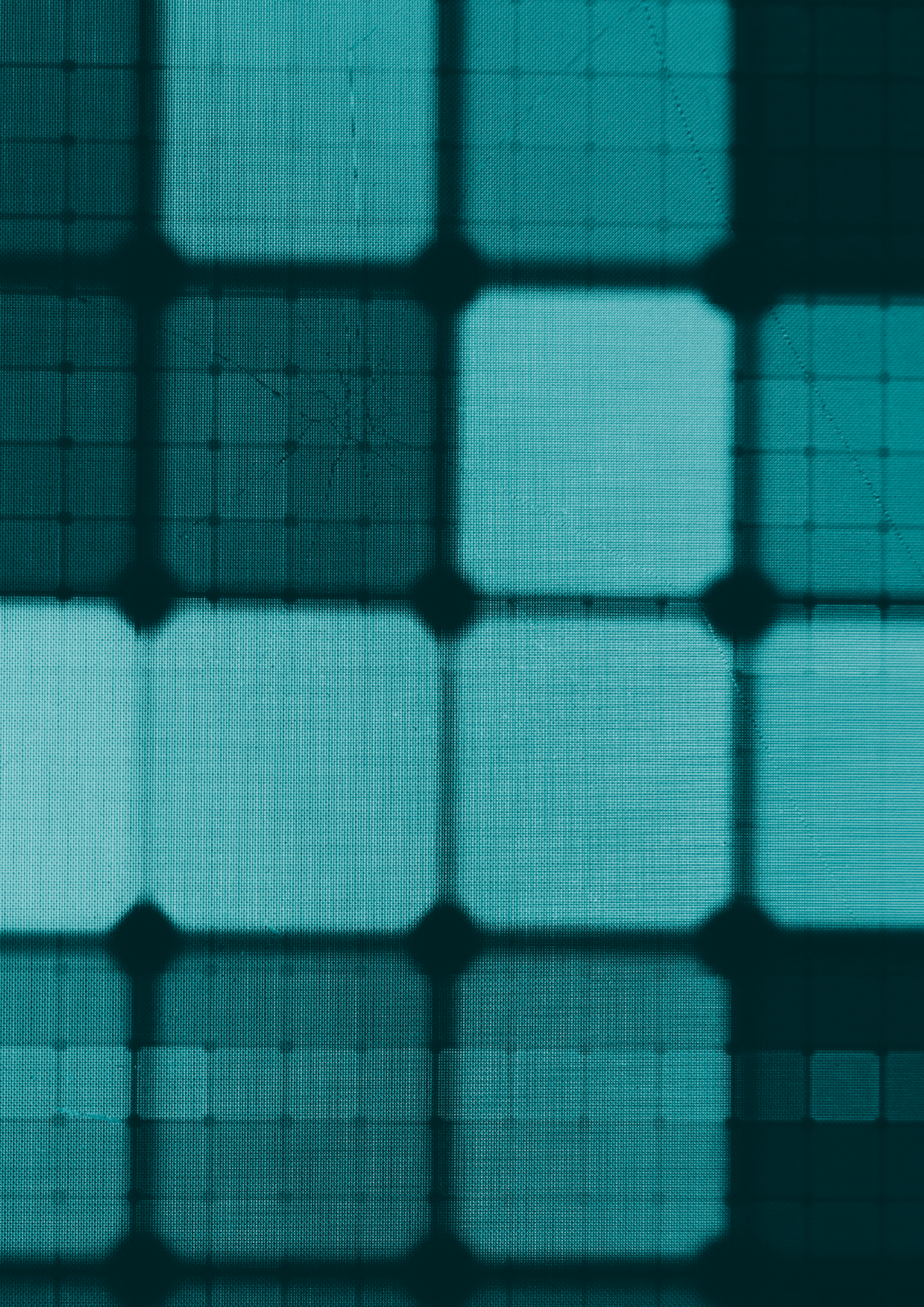
Het doel van dit rapport is om een gevoel van urgentie te creëren voor het thema datadelen. Hiertoe duiden wij wat er op het spel staat en geven we handvaten voor overheid en bedrijfsleven om in beweging te komen. In hoofdstuk 2 bespreken

we de weg naar en werking van de datagedreven wereld. Hoofdstuk 3 schetst de risico's die verbonden zijn aan een afwachtende houding: wat er gebeurt als China en de Verenigde Staten hun voorsprong op het gebied van artificiële intelligentie verder uitbouwen. Hoofdstuk 4 zet vervolgens uiteen welke randvoorwaarden nodig zijn voor datadeling op kleine schaal. Hoofdstuk 5 neemt in beschouwing hoe datadeling kan functioneren als een volwassen meerzijdige markt, waar velen elkaar kunnen vinden. Hoofdstuk 6 gaat in op de concrete stappen die nodig zijn om datadelen op schaal in een meerzijdige markt te realiseren: wat is nodig om adoptie van datadelen in de komende jaren fors te verhogen? Hoofdstuk 7 verkent waar

Nederland heeft een goede uitgangspositie om datadelen op schaal te realiseren. Maar politiek en bedrijfsleven moeten wel met spoed in actie komen

Nederland en Europa nu staan wat betreft datadelen en vat samen voor welke uitdaging overheid en bedrijfsleven staan. Tot slot beschrijft de appendix hoe bedrijven zich zullen gedragen in een wereld waar de condities zijn gecreëerd voor veilig, eenvoudig en grootschalig datadelen – en welke maatregelen noodzakelijk zijn om te zorgen dat alle bedrijven van groot tot klein eerlijk kunnen concurreren in een datagedreven economie.

^b In tegenstelling tot de European Data Strategy, die van meer proactiviteit en ambitie getuigt. Eind 2020 heeft de Europese Commissie met de Digital Governance Act ook een aantal goede voorstellen gedaan om datadelen te bevorderen.

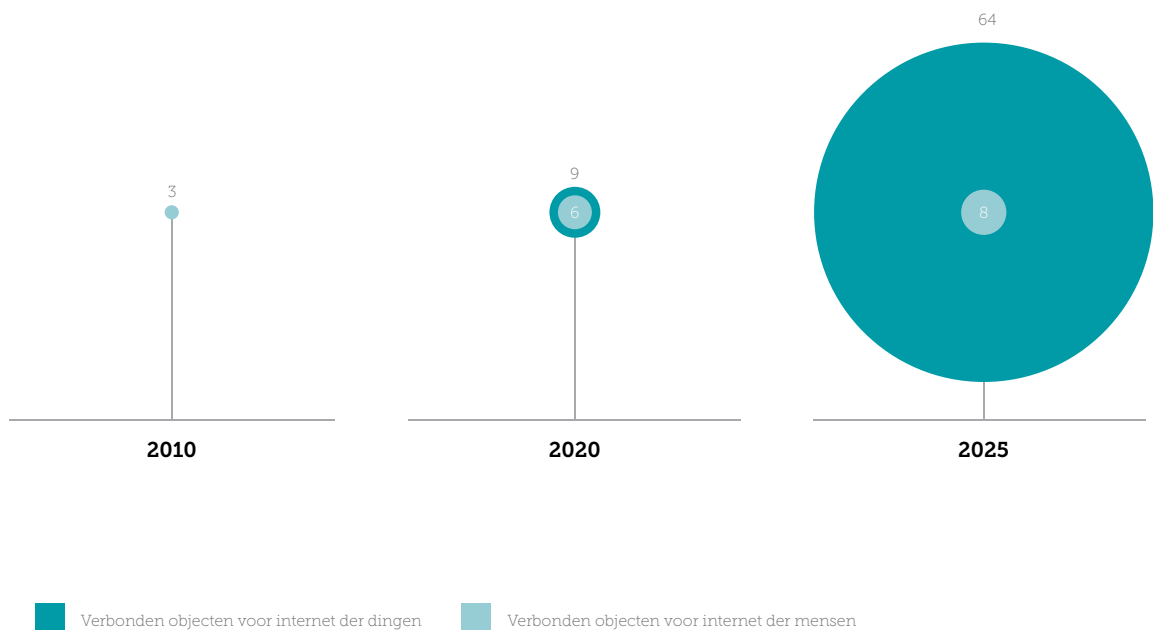


2. DE DATAGEDREVEN WERELD

De transitie naar het internet der dingen

De transitie die zich de komende tien jaar gaat voltrekken is de uitbreiding van een internet der mensen naar een internet der dingen. Traditionele verbonden objecten als computers, tablets en smartphones hebben een interface voor mensen, maar de groep verbonden objecten die dit niet heeft neemt razendsnel toe. We hebben het dan over bijvoorbeeld voertuigen, huishoudelijke apparaten en gebouwen, maar ook industriële installaties, medicatie, kleding en nutsvoorzieningen.

Figuur 2. De explosieve groei van het internet zal gedreven worden door verbonden objecten zonder menselijke interface (in miljarden objecten)



Bron: DenkWerk-analyse op basis van data van Gartner, IDC, GSMA Intelligence en Inside Intelligence.

De term 'internet der dingen' is al sinds 2010 in gebruik, toen Google begon met het indexeren van de fysieke wereld tijdens het Google Street-View project. Het werd snel overgenomen door gerenommeerde onderzoeksbureaus als IDC en Gartner, en in 2010 werd het ook door China genoemd als strategische prioriteit in hun vijfjarenplan. Ondanks dat de term 10 jaar oud is, staan we nog aan het begin van de transformatie. Er zijn vier ontwikkelingen die de transitie zullen drijven.

- **Verbetering van connectiviteit.** Het internet der dingen vereist een goede verbinding, omdat er enorme hoeveelheden data moeten worden verzonden. De uitrol van 5G is hierin een grote stap. Een andere belangrijke ontwikkeling is de opkomst van 'edge computing', de tegenhanger van de cloud^c. Bij edge computing worden berekeningen dicht bij de verbonden objecten uitgevoerd, in plaats van op afstand. Dit leidt tot snellere responstijd, minder belasting van internetverbindingen en een veerkrachtiger netwerk.
- **Toename van rekenkracht.** Er ontstaat een enorme hoeveelheid data in een datagedreven wereld, die in steeds geavanceerdere algoritmen wordt toegepast. Daarvoor is veel rekenkracht nodig. Er zijn nog steeds veel innovaties op het gebied van chips die rekenkracht vergroten. De grote revolutie waarop wordt geanticipeerd is de opkomst van de kwantumcomputer, waardoor de rekenkracht voor bepaalde toepassingen extreem toeneemt.
- **Innovatie in artificiële intelligentie.** Door ontwikkelingen op het vlak van artificiële intelligentie komen er steeds meer en effectievere toepassingen van data (zie eerder DenkWerk-rapport 'Artificial Intelligence').

- **Adoptie door bedrijven.** Naast deze technologische ontwikkelingen is het ook belangrijk dat bedrijven hun cultuur, bedrijfsmodel en operaties aanpassen aan de nieuwe mogelijkheden die het internet der dingen biedt^d.

Toepassingen van data in de datagedreven wereld

Objecten worden verbonden met het internet der dingen om data te genereren óf toe te passen. De data die wordt gegenereerd door verbonden objecten wordt pas waardevol als ze wordt toegepast; het genereren is dus nodig om de brandstof te creëren, terwijl het toepassen de waarde creëert. Er zijn vier belangrijke toepassingen voor data die is gegenereerd door verbonden objecten. Wij zullen deze toepassingen illustreren aan de hand van voorbeelden van een 'intelligente boerderij', ook wel aangeduid als precisielandbouw.

- **Analyseren.** Op de intelligente boerderij meten veldsensoren bijvoorbeeld de grondkwaliteit en de aanwezigheid van pesticiden. Door deze data van honderden boerderijen te bundelen en te combineren met andere datasets, kunnen boeren beter analyseren wat ze moeten doen om hun akkers optimaal vruchtbaar te houden.
- **Monitoren.** De veldsensoren op akkers kunnen bijvoorbeeld de vochtigheid van de aarde en de buitentemperatuur monitoren, om in combinatie met de weersvoorspellingen een bewateringsplan te maken voor de gewassen.
- **Voorspellen.** Data uit de oormerken van zwangere schapen kan worden gebruikt om te voorspellen wanneer ze hun lammeren zullen baren. Dit helpt bij de verzorging van de schapen.

^c Onderzoeksbureau IDC voorspelt dat in 2022 40% van de bedrijven hun cloudarchitectuur combineren met edge computing (IDC FutureScape: Worldwide IT Industry predictions, 2019).

^d Een gebrek aan digitaal leiderschap en flexibiliteit in het bedrijfsmodel zorgde ervoor dat de eerste golf van innovaties beperkt bleef tot relatief simpele toepassingen zoals preventief onderhoud en optimalisatie van logistiek (zie de studie 'Beyond predictive maintenance – the art of the possible with IoT' van BCG en Microsoft).

- **Automatiseren.** Wanneer verbonden objecten automatisch beslissingen kunnen nemen op basis van datastromen spreken we van intelligente objecten. Intelligente objecten kunnen beter presteren als hun algoritmen toegang hebben tot meer data dan dat ze zelf hebben gegenereerd. Zo maakt een zelfrijdende tractor bijvoorbeeld niet enkel gebruik van zijn eigen sensoren, maar ook van GPS-data en data van veldsensoren. Intelligente objecten vormen zo een geïntegreerd systeem.

Al deze toepassingen maken naast data ook gebruik van algoritmen. Een algoritme is een beslisregel waarmee een systeem automatisch beslissingen kan nemen. Vroeger waren deze beslisregels expliciet door de programmeur gecoördeneerd, maar tegenwoordig leren steeds meer algoritmen zichzelf de beste beslisregel aan, gegeven een doelstelling en inputdata. Dit fenomeen heet machine learning.

Het wordt in de toekomst een belangrijk concurrentievoordeel om nieuwe toepassingen te vinden voor data, vaak buiten de sector waar de data wordt gegenereerd. Bedrijven kunnen hierdoor betere proposities ontwikkelen dan concurrenten die zich beperken tot sectorspecifieke data. Figuur 3 toont voorbeelden van sectoren waar al toepassingen met cross-sectorale data zijn ontwikkeld. Uit onderzoek van het BCG Henderson Institute blijkt bijvoorbeeld dat ~85% van de toepassingen in intelligente steden gebruik zal maken van cross-sectorale data^e.

Innovatieve toepassingen van data bieden niet alleen economische kansen, maar spelen ook een sleutelrol in het oplossen van de grootste uitdagingen waar de mensheid voor staat. De Verenigde

85% van de toepassingen in intelligente steden zal gebruik maken van cross-sectorale data

Naties organiseren sinds 2016 jaarlijks het UN World Data Forum en hebben het Global Pulse programma gelanceerd – beide erop gericht om met private en publieke partijen kennis uit te wisselen over hoe data kan worden gebruikt om de Sustainable Development Goals te realiseren. Er worden reeds concrete successen geboekt op het vlak van financiële inclusiviteit, noodhulp bij rampen, ontbossing, volksgezondheid en klimaatverandering^f.

^e Zie het artikel 'The Risks and Rewards of Data Sharing for Smart Cities' van het BCG Henderson Institute.

^f Zie voor meer informatie het artikel 'Sharing Data to Address Our Biggest Societal Challenges' van het BCG Henderson Institute.

Figuur 3. Huidige toepassingsgebieden van het internet der dingen

SECTOR	
 Intelligente industrie	<ul style="list-style-type: none">• Start-up Guardhat produceert intelligente veiligheidshelmen die bewegende objecten detecteren waardoor werkplaatsongevallen met 20% kunnen verminderen.• Trumpf biedt zijn platform Axoom aan voor analyse o.h.g.v. preventief onderhoud van industrieel gereedschap.
 Intelligente gezondheidszorg	<ul style="list-style-type: none">• Medtronic heeft een product ontwikkeld dat het glucoseniveau van diabetespatiënten monitort en deelt met een arts.• Proteus Digital Health produceert medicatie met inslikbare sensors die fysiologische data stuurt naar de telefoon van een patient om de effectiviteit van de behandeling te meten.
 Intelligente energie	<ul style="list-style-type: none">• Een deel van de Volkswagen auto's in Europa verzamelt weergegevens, welke worden gedeeld met TenneT, een transmissienetbeheerder. TenneT gebruikt de data om te voorspellen waar en hoeveel zonne-energie het elektriciteitsnet binnenstroomt.
 Intelligente mobiliteit	<ul style="list-style-type: none">• Surtrac ontwikkelt software om verkeersdata te verzamelen en in real-time te analyseren zodat stoplichten zodanig van kleur verspringen dat files in de stad geminimaliseerd worden.• Nederlandse start-up Nedap ontwikkelt software om auto's in een stad real-time te geleiden naar beschikbare parkeerplaatsen.
 Intelligente retail	<ul style="list-style-type: none">• Rebecca Minkoff gebruikt intelligente spiegels in pashokjes om klanten tijdens het passen te laten zien welke producten beschikbaar zijn en in welke maten. Klanten kunnen artikelen laten bezorgen naar het pashokje.• De Coca-Cola Freestyle frisdrankmachine dataficeert klantvoorkeuren in groot detail voor het ontwikkelen van nieuwe recepten.
 Intelligente verzekeringen	<ul style="list-style-type: none">• Het Amerikaanse bedrijf Farmers Insurance gebruikt onbemande drones in plaats van menselijke inspecteurs om schade aan woonhuizen te beoordelen na natuurrampen, waardoor vergoedingen sneller worden uitgekeerd.• Metromile biedt een prijs-per-kilometer verzekering op basis van gebruiksgegevens van de auto.
 Intelligente natuur	<ul style="list-style-type: none">• Het Krueger Park in Zuid-Afrika heeft met Cisco een programma opgezet om met een sensorsysteem 24 uur per dag het park te surveilleren, wat leidde tot een 96% daling in stroperijen.• Start-up The Yield gebruikt sensordata om ziekten in oesterteelt te bestrijden en een 30% sterftereductie te realiseren.
 Intelligente steden	<ul style="list-style-type: none">• BigBelly levert afvalverwerkingsdiensten aan steden door intelligente en verbonden afvalbakken te installeren. Het afval wordt hierdoor pas opgehaald wanneer de afvalbakken vol zijn, volgens een automatisch geoptimaliseerde route. BigBelly is actief in meer dan 55 landen waaronder Nederland.
 Intelligente huizen	<ul style="list-style-type: none">• Philips Hue is een intelligent verlichtingssysteem dat het type licht afstelt op bijvoorbeeld het moment van de dag en de bezigheden van de bewoners, zodat de leefervaring verbeterd en de lampen minder energie gebruiken.• Wattio is een Spaanse start-up die helpt om energiegebruik van huishoudelijke apparaten te monitoren en controleren.

Bron: BCG Henderson Institute en StartUS Insights.

Schaalvoordelen en synergievoordelen van data

Bedrijven met toegang tot meer en rijkere data zullen groot voordeel behalen in deze datagedreven economie. Dit komt door schaalvoordelen en synergievoordelen⁹ in het gebruik van data.

Met toegang tot grotere hoeveelheden relevante data kunnen algoritmen zichzelf beter trainen, waardoor schaalvoordelen ontstaan (zie ook het DenkWerk rapport 'Artificial intelligence'). Deze beter getrainde algoritmen vormen een aantrekkelijker product, dat hiermee een grotere gebruikersgroep aantrekt, wat weer leidt tot een grotere toestroom van trainingsdata – dit is een positieve

feedbackloop die de koploper in een datagedreven markt aan kop houdt. Google Search is een goed voorbeeld. Omdat Google Search de beste zoekmachine was, werd Google Search het meest gebruikt waardoor Google ook de meeste data verkreeg over zoekgedrag. Deze data werd gebruikt om de zoekmachine verder te verbeteren, waardoor de voorsprong van Google nog groter werd – enzovoort.

Figuur 4 laat zien hoeveel data er in de VS, Europa en China gegenereerd wordt uit economische activiteit. Europa loopt daarin achter. China loopt sterk voor met het genereren van data uit het internet der dingen. Daar komt bovenop dat een groot deel van de data die in Europa gegenereerd wordt in handen is van Amerikaanse en Chinese

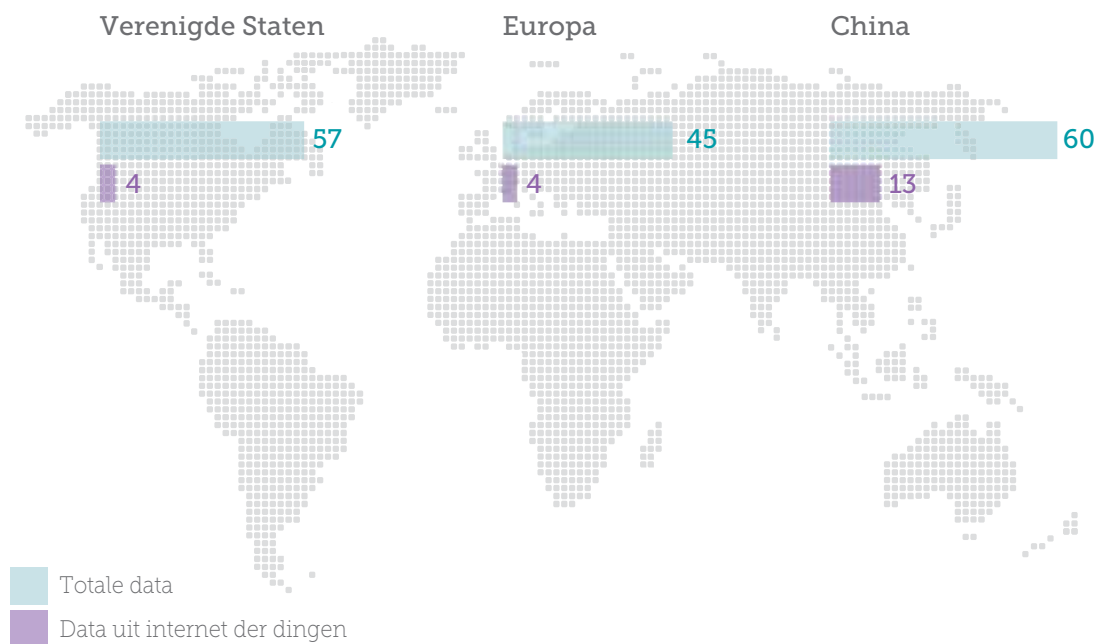
⁹ Synergievoordelen ontstaan wanneer twee economische activiteiten efficiënter of beter worden door ze te combineren, waardoor de combinatie waardevoller is dan de som van de delen.

bedrijven die actief zijn in de Europese online wereld. Relatief heeft Europa dus minder toegang tot data op schaal.

Naast schaalvoordelen gelden er ook sterke synergievoordelen voor data. Door data uit verschillende domeinen te bundelen kan een algoritme zijn beslissingen baseren op meer verschillende factoren. Het heeft meer informatie tot zijn beschikking en neemt dus betere beslissingen. Een voorbeeld zijn de algoritmen van hedgefondsen. In plaats van enkel uit te gaan van traditionele data zoals prijzen en economische statistieken, wordt nu op grote schaal geïnvesteerd in algoritmen die ook gebruik maken van andere databronnen zoals content op sociale media. Deze algoritmen kunnen voor een groot aantal markten en bedrijven veranderingen opmerken in consumentenvertrouwen en trends, waarop het hedgefonds haar investeringsportfolio kan aanpassen.

Onze toekomst is een datagedreven samenleving waarin intelligentie verweven raakt met onze fysieke werkelijkheid. Het internet is in deze datagedreven samenleving het zenuwstelsel dat alles met elkaar verbindt. Maar is het internet zoals we dat vandaag kennen wel voorbereid op die rol? Wat zijn de implicaties van de problemen in de digitale infrastructuur, zoals beschreven in het rapport 'De online wereld.nl', in een datagedreven wereld? En stel, de online wereld van vandaag wordt gerepareerd door alle aanbevelingen uit 'De online wereld.nl' te implementeren, zijn we daarmee ook voldoende voorbereid op de datagedreven wereld, of schuilen er nieuwe risico's? Hoofdstuk 3 beschrijft hoe de transformatie eruit zal zien als Nederland afwachting blijft en onvoldoende de regie pakt.

Figuur 4. De hoeveelheid data die in de VS, Europa en China gegenereerd wordt uit economische activiteit (in TB per miljoen euro BNP)



Bron: World Bank, International Data Corporation



3. HET GEVAAR VAN AFWACHTEN

In 2002 stuurde Jeff Bezos, de oprichter van Amazon, een memo naar al zijn werknemers met een simpele boodschap. Alle software functionaliteiten die door het bedrijf werden gebouwd, en alle datastromen die door het bedrijf werden gegenereerd, moesten zo worden ingericht dat ze geschikt waren voor interne deling. Met deze gouden regel moest het altijd mogelijk zijn om data van de ene bedrijfstak toe te passen in proposities elders in het bedrijf. Iedereen die zich niet aan deze regel hield zou op staande voet worden ontslagen. Amazon is vandaag een bedrijf dat zich in honderden verschillende markten roert en vaak succesvol is door datastromen te bundelen^h. Het Bezos-mandaat wordt gezien als een cruciale oorzaak van dit succes.

Hoofdstuk 2 beschrijft hoe schaalvoordelen en synergievoordelen van data de sleutel zijn om te winnen in de datagedreven economie. Het Bezos-mandaat is een goede illustratie van de manier waarop GAFAM – een acroniem voor Google, Apple, Facebook, Amazon en Microsoft – data uit honderden verschillende bedrijfsactiviteiten combineert om nieuwe datagedreven proposities te ontwikkelen. Zo behaalt GAFAM schaalvoordelen en synergievoordelen van data binnen hun eigen bedrijf. We spreken daarom van ‘dataconglomeraten’.

Dataconglomeraten gaan voorop in de transitie

Dataconglomeraten verzamelen zoveel mogelijk data en gebruiken en hergebruiken deze data voortdurend om superieure algoritmen te ontwikkelen in allerlei markten. Zo heeft Google bijvoorbeeld niet alleen de sterkste zoekalgoritmen voor Google Search, maar is het nu ook leidend in de race naar een zelfrijdende auto. In een artikel van de Financial Times uit februari 2019 wordt een vergelijking gemaakt tussen zelfrijdende auto's van verschillende fabrikanten. Als maatstaf wordt gekeken naar het aantal kilometers dat ze gemiddeld

kunnen rijden op de openbare weg voordat een menselijke ingreep nodig is. Voor Mercedes-Benz was dit 2,4 kilometer en voor Toyota 4 kilometer. Google's zusterbedrijf Waymo kon in 2018 gemiddeld 17.732 kilometer rijden voordat een menselijke ingreep nodig was, en in 2019 zelfs 21.274 kilometer. De belangrijkste reden voor dit enorme verschil is dat Google gebruik maakt van data uit andere bedrijfsonderdelen. Zo heeft Google met projecten als Google StreetView en Google Earth een zeer gedetailleerde indexering van onze fysieke werkelijkheid gemaakt, met informatie zoals de breedte van straten. Daarnaast hebben ze zeer geavanceerde artificiële intelligentie ontwikkeld, die ze kunnen hergebruiken voor zelfrijdende auto's. Zo maakt Waymo gebruik van veel AI-technieken die zijn ontwikkeld door DeepMind, het door Google gekochte bedrijf dat in 2016 het nieuws haalde door de wereldkampioen in het Chinese bordspel Go te verslaan. Google loopt voorop met zelfrijdende auto's, maar er zijn ook dataconglomeraten actief op andere innovatiefronten in de auto-industrie zoals elektrisch rijden, spraakassistentie en online dienstverlening in de auto. Het is niet voor niets dat de nieuwe Ford CEO bij zijn aantreden in augustus 2020 stelde: *“We weten dat onze concurrentie vandaag uit Amazon, Baidu, Tesla en Apple bestaat”*.

Zelfrijdende auto's van Mercedes-Benz en Toyota kunnen respectievelijk 2,4 en 4 kilometer rijden zonder menselijke ingreep, vergeleken met 21.274 kilometer voor de zelfrijdende auto van Google

^h Zie hoofdstuk 6 van 'De online wereld.nl'.

Figuur 5. Het Bezos-mandaat

All teams will henceforth expose their data and functionality through service interfaces.

Teams must communicate with each other through these interfaces.

There will be no other form of interprocess communication allowed: no direct linking, no direct reads of another team's data store, no shared-memory model, no back-doors whatsoever. The only communication allowed is via service interface calls over the network.

It doesn't matter what technology they use. HTTP, Corba, Pubsub, custom protocols - doesn't matter.

All service interfaces, without exception, must be designed from the ground up to be externalizable. That is to say, the team must plan and design to be able to expose the interface to developers in the outside world. No exceptions.

Anyone who doesn't do this will be fired.

Thank you; have a nice day!

-JEFF BEZOS

Een ander voorbeeld van een markt waar dataconglomeraten winnende proposities bouwen is in de gezondheidszorg. The Economist publiceerde op 5 december 2020 een artikel genaamd 'The Dawn of Digital Medicine', waarin wordt beschreven hoe dataconglomeraten steeds actiever worden in de gezondheidszorg. Amazon's stemassistent Alexa kan bijvoorbeeld iemands hoest analyseren om de waarschijnlijke oorzaak ervan te bepalen. In november 2020 is Amazon zijn eigen digitale apotheek gestart. Ook het Chinese dataconglomeraat Alibaba is actief in de sector, onder de naam AliHealth. Apple's smartwatch heeft een verscheidenheid aan klinisch bewezen medische toepassingen, en er zijn bijna 50 duizend gezondheidsapps beschikbaar in de App Store. Google's zusterbedrijf Verily is een onderzoekslab, toegewijd aan de levenswetenschappen. Door de initiatieven van dataconglomeraten in de gezondheidszorg stelt Pamela Spence van EY dat deze sector evolueert van "een klinische wetenschap ondersteund door data naar een datawetenschap ondersteund door clinici"³.

We zien dus al dat dataconglomeraten met datagedreven oplossingen gaan concurreren buiten hun traditionele markten. Ze hebben toegang tot een enorme hoeveelheid data, omdat ze actief zijn in de digitale infrastructuur: de slagaders van datastromen. Hieromheen hebben ze ecosystemen gebouwd die op allerlei gebieden data genereren, wat leidt tot een reusachtig data-universum. Door deze data toe te passen voor analyses, monitoring, voorspellingen en automatiseringen hebben zij de sleutel tot winnende proposities in alle sectoren. Maar wat betekent dit voor Europese bedrijven? Deze vraag stellen we eerst in het 'status quo scenario' waar GAFAM vrij is haar ecosystemen te ontplooiën. Daarna stellen we de vraag in het scenario dat bedrijven als GAFAM strak worden gereguleerd volgens de aanbevelingen van 'De online wereld.nl' – dit noemen wij het 'gereguleerde scenario'.

Het status quo scenario – dataconglomeraten als poortwachters van de economie

Dataconglomeraten hebben een unieke mogelijkheid om data uit verschillende markten te verzamelen en bundelen voor vernieuwende proposities. In het scenario waarin ze deze strategie ongeremd mogen volgen, krijgen ze de sleutel in handen voor cruciale concurrentievoordelen in alle sectoren. Net zoals GAFAM nu poortwachtersposities vervult in internetmarkten, zal ze dat straks in een datagedreven wereld ook doen in bijvoorbeeld de auto-industrie, gezondheidszorg en energiesector. GAFAM kan deze poortwachterspositie met twee verschillende strategieën te gelde maken: door zelf een propositie te ontwikkelen en in de markt te zetten, zoals we nu zien bij zelfrijdende auto's en digitale gezondheidszorg, of door toegang tot de data en de consument tegen hoge commissies te verkopen aan traditionele spelers,

Net zoals GAFAM nu poortwachtersposities vervult in internetmarkten, zal ze dat straks in een datagedreven wereld ook doen in bijvoorbeeld de auto-industrie, gezondheidszorg en energiesector

zoals Apple doet met de HomeKit propositie voor intelligente huizen. Dit toekomstbeeld is extreem problematisch voor Europa en voor Nederland. Het betekent dat we economisch gezien ver achter raken op de Verenigde Staten en China. En daarnaast leidt het ertoe dat we een deel van de zeggenschap verliezen over de vormgeving van onze leefwereld.

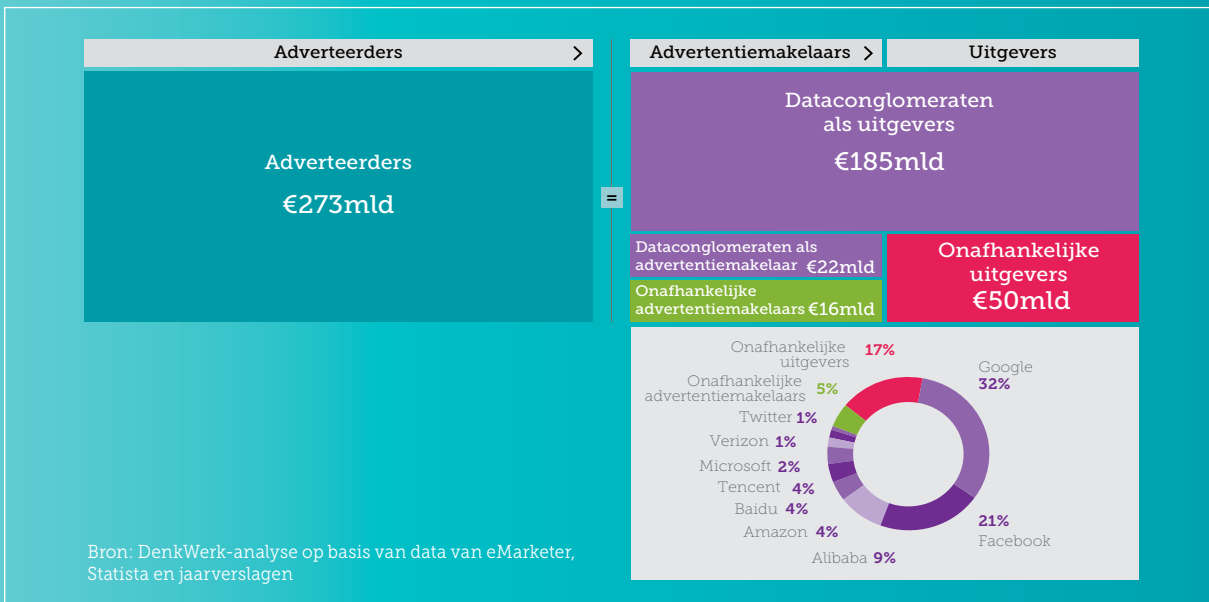
De dominantie van dataconglomeraten in de online advertentiemarkt

Een markt waar dataconglomeraten reeds een groot deel van de waarde naar zich toe hebben getrokken is de online advertentiemarkt. Van alle inkomsten die gegenereerd worden uit online advertenties ontvangen dataconglomeraten bijna 80% (zie figuur 6). Dataconglomeraten verdienen inkomsten aan advertenties in de hoedanigheid van uitgever en advertentiemakelaar. Als uitgever tonen ze een advertentie op hun eigen applicatie, zoals Google op Youtube, Microsoft op LinkedIn of Facebook op Instagram. Dataconglomeraten verdienen jaarlijks €185 miljard op deze manier. Dit grote aandeel hebben ze te danken aan hun dominante posities in de markten van bijvoorbeeld sociale media, zoekmachines, videoplatforms of nieuws. Maar dataconglomeraten verdienen ook op een andere manier geld in de online advertentiemarkt: door op te treden als advertentiemakelaars. Met name Google en Facebook bieden verschillende diensten aan zodat de juiste adverteerders en uitgevers met elkaar verbonden worden. Google doet dat via zijn 'Google Network Properties' (voorheen was dit met name AdSense) en Facebook via zijn Audience Network platform. Ze kunnen een groot deel van de marge afromen doordat ze de beste data hebben: dataconglomeraten als Google en Facebook weten beter

dan adverteerders of uitgevers welke advertenties effectief zullen zijn bij welke bezoeker.

Dataconglomeraten claimen een groot deel van de online advertentiemarkt door te concurreren als uitgever én door marge af te romen bij andere uitgevers. Het is dus een goede illustratie van de manieren waarop dataconglomeraten hun poortwachtersposities te gelde kunnen maken. In de toekomst kan de situatie nog extremer worden. Onafhankelijke uitgevers en advertentiemakelaars zijn vaak van 3rd party cookies afhankelijk om data te verzamelen van gebruikers zodat zij hen kunnen koppelen aan de juiste advertentie (en dus een hoge prijs kunnen vragen aan adverteerders). Maar de browsers Safari en Firefox hebben 3rd party cookies reeds verboden en Google heeft aangekondigd om binnenkort op Chrome hetzelfde te doen. Dataconglomeraten kunnen zelf data blijven verzamelen, omdat zij vaak betrokken zijn als 1st party via diensten in hun enorme ecosysteem. Google is bijvoorbeeld betrokken als 1st party wanneer een website gebruik maakt van AdSense, een Youtube plug-in, een Maps plug-in of een loginmechanisme via het Google-account. De datavoorsprong van dataconglomeraten in de advertentiemarkt zal dus nog verder groeien.

Figuur 6. Dataconglomeraten hebben een groot deel van de online advertentiemarkt naar zich toegetrokken en romen marge af van onafhankelijke uitgevers



Wanneer Europese bedrijven voor de essentiële grondstof 'data' afhankelijk worden van dataconglomeraten, betekent dat grote schade voor de concurrentiepositie van de Europese economie. 'Data is de nieuwe olie', schreef The Economist in mei 2017. En net zoals controle over data zal leiden tot poortwachtersposities, heeft controle over olie ook altijd geleid tot poortwachtersposities. Inderdaad zijn landen die over veel olie beschikken daar veelal steenrijk van geworden. Toch is poortwachterschap over data een veel belangrijker economisch voordeel dan poortwachterschap over olie. Dit komt ten eerste omdat oliebezit van nature gefragmenteerd is, terwijl controle over data een natuurlijke tendens heeft naar concentratie¹. Dat betekent dat aanbieders van data een sterkere positie hebben en meer waarde in de waardeketen naar zich toe kunnen trekken dan aanbieders van olie. Nog belangrijker is dat olie een simpele commodity is en data niet. Olie heeft enkel waarde als voedingsbron voor economische operaties; het dient om 'de fabriek' draaiende te houden. Data daarentegen is niet alleen een belangrijke input voor operaties, maar ook voor innovatie. Olie-poortwachters acteren in het deel van de waardeketen waar extreme prijsconcurrentie heerst door gebrek aan differentiatie. Datapoortwachters acteren in het deel van de waardeketen waar innovatie plaatsvindt en concurrentievoordelen worden bepaald. Dataconglomeraten zijn dus niet alleen poortwachter van een simpele commodity, maar ook van innovatie en concurrentievoordelen. Dataconglomeraten kunnen daardoor een groot deel van de marge afvangen in waardeketens. Een voorbeeld is Google Plex, een nieuwe dienst voor de financiële wereld. Het Financieele Dagblad schreef daarover het volgende⁴: "Google verzorgt de gehele techniek: bouwt de app, zorgt voor fraai design met slimme zoekfuncties en regelt de beveiliging tegen cybercriminelen. De bank blijft gewoon bank, teruggebracht tot kern van bankieren: geld inlenen tegen een gunstiger tarief dan uitlenen." Met andere woorden: bankieren wordt teruggebracht tot het houden en managen van een balans, een commodity. Alle innovatie en

differentiatie wordt afgevangen door Google, en daarmee ook de waarde in de waardeketen. In een datagedreven wereld die gedomineerd wordt door dataconglomeraten, verandert Europa in een win-gewest voor de Verenigde Staten en China.

Dataconglomeraten zijn niet alleen poortwachter van een simpele commodity, maar ook van innovatie en concurrentievoordelen

De tweede grote zorg in het status quo scenario – waarin dataconglomeraten ongeremd hun ecosystemen kunnen ontplooiën – is dat Europa een deel van de zeggenschap verliest over de vormgeving van zijn leefwereld. Als de digitale infrastructuur van onze intelligente steden, intelligente bedrijven en intelligente natuur in handen is van buitenlandse entiteiten, verliezen we daarover een deel van de controle. Dit is problematisch omdat Europa afhankelijk wordt van buitenlandse bedrijven en dus geen goed uitgangspunt heeft in geval van bijvoorbeeld (handels)conflicten of onderhandelingen. Op subtielere wijze leidt het er ook toe dat Europa sturing verliest op de alledaagse vormgeving van zijn wereld. Net zoals we in 'De online wereld.nl' zagen hoe platformmechanismen onderdeel worden van de maatschappelijke orde, zal dit straks ook gelden voor algoritmen in de datagedreven wereld. Er wordt niet meer in Europa maar van buitenaf bepaald hoe bijvoorbeeld de algoritmen werken die het onderwijs personaliseren, bepalen welke risico's zelfrijdende auto's nemen en diagnosticeren wanneer iemand ziek is en welke leefstijlinterventies nodig zijn voor genezing. We kunnen natuurlijk nog wel wetten invoeren, maar daarmee stellen we enkel kaders.

¹ Het aanbod van olie kent natuurlijke fragmentatie, doordat het geografisch verspreid is en doordat de schaal van ontginners in bedwang wordt gehouden door operationele complexiteit die leidt tot schaalnadelen. Dataconglomeraten kunnen 'schalen zonder massa' en dus een veel hogere marktconcentratie behalen (zie 'De online wereld.nl').

Het is minstens zo belangrijk wie binnen die kaders de dagelijkse beslissingen nemen. Nu zijn dat nog Europese bedrijven en burgers, straks zijn dat voor een belangrijk deel intelligente objecten gecodeerd door buitenlandse programmeurs. Europa zou in dit toekomstscenario kunnen worden gezien als

Om onze concurrentiepositie te beschermen in de transitie naar de datagedreven economie, is het faciliteren van datadelen op schaal een absolute must

een intelligente leefwereld waarvan het brein in China en de Verenigde Staten wordt beheerd.

Het gereguleerde scenario – het beteugelen van dataconglomeraten is niet genoeg

Dataconglomeraten zijn het best gepositioneerd om schaalvoordelen en synergievoordelen van data naar zich toe te trekken, en daarmee leidend te worden in alle sectoren van de datagedreven economie. De Digital Markets Act van de Europese Commissie en in sterkere mate de aanbevelingen in 'De online wereld.nl', pleiten ervoor dat dataconglomeraten in de digitale infrastructuur streng worden gereguleerd. Genoemde maatregelen zijn onder andere: onthouding van excessieve dataverzameling, opsplitsing van activiteiten in het ecosysteem en openstelling van data aan mogelijke concurrenten. Implementatie van deze maatregelen lost zeker een gedeelte van de problematiek in de datagedreven toekomst op. Maar geen enkele oplossing die alleen de activiteiten van dataconglomeraten binnen Europa beteugelt is voldoende. Dataconglomeraten blijven buiten Europa namelijk data bundelen op schaal, waardoor ze in staat zijn betere algoritmen te trainen. Zelfs als ze deze algoritmen niet als propo-

in de Europese markt mogen zetten, kunnen ze de superieure algoritmen verkopen aan derden. Wanneer Europese bedrijven partnerschappen aangaan om deze superieure algoritmen toe te kunnen passen, wordt hun marge afgeroomd; en wanneer Europese bedrijven dat niet doen, verliezen ze de concurrentieslag met niet-Europese bedrijven die dat wél doen. We kunnen bijvoorbeeld verbieden dat Google en Apple zelfrijdende auto's verkopen in Europa, maar niet dat ze de data en algoritmen verkopen aan General Motors of Ford om zelfrijdende auto's te ontwikkelen en vervolgens te exporteren naar Europa. En Europese bedrijven verliezen ook hun positie om te kunnen exporteren naar niet-Europese markten, omdat ze daar wél moeten concurreren met dataconglomeraten.

Het helpt dus niet om enkel de dataconglomeraten aan banden te leggen. Nederlandse bedrijven moeten zélf ook de transitie doormaken en mee gaan doen in de datagedreven economie. Hiervoor is nodig dat Nederlandse bedrijven zelf toegang hebben tot schaalvoordelen en synergievoordelen van data. Daarmee kunnen Nederlandse organisaties blijven innoveren, concurreren en bijdragen aan maatschappelijke uitdagingen. Het gaat er niet zozeer om dat Nederland kan concurreren in het ontwikkelen van de beste algoritmen, maar wel dat Nederland kan concurreren mét de beste algoritmen, door deze toe te passen in producten en diensten. Dat kan alleen met toegang tot schaalvoordelen en synergievoordelen van data. Europa zal niet in staat zijn zelf een dataconglomeraat te bouwen, noch is dat wenselijk volgens Europese waarden van marktconcurrentie. Dat betekent dat organisaties toegang nodig zullen hebben tot elkaars data om op deze manier de beste toepassingen te ontwikkelen. Om onze concurrentiepositie te beschermen in de transitie naar de datagedreven economie, is het faciliteren van datadelen op schaal dus een absolute must. De volgende hoofdstukken onderzoeken wat nodig is om datadelen op schaal in Nederland en Europa te bewerkstelligen.





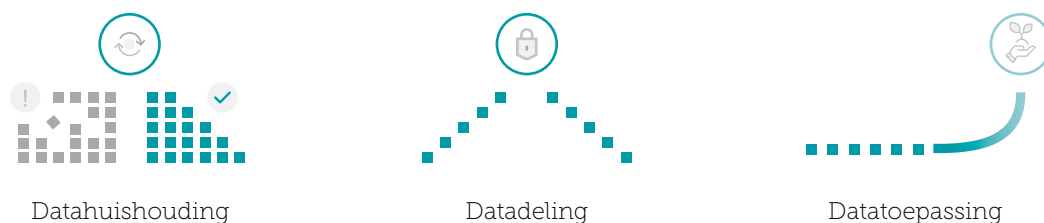
4. RANDVOORWAARDEN VOOR DATADELEN

Om de vruchten te plukken van een datagedreven wereld is het noodzakelijk dat data kan stromen naar plaatsen waar het waardevol kan worden toegepast. In geval van dataconglomeraten wordt dit georganiseerd binnen één bedrijf. Andere bedrijven zijn genoodzaakt om data met elkaar te delen. Het delen van data is echter geen sinecure. Uit onderzoek van de Europese Commissie in 2018 bleek dat 60% van de Europese bedrijven helemaal geen data deelt⁵. Bedrijven die dat wel doen, delen in de meeste gevallen slechts een beperkte hoeveelheid in een klein netwerk.

Wat is er nodig om data te laten stromen naar waardevolle toepassingen? Het begint met de kwaliteit van de datahuishouding van organisaties zelf. Als de datahuishouding niet goed is georganiseerd kunnen andere organisaties geen gebruik maken van de data. Datasets kunnen vervuild zijn,

60% van de Europese bedrijven deelt helemaal geen data

Figuur 7. Er zijn drie stappen nodig om data te laten stromen naar diens waardevolle toepassingen



⁵ Deze alinea volgt de logica van het 'Triple A' model dat is ontwikkeld door adviesbureau INNOPAY. Het Triple A model stelt dat Availability, Accessibility en Applicability nodig zijn voor het effectief stromen van data.

er kan bedrijfsspecifieke kennis nodig zijn om de data correct te interpreteren en data kan opgeslagen zijn in een niet machine-leesbaar format. De volgende uitdaging is het daadwerkelijk beschikbaar stellen van de data aan derden. Er moeten hiervoor verschillende technische afspraken worden gemaakt, onder andere over compatibiliteit, beveiliging, identificatie en authenticatie. Maar daarnaast zijn ook operationele en juridische afspraken nodig om risico's zoals onrechtmatig gebruik en privacyschending te minimalise-

De Europese datahuishouding kan worden beschreven als een Babylonische kakofonie

ren^k. Deze technische, operationele en juridische afspraken kunnen onderling worden gemaakt, of ze kunnen worden gefaciliteerd door een intermediair. Tot slot moet in de derde stap de data op een waardevolle manier worden toegepast door de ontvangende partij.

In dit rapport leggen wij de nadruk op de middelste stap: het beschikbaar stellen van data aan andere organisaties. Het verbeteren van de datahuishouding en het toepassen van data zijn namelijk voor veel organisaties al een hoge prioriteit – met de kanttekening dat ook hier nog veel te winnen is. Datadelen lijdt onder een coördinatieprobleem en een gebrek aan urgentie. Er is in Nederland nog te weinig aandacht en financiering voor vanuit bedrijven, terwijl het op systeemniveau een essentiële stap is richting de datagedreven economie.

Dit hoofdstuk zal de randvoorwaarden beschouwen voor datadelen in kleine kring. Hoofdstuk 5 onderzoekt hoe dezelfde randvoorwaarden kunnen worden gerealiseerd voor datadelen op grote schaal en cross-sectoraal – wat uiteindelijk

noodzakelijk is om schaalvoordelen en synergievoordelen van data te benutten. Wij zien twee randvoorwaarden voor datadelen die samen vanuit het belangrijkste zijn. Ten eerste heeft Europa een gemeenschappelijke data-taal nodig. Er bestaat op dit moment een sterke heterogeniteit van technische formats en data-taxonomieën, wat datadelen in de weg staat. Ten tweede moet vertrouwen in veilig datadelen gewaarborgd worden door het systeem en dus niet afhankelijk zijn van de wederpartij.

Een gemeenschappelijke data-taal

De Europese datahuishouding kan worden beschreven als een Babylonische kakofonie^l. Ieder bedrijf slaat data op in eigen technische formats – de syntaxis van data. De betekenis van die data verschilt ook weer voor ieder bedrijf, omdat ze verschillende 'data-taxonomieën' hanteren – de semantiek van data. Met taxonomie verwijzen we naar de interpretatie. Het gegeven dat een persoon 'risicoklasse A' is kan bijvoorbeeld voor de ene verzekeraar iets heel anders betekenen dan voor de andere. Door het verschil in technische formats en data-taxonomieën spreken bedrijven geen gemeenschappelijke taal om data in te delen.

Een gebrek aan een gemeenschappelijke data-taal is een enorm obstakel voor het delen van data. Een voorbeeld hiervan vinden we wederom bij zelfrijdende auto's. BMW en Mercedes-Benz realiseerden zich in juni 2019 allebei dat ze meer data nodig hadden om de strijd aan te gaan met partijen als Google en Apple. Ze besloten daarom met elkaar in zee te gaan en onder andere testdata te delen. Maar na een aantal maanden viel het partnerschap uiteen, omdat bij afwezigheid van één data format en data-taxonomie het te moeilijk en te duur bleek om de data met elkaar te delen. Een ander voorbeeld vinden we in de chemische industrie, waar nagenoeg geen sensordata van bedrijven wordt gebundeld. Ieder bedrijf gebruikt namelijk net andere meetmethoden en parameters om hun processen te dataficeren⁶. Dataconglomeraten hebben geen last van dit probleem, omdat ze intern één standaard hanteren. Dit is precies wat Jeff Bezos met zijn mandaat

^k INNOPAY ontwikkelde het 9-blokjesmodel waarin alle uitdagingen voor Accessibility in kaart worden gebracht.

^l Dit probleem bestaat op zekere hoogte ook in andere delen van de wereld.

uit 2002 wilde bereiken: dat verschillende takken van zijn bedrijf data konden delen en dus konden samenwerken. Dataconglomeraten dwingen niet enkel één taal af binnen hun bedrijf, maar ook bij al hun partners. Als een bedrijf bijvoorbeeld zijn producten wil aanbieden in de webshop van Amazon, moet het communiceren in de taal van Amazon. Als Microsoft-partners hun softwareproducten willen integreren in Microsoft suites, moet dat in de taal van Microsoft. Hierdoor is in de Verenigde Staten, en op een vergelijkbare manier ook in China, een veel hogere mate van homogeniteit in technische formats en data-taxonomieën ontstaan.

Om de datagedreven economie te ontsluiten zijn nieuwe standaarden nodig, die een grote variëteit aan datatypen beschrijven

Een verschil in format en taxonomie kan worden overbrugd met een vertaalsleutel: een specificatie hoe data van de ene organisatie omgezet kan worden naar het format en taxonomie van de andere organisatie. Het ontwikkelen van vertaalsleutels is echter tijdrovend en kostbaar. Bovendien dient de vertaalsleutel aangepast te worden bij iedere wijziging die een organisatie doorvoert in zijn datahuishouding. Dit is voor één-op-één datadeling nog een mogelijkheid, maar het is absoluut niet schaalbaar. Het is dus beter om toe te werken naar een gemeenschappelijke taal – een standaard.

In het verleden zijn op veel terreinen succesvolle standaarden ontwikkeld. Organisaties als de Internationale Organisatie voor Standaardisatie (ISO) en het Nederlands Normalisatie Instituut beheren duizenden standaarden omwille van compatibiliteit^m.

^m Naast standaarden omwille van compatibiliteit zijn er ook standaarden omwille van kwaliteit, zoals brandwerendheid, waterkwaliteit en voedselveiligheid. Bedrijven kunnen zich door geaccrediteerde inspectiebureaus laten certificeren op basis van deze kwaliteitsnormen. In 'De online wereld.nl' doen wij de aanbeveling om organisaties zich verplicht te laten certificeren voor digitale wet- en regelgeving, waaronder GDPR.

Zo is er een standaard voor de omtrek van bouten en papierformaten (zoals A4), zodat hiermee rekening kan worden gehouden bij het ontwerpen van respectievelijk moersleutels en printers. Ook in het digitale domein bestaan veel standaarden omwille van compatibiliteit. Zo werden in 1987 de technische kenmerken vastgelegd voor digitale mobiele telefonie met de GSM-standaard. De Nederlandse overheid heeft begin deze eeuw Standard Business Reporting geïntroduceerd (SBR). SBR is een afsprakenpakket om zakelijke informatie zoals jaarrekeningen, belastingaangiften en vastgoedtaxaties op een uniforme manier te delenⁿ. Een alliantie van Nederlandse banken lanceerde in 2005 iDEAL, een standaard voor internetbetalingen. Ook zijn er al standaarden geïntroduceerd specifiek als onderdeel van datadeelsystemen. In Nederland kennen we voorbeelden in de verzekeringswereld (Assurantie Data Netwerk), energiesector (Energie Data Services Nederland) en voor hypotheek (Hypotheek Data Netwerk)^o. Ook grote bedrijven proberen standaarden te introduceren door deze op te leggen aan partners in hun waardeketen, zoals Volkswagen doet met de Volkswagen Industrial Cloud^p.

Deze bestaande standaarden werken goed binnen hun huidige specifieke toepassing en domein, maar zijn te beperkt om ook datadelen in andere domeinen en voor andere toepassingen mogelijk te maken. De datagedreven wereld vereist het delen van grote hoeveelheden en diverse data, met enorme verscheidenheid aan datatypen. Bestaande standaarden hebben daarvoor een te klein bereik. Om de datagedreven economie te ontsluiten zijn dus nieuwe standaarden nodig, die een veel grotere variëteit aan datatypen beschrijven. Hoofdstuk 6 beschrijft de route die we het beste kunnen nemen om tot deze standaarden te komen.

Vertrouwen voor datadelen

De andere randvoorwaarde voor datadelen is dat de delende partij voldoende vertrouwen moet

ⁿ SBR bestaat naast een standaardtaal ook uit software en digitale infrastructuur voor rapportages: Digipoort voor rapportages aan overheid en Bancaire Infrastructuur Voorziening (BIV) voor rapportages aan private partijen.

^o Zie het rapport 'De overheid als partner bij datadelen' van Digicampus voor meer toelichting en voorbeelden.

^p Zie 'Orchestrating the Value in IoT Platform-Based Business Models' van het BCG Henderson Institute voor meer informatie.

hebben dat er niet verkeerd met de data wordt omgegaan. Op dit moment betekent dat meestal dat de delende partij vertrouwen moet hebben in de wederpartij. Wanneer een organisatie data deelt met een andere organisatie, stelt het zichzelf bloot aan een aantal belangrijke risico's. Zo kan de ontvangende organisatie de data gebruiken om een concurrerende propositie te ontwikkelen. Deze organisatie kan de data ook verkopen aan een concurrent van de delende organisatie, of er gewoon slordig mee omgaan waardoor klantgegevens op straat komen te liggen. Om data te delen moet een organisatie dus vertrouwen hebben in de partij waarmee de data wordt gedeeld. Daarmee is de betrouwbaarheid van datadelen een essentiële randvoorwaarde om datadelen mogelijk te maken.

Vertrouwen is noodzakelijk ongeacht de manier waarop data wordt gedeeld. De Nederlandse AI coalitie onderscheidt in haar rapport 'Verantwoord datadelen voor AI' drie modellen voor datadeling, met als belangrijkste verschil de vertrouwensbalans tussen partijen: de data brengen naar het algoritme, het algoritme brengen naar de data, of beide brengen naar een neutrale omgeving^q. Ook wanneer het algoritme naar de data gebracht wordt is vertrouwen nodig, omdat algoritmen vaak intellectueel eigendom bevatten en informatie weggeven over data in de trainingset. Het gekozen model heeft implicaties voor wie de data en het algoritme kunnen inzien, en wie dus in feite wie moet vertrouwen. Maar voor alle drie de modellen voor datadeling is vertrouwen vereist.

Er zijn vier belangrijke manieren om vertrouwen te creëren en daardoor datadelen mogelijk te maken⁷:

- **Informeel vertrouwen.** Een organisatie kan vertrouwen hebben in een zorgvuldige en goedaardige omgang met gedeelde data zonder een interventie te doen om risico's weg te nemen. Dit is bijvoorbeeld het geval bij een langdurig partnerschap of een uitstekende reputatie van de ontvangende organisatie.

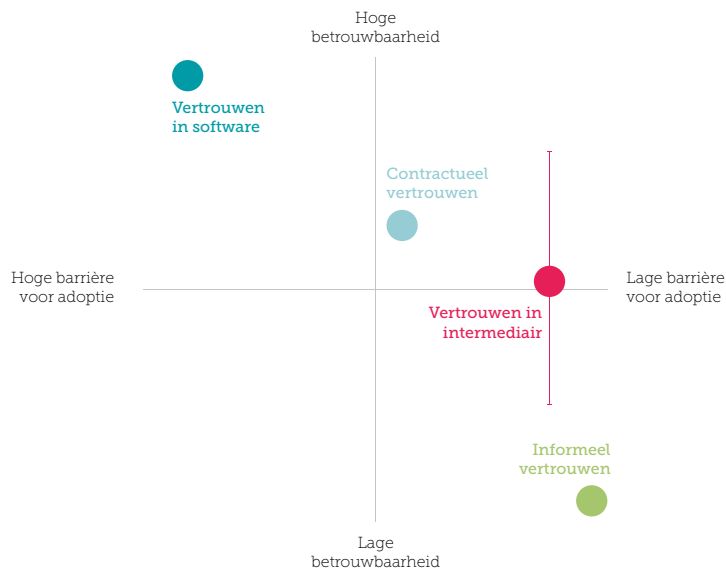
- **Contractueel vertrouwen.** Organisaties kunnen onderling afspraken maken over hoe er met de gedeelde data moet worden omgegaan, inclusief monitoring en een sanctieregime voor overtredingen. Deze vorm van vertrouwen is de basis voor afsprakenstelsels^r.
- **Vertrouwen in intermediair.** Partijen kunnen ervoor kiezen om vertrouwen in elkaar te vermijden en in plaats daarvan een derde partij te vertrouwen. Deze vertrouwensvorm speelt met name een rol wanneer ervoor wordt gekozen om algoritme en data te laten interacteren in een neutrale omgeving.
- **Vertrouwen in software.** Er is tegenwoordig veel software beschikbaar om risico's weg te nemen^s. De belangrijkste software op het gebied van vertrouwen is de blockchain⁸. Met traditionele beveiliging kan een organisatie een 'register' opstellen voor wie de data mag zien en wat er mee gedaan mag worden. Maar op het moment dat een andere partij de data vervolgens ontvangt, kan deze veranderingen in dat register doorvoeren zonder dat de oorspronkelijke eigenaar van de data het doorheeft. Met blockchain wordt het register decentraal opgeslagen, waardoor alle deelnemende partijen in het datadeelsysteem hetzelfde register hebben: iedereen weet wie toegang tot de data heeft en voor welke doeleinden. De oorspronkelijke eigenaar van de data behoudt dus controle ook nadat de data is gedeeld.

^q In feite is 'data beschikbaar stellen' dus een accuratere omschrijving dan 'data delen'.

^r Een contract levert alleen vertrouwen op wanneer een sterke onafhankelijke partij toeziet op de naleving ervan. Kevin Werbach spreekt daarom ook wel van 'Leviathan vertrouwen', naar het machtswezen dat in Hobbes' filosofie toeziet op de naleving van afspraken en daarmee een mogelijkheid tot samenleven creëert voor de mens.

^s Zie voor een uitgebreidere beschouwing van vertrouwen door software o.a. het rapport 'Verantwoord datadelen voor AI' van de Nederlandse AI Coalitie. Hier wordt de staat van verschillende nieuwe technieken geëvalueerd, waaronder federatief leren.

Figuur 8. Betrouwbaarheid versus implementeerbaarheid van vertrouwensvormen



Bron: DenkWerk-analyse

De ene manier om vertrouwen te creëren is niet beter dan de andere, maar ze zijn geschikt voor verschillende situaties (zie figuur 9)^t. Er bestaat op dit moment een afweging tussen de betrouwbaarheid van een vertrouwensvorm en de barrière voor adoptie van de vertrouwensvorm. Informeel vertrouwen is het makkelijkst te realiseren, maar is ook het minst solide en kent daarmee de grootste risico's.

Contractueel vertrouwen kent hogere barrières voor adoptie, want het is arbeidsintensief om over contracten te onderhandelen en deze te monitoren. Het is betrouwbaarder dan informeel vertrouwen, maar kent wel nog steeds het risico van contractbreuk. De betrouwbaarheid en barrière voor adoptie van intermediairs wisselt sterk per intermediair. Maar over het algemeen zijn intermediairs goed in het creëren van gebruiksvriendelijkheid (dit stimuleert namelijk adoptie van hun platform)⁹. De betrouwbaarheid van intermediairs

^t Vaak wordt gebruik gemaakt van een combinatie van vertrouwensvormen. Intermediairs verhogen vaak het vertrouwen van gebruikers via software en contracten.

Organisaties zullen enkel data delen wanneer ze het gevoel hebben dat hun datasoevereiniteit daarbij niet in het geding komt

in de online wereld is wisselvallig; zo beschadigen veel platforms in de online wereld op dit moment het vertrouwen dat gebruikers in hen hebben door privacygevoelige data te verzamelen en te hergebruiken voor indirecte verdienmodellen^u. Software kan zorgen voor een zeer hoge betrouwbaarheid, maar is voor veel organisaties moeilijk te

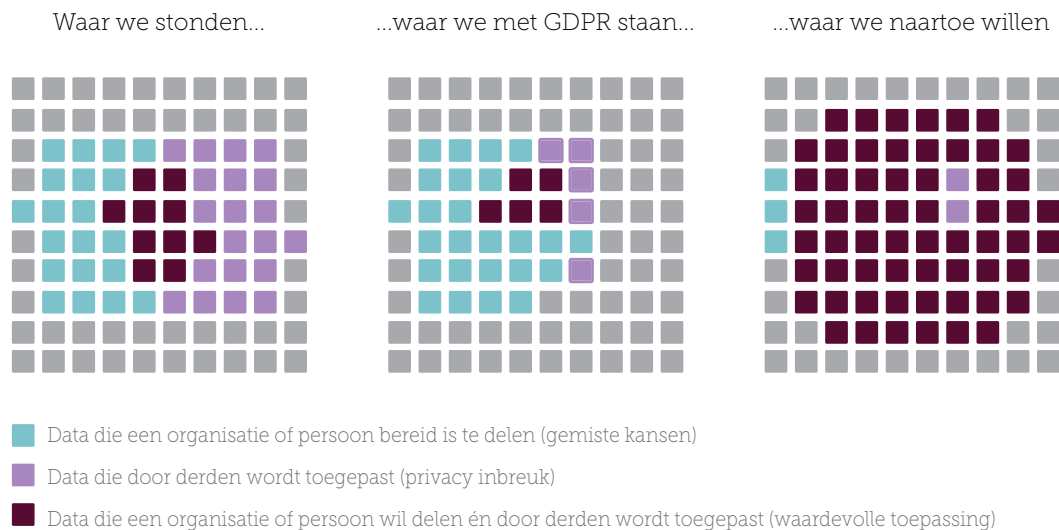
^u Zie hoofdstuk 6 van 'De online wereld.nl'.

implementeren, omdat het gepaard gaat met hoge kwaliteitseisen aan de datahuishouding. Omdat verschillende vertrouwensvormen geschikt zijn voor verschillende situaties, maken de meeste datadeelsystemen op dit moment gebruik van een combinatie van vertrouwensvormen^v.

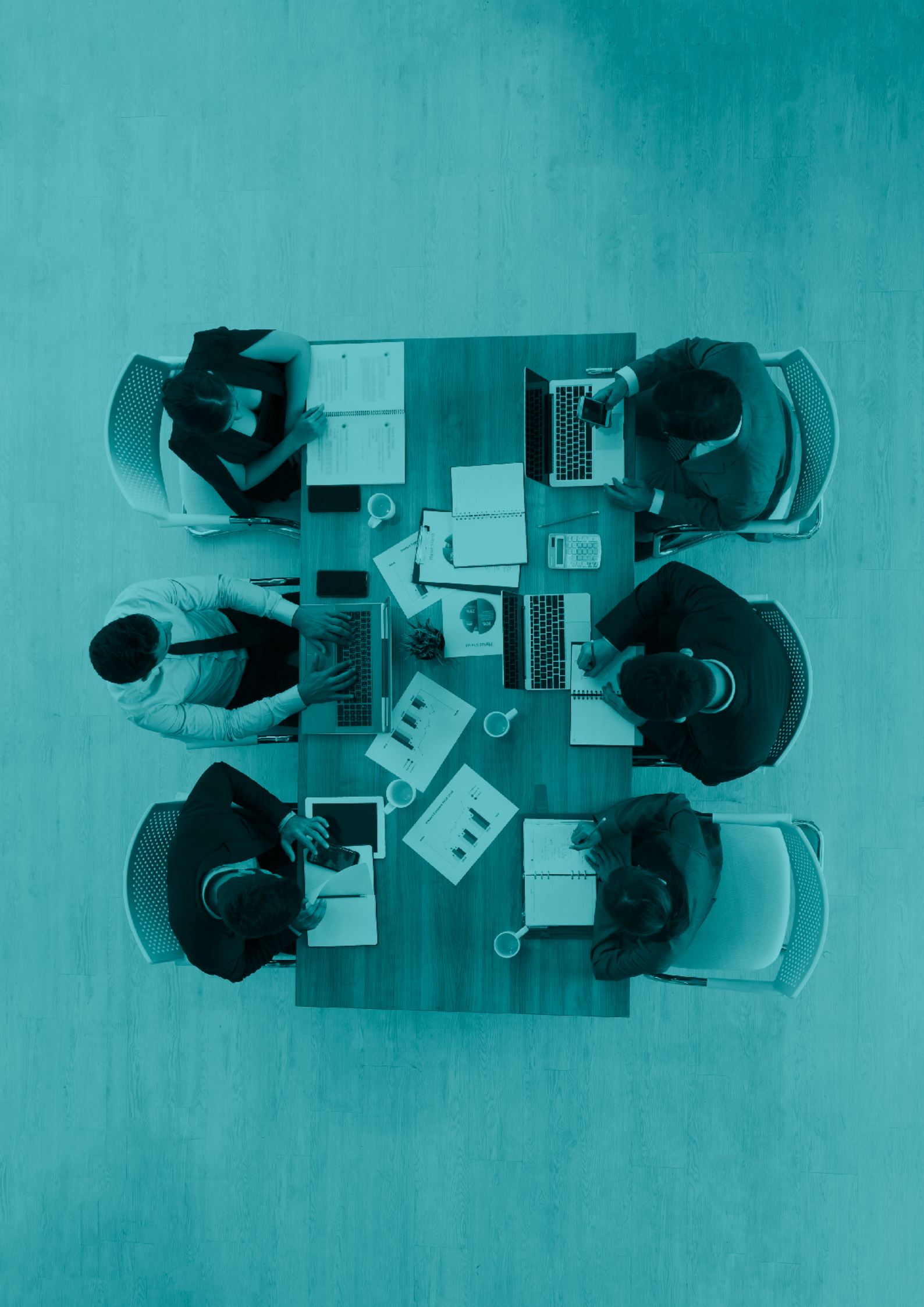
Wanneer een organisatie of een persoon te allen tijde controle houdt over zijn data – en zijn vertrouwen dus niet geschonden kan worden nadat hij data heeft gedeeld – spreken we van *datasoevereiniteit*. Datasoevereiniteit houdt in dat iedereen kan bepalen wie er bij zijn data kan en wat er met die data mag gebeuren. Organisaties zullen enkel data delen wanneer ze het gevoel hebben dat hun

datasoevereiniteit daarbij niet in het geding komt. Voor burgers is datasoevereiniteit op dit moment enkel een droom. Voor vrijwel alle internetdiensten is het noodzakelijk om data 'af te staan' aan platforms en andere bedrijven, zonder controle over deze data te behouden. Hoewel GDPR burgers veel rechten geeft over hun data, hebben zij in de praktijk vaak geen mogelijkheid om hun rechten uit te oefenen¹⁰. Hoe betrouwbaarder een datadeelsysteem is, hoe meer data er gedeeld kan worden terwijl de delende partij zijn datasoevereiniteit behoudt.

Figuur 9. GDPR heeft ervoor gezorgd dat er minder privacy-inbreuk plaatsvindt, maar niet dat data ook waardevol wordt toegepast. Met datasoevereiniteit kunnen beide worden gecombineerd.



^v Zie bijvoorbeeld het rapport 'De overheid als partner bij datadelen' van Digicampus.



5. DATADELEN OP SCHAAL

Hoofdstuk 4 beschreef de uitdaging van datadelen in kleine groepen. Maar het heeft de allerhoogste prioriteit dat Europa datadelen *op schaal* realiseert. Dat is namelijk noodzakelijk om de welvaart van Europa te beschermen en om meester te blijven over de algoritmen die vormgeven aan onze leefwereld. Gelukkig staan we nog aan het begin van de transitie en hebben we een kans om datadelen op schaal op tijd te realiseren.

Datadelen als meerzijdige markt

Datadeling op schaal houdt in dat een omvangrijke en diverse groep van personen en organisaties elkaar kan vinden om data te delen, ongeacht de locatie of sector waarin ze zich bevinden. Het is een veel-tot-veel interactie, of in andere woorden: een meerzijdige markt. Een entiteit die aanbieders en vragers van data bij elkaar brengt om data te delen – en daarmee gestalte geeft aan de meerzijdige markt – noemen we een datadeelsysteem. Een datadeelsysteem is dus een soort Amazon, maar in plaats van producten kan men er op zoek gaan naar de juiste data voor een specifiek doel – aangeboden door een grote veelzijdigheid aan organisaties die hun data beschikbaar stellen.

De meerzijdige markt voor datadelen is nog jong en dus onontgonnen terrein. Er zijn nog geen dominante datadeelsystemen, maar wel een groot aantal groeiende initiatieven. Deze initiatieven verschillen sterk van elkaar in bijvoorbeeld governancestructuur en voorspellen dus ieder een wezenlijk andere toekomst voor de datagedreven wereld. Op lokaal niveau is er bijvoorbeeld de Amsterdam Data Exchange (AMdEX). Ook bestaan er coöperaties met vaak een sectorale focus, zoals JoinData voor landbouw, MedMij in de zorg en iSHARE in de logistiek. In het private domein worden datamarktplaatsen gebouwd door tal van

kleine spelers, zoals Datum, Datawallet, Datacoup en Citizen.me voor persoonlijke data; The DX Network en Ocean voor bedrijfsdata; en IOTA Data Market en Streamr voor sensordata. Grote bedrijven hebben consortia opgericht met datadeelsystemen, met als voornaamste voorbeeld HERE Technologies van onder andere de Duitse autofabrikanten, Intel, Mitsubishi en NTT. Industriële grootmachten als Siemens en John Deere organiseren datadeling in hun eigen sector, respectievelijk in industriële machines en precisielandbouw.

Datadeling op schaal houdt in dat een omvangrijke en diverse groep van personen en organisaties elkaar kan vinden om data te delen, ongeacht de locatie of sector waarin ze zich bevinden

Ook de grote cloudbaanbieders hebben geïntegreerde datamarktplaatsen gebouwd, zoals AWS Data Exchange en Google Cloud Marketplace. Een andere grote aanbieder is Snowflake, het Amerikaanse data-warehousing bedrijf dat in 2014 zijn deuren opende en vandaag meer dan 100 miljard dollar waard is. Snowflake biedt naast diensten voor databeheer nu ook diensten voor datadeling en dataverkoop via Snowflake Data Marketplace en Snowflake Data Exchange.

Alle bovengenoemde initiatieven – van AMdEX tot Datawallet en HERE tot Snowflake – bouwen een datadeelsysteem met verschillende spelregels en governancestructuur en proberen dus een bepaalde richting te geven aan de meerzijdige markt voor datadelen. Uiteindelijk zullen er

te delen (omwille van synergievoordelen), met andere woorden: er bestaan sterke netwerkeffecten. Doordat er sterke netwerkeffecten bestaan in een datadeelsysteem, zullen er uiteindelijk slechts weinig voortbestaan. Dit betekent dat datadeelsystemen met een beperkte geografische dekking, zoals AMdEX, en datadeelsystemen met een beperkte sectorale dekking, zoals JoinData, weliswaar belangrijk zijn om mee te “starten en oefenen”, maar uiteindelijk niet in staat zijn datadeling op schaal te realiseren.

Doordat er sterke netwerkeffecten bestaan in een datadeelsysteem, zullen er uiteindelijk slechts weinig datadeelsystemen overblijven

weinig datadeelsystemen overblijven. Aanbieders en vragers van data zijn gebaat bij een groot aantal partijen om mee te delen (omwille van schaalvoordelen) en een diverse groep om mee

Datadelen verdient een decentraal model

Het belangrijkste verschil in bestuursstructuur tussen verschillende datadeelsystemen is of het gaat om een centraal model (platform) of decentraal model (standaarden en open platforms)^y. In het decentrale model is er niet één platform waarop gebruikers elkaar vinden, maar zijn er meerdere 'open platforms' die toegang verschaffen tot het datadeelsysteem. De netwerkeffecten zitten in dit model dus ook niet vast aan één platform, maar zijn de optelsom van alle deelnemende

hiervoor is dat deze grote dataconglomeraten veel complementaire proposities hebben, zoals cloud-diensten voor databeheer en datatoepassing (zoals AWS, Google Cloud en Microsoft Azure). Bovendien hebben deze dataconglomeraten een grote klantenbasis aan wie ze datadeeldiensten kunnen kruisverkopen. Hoewel datadelen dus een alternatief moet worden voor de dominantie van dataconglomeraten, bestaat er een risico dat deze dataconglomeraten óók het datadelen naar zich toe gaan trekken. Dit is ten eerste schadelijk omdat deze platforms data kunnen onttrekken aan het datadeelverkeer op hun platform^z. Daarnaast kunnen deze dataconglomeraten de spelregels op hun datadeelplatform zodanig inrichten dat datadelen niet gestimuleerd maar juist bemoeilijkt wordt. Dit creëert namelijk ruimte voor henzelf om unieke dataproposities te ontwikkelen (zie hoofdstuk 3). Het is bijvoorbeeld de vraag in hoeverre Google bereid is datadeling in de autosector maximaal te ondersteunen op haar datadeelplatform als dit leidt tot concurrentie bij het ontwikkelen van de zelfrijdende auto.

Organisaties kunnen in een decentraal model kiezen via welk open platform zij deelnemen aan het datadeelsysteem. Open platforms in een decentraal datadeelsysteem concurreren niet met elkaar op basis van netwerkeffecten, maar op basis van de kwaliteit van dienstverlening en de spelregels die door het platform worden ondersteund. Dit levert een groot maatschappelijk voordeel op, omdat platforms op deze wijze 'gedwongen' worden om de spelregels te ondersteunen die hun gebruikers wensen. De spelregels zijn van groot belang voor wie met elkaar data zullen delen, onder welke voorwaarden dat gebeurt en hoe de waarde daarvan wordt verdeeld.

Een ander belangrijk voordeel van het decentrale model is dat er geen platform lock-in^{aa} ontstaat. Net zoals veel micro-entrepreneurs in de platformeconomie nu 'vastzitten' aan een platform, en daardoor kwetsbaar zijn voor oneerlijke

z Waarschijnlijk zullen ze geen toegang hebben tot de gedeelde data zelf, maar wel tot de metadata: wie deelt met wie, op welk tijdstip, voor welke prijs, hoe groot is de dataset, et cetera. In het verleden is aangetoond dat dataconglomeraten ook metadata verzamelen bij gebruik van bijvoorbeeld clouddiensten en email.

aa Zie hoofdstuk 4 van 'De online wereld.nl' voor nadere toelichting.

Open platforms in een decentraal datadeelsysteem concurreren niet met elkaar op basis van netwerkeffecten, maar op basis van de kwaliteit van dienstverlening en de spelregels die door het platform worden ondersteund

platforms samen. Een decentraal datadeelsysteem maakt gebruik van een gestandaardiseerde data-taal en een afsprakenpakket, waardoor interoperabiliteit tussen de deelnemende platforms ontstaat. Hoewel er uiteindelijk weinig datadeelsystemen zullen bestaan in de meerzijdige markt, is het wel mogelijk dat er veel open platforms bestaan binnen een decentraal datadeelsysteem.

In een centraal georganiseerde meerzijdige markt, waar enkele dominante platforms alle spelregels uitmaken, hebben deelnemende organisaties weinig te kiezen met betrekking tot hoe datadeling plaatsvindt. Het grote probleem hiervan is dat de 'natuurlijke winnaars' in het centrale model – de platforms die de spelregels gaan bepalen – de grote dataconglomeraten zullen zijn. De reden

y Een voorbeeld van een decentraal georganiseerde meerzijdige markt is bellen: iedereen kan elkaar bellen, ongeacht telecomprovider. Een voorbeeld van een centraal georganiseerde meerzijdige markt is videobellen. Hier kan men alleen gebruikers van hetzelfde platforms bereiken (zoals op Skype, Hangouts, Facetime en Zoom).

Figuur 10. Modellen voor een meerzijdige markt



Model 1: Platform

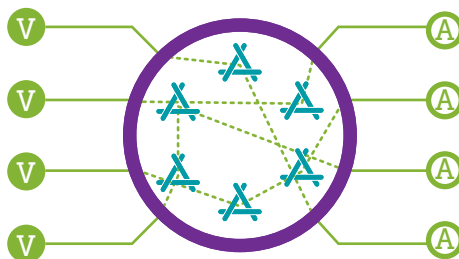
Vraag en aanbod in een meerzijdige markt vinden elkaar binnen een ommuurd platform



Model 2: Open standaarden

Vraag en aanbod voor data vinden elkaar via een gestandaardiseerde data-taal, waarbij een verscheidenheid aan open platforms diensten verleent om datadelen mogelijk te maken

-  Standaard
-  Datadeelplatform



praktijken, kunnen organisaties met hun datahuishouding straks vast komen te zitten in een datadeelplatform. In een decentraal datadeelstelsel is het altijd mogelijk om te switchen naar een concurrerend platform, waardoor de organisatie controle behoudt over zijn eigen data. In dit

model is er dus sprake van datasoevereiniteit. Het platform kan de gebruikers dan niet 'gijzelen' en oneerlijk behandelen.



6. IMPLEMENTATIE EN ADOPTIE VAN DATADELEN

Europa heeft op korte termijn een meerzijdige markt nodig voor datadelen. Organisaties uit verschillende sectoren en locaties kunnen elkaar hier vinden om grote hoeveelheden en diverse data te delen. Dit geeft Europese bedrijven de kans om te blijven innoveren, met name door artificiële intelligentie toe te passen in hun producten en diensten. Zonder onderling data te delen worden Europese bedrijven afhankelijk van dataconglomeraten en dus wordt hun marge afgeroomd.

Deze meerzijdige markt voor datadelen kan ontstaan wanneer aan twee randvoorwaarden wordt voldaan^{ab}: een gestandaardiseerde data-taal en schaalbaar vertrouwen voor datadelen. Hoofdstuk 5 liet zien dat er reeds veel opkomende datadeel-systemen zijn die pogen om deze meerzijdige markt voor datadelen te bouwen. Maar er is nog een lange weg te gaan. Te weinig bedrijven delen te weinig data. Nederland heeft dus een strategie nodig om de randvoorwaarden voor datadelen op schaal zo snel mogelijk te realiseren. Daarin ligt een grote rol voor zowel overheid als bedrijfsleven.

Nederland heeft een strategie nodig om de randvoorwaarden voor datadelen op schaal zo snel mogelijk te realiseren

^{ab} Daarnaast is het ook noodzakelijk dat bedrijven investeren in hun eigen datahuishouding en hun vermogen om data innovatief toe te passen. Een belangrijk onderdeel hiervan is upskilling van personeel op het vlak van digitale vaardigheden (zie Denkwerk rapport 'Arbeid in Transitie').

Het organiseren van vertrouwen voor datadelen

Eén van de randvoorwaarden voor succesvolle datadeelsystemen is effectief en schaalbaar vertrouwen creëren. Zoals beschreven in hoofdstuk 4, zijn er verschillende vertrouwensvormen die ieder geschikt zijn voor verschillende omstandigheden. Datadeelsystemen maken vaak gebruik van een mix aan vertrouwensvormen om datadelen te realiseren. Door deel te nemen aan deze datadeelsystemen en actief te experimenteren met en te investeren in de verschillende vertrouwensvormen dragen bedrijven bij aan het realiseren van datadelen op schaal. De overheid kan met beleid bijdragen aan de effectiviteit van ieder van de vertrouwensvormen.

- Stimuleren van vertrouwen door contracten.

De overheid kan een ondersteunende rol spelen bij de totstandkoming van afsprakenstelsels. Zo heeft het Ministerie van Economische Zaken een belangrijke rol gespeeld bij de totstandkoming van eHerkenning en de Data Sharing Coalition¹³. Er zijn in Nederland echter nog veel te weinig afsprakenstelsels en het niveau van samenwerking is nog onvoldoende.

- Reguleren van vertrouwen door intermediairs.

De overheid kan bijdragen aan de betrouwbaarheid van intermediairs door deze te reguleren. In de financiële sector bestaat er regelgeving en toezicht om betrouwbaarheid van financiële intermediairs (zoals banken) te waarborgen. De onlangs aangekondigde Digital Services Act stelt regelgeving op voor socialemediaplatforms. Zo zijn er ook spelregels nodig voor datadeelplatforms. In november 2020 bracht de Europese Commissie een voorstel uit voor de Data Governance Act^{ac}. In het derde hoofdstuk hiervan staan enkele waardevolle richtlijnen om data-intermediairs te reguleren, met

^{ac} Zie hoofdstuk 7 voor een uitgebreidere toelichting op de Data Governance Act.

name met betrekking tot hun neutraliteit. Wij moedigen de Nederlandse overheid aan om deze richtlijnen snel te implementeren en om aanvullende maatregelen te formuleren.

- **Investeren in vertrouwen door software.** Software als vertrouwensvorm is jong en er is veel ruimte voor verbetering, bijvoorbeeld op het gebied van blockchaintechnologie en algoritmen die federatief kunnen leren. De overheid kan dergelijk onderzoek co-financieren. Een belangrijke stap is ook de introductie van het digitale paspoort.

Het digitale paspoort

Wanneer een organisatie data deelt, moet het de ontvangende partij in meerdere opzichten vertrouwen. Zo moet er vertrouwen zijn dat deze partij zegt wie hij is (authenticatie) en dat hij enkel doet met de data wat is afgesproken (autorisatie). Autorisatie is een complex proces waar geen makkelijke oplossing voor bestaat. Er is dus een combinatie van vertrouwensvormen nodig om dit in te richten. Voor authenticatie bestaat mogelijk wél een 'simpele' oplossing: het digitale paspoort.

Ook in de fysieke wereld gebruiken wij een paspoort voor authenticatie. Zo heeft iemand zijn paspoort nodig om een rijbewijs aan te vragen, te vliegen of om een bankrekening te openen. Het paspoort zorgt ervoor dat publieke en private instituten erop kunnen vertrouwen dat een persoon is wie hij zegt te zijn, waardoor dienstverlening aan deze persoon soepel en snel kan plaatsvinden. In de online wereld bestaat een dergelijke universele authenticatiemethode niet. Oorspronkelijk had iedere website en iedere applicatie daarom zijn eigen wachtwoordstelsel. Dit bleek vaak niet handig. De grote dataconglomeraten zijn daarom in dit gat gesprongen. Tegenwoordig ondersteunen de meeste websites en applicaties de mogelijkheid om in te loggen met een Apple ID, Google-account of Facebook-account. Deze dataconglomeraten kunnen veel data verzamelen over

het hele internet doordat andere websites hun inlogstelsel gebruiken voor authenticatie.

Er gaan veel stemmen op dat het verzorgen van een universele authenticatiemethode in de online wereld geregeld zou moeten worden door de overheid, net zoals in de fysieke wereld. In feite heeft Nederland deze dienst al opgetuigd, in de vorm van DigiD. DigiD is nu alleen nog niet toe te passen buiten overheidswebsites. België loopt daarin voor. Daar is de digitale identificatie-app 'Itsme' ook toepasbaar voor burgers om zich te identificeren bij verzekeraars, banken en andere private bedrijven. Ook in Estland hebben alle burgers een e-identity, waarmee ze digitale handtekeningen kunnen zetten, kunnen stemmen en kunnen reizen door Europa¹⁴. Taiwan rolt een digitaal paspoort uit in juli 2021¹⁵.

Het Ministerie van Binnenlandse Zaken is bezig met de ontwikkeling van het digitale paspoort. Dankzij de Europese eIDAS-verordening zijn andere Europese lidstaten nu al verplicht om een eventueel Nederlands digitaal paspoort te herkennen en accepteren. Het digitale paspoort zal een belangrijke katalysator zijn voor datadelen, datasoevereiniteit en blockchaintoepassingen.

Co-creatie van standaarden

Naast vertrouwen, is standaardisatie van de data-taal nodig. Het internet der dingen floreert alleen met universele standaarden voor de syntaxis en semantiek van data, zodat data kan stromen naar zijn waardevolle toepassingen. Idealiter gebeurt dit op Europees niveau, maar als in Europa onvoldoende draagvlak bestaat voor een daadkrachtige standaardisatie met strenge proceseisen, adviseren wij om in Nederland te starten. In een latere fase kan Nederland dan homogenisering nastreven met andere Europese voorlopers en uiteindelijk met de Europese Unie. Er is een relatief kleine set aan datatypen die voorkomt in allerlei sectoren, zoals bijvoorbeeld tijds- en plaatsbepalingen. De overheid moet hiervoor in samenwerking met het desbetreffende normalisatie-instituut de standaard definiëren. De grotere uitdaging ligt in het definiëren van standaarden voor sectorspecifieke data. Hiervoor is een strategie nodig met twee elementen: een samenwerkingsmodel en een prioritering^{ad}.

Het internet der dingen floreert alleen met universele standaarden

Bij het definiëren van een samenwerkingsmodel is een balans nodig tussen de sectorspecifieke expertise die bedrijven meenemen en de bemiddelende en bespoedigende rol die neutrale partijen als de overheid kunnen spelen. In het verleden is gebleken dat het volgende samenwerkingsmodel goed werkt:

- **De overheid** stelt proceseisen zoals deadlines en inspraakprocedures op. De overheid is niet de aangewezen partij om het standaardisatieproces te leiden, maar wel om het af te dwingen^{ae}.

^{ad} De inzichten voor deze strategie zijn voor een belangrijk deel ontleend aan internationale projectervaring van BCG.

^{ae} De overheid heeft eerder bijvoorbeeld standaardisatie afgedwongen in het geval van eHerkenning (2010) en de OV-chipkaart (2004).

- **Een onafhankelijke partij** zoals een normalisatie-instituut of non-profit organisatie voert de dagelijkse regie over het proces. Het doel is om tot een goed functionerende standaard te komen met een brede consensus onder belanghebbenden in de sector. Hiertoe zit de onafhankelijke partij werkgroepen voor met bedrijven om standaarden uit te werken, welke vervolgens breed in de sector worden gevalideerd^{af}.
- **Bedrijven** uit de sector en technologie-bedrijven zoals sensorfabrikanten leveren input in de werkgroepen op basis van expertise en ervaring. Ook valideren zij de kwaliteit.

Een vergelijkbaar samenwerkingsmodel is gebruikt voor de ontwikkeling van GSM, eHerkenning en Standard Business Reporting⁴⁶. Bij het standaardisatieproces is het belangrijk om nauw samen te werken met bestaande sectorale datadeelsystemen om voort te bouwen op de opgedane ervaringen van eerdere initiatiefnemers.

De prioritering van verschillende domeinen voor standaardisatie hangt ten eerste af van de geografische spreiding van verbonden objecten in dat domein^{ag}. Een domein waar de verbonden objecten namelijk breed verspreid zijn over onze leefwereld, zoals het elektriciteitsnet, de autovloot, logistiek of toepassingen in huizen, kan bruikbare data genereren voor een grote groep toepassingen. Data die gegenereerd wordt op enkele specifieke locaties is vaak moeilijker te hergebruiken voor andere sectoren. Domeinen zoals de chemische industrie hebben dus een lagere prioriteit. Een tweede factor is het publieke belang. Datadeling ten behoeve van gezondheidszorg, veiligheid en klimaat heeft vanuit overheidsperspectief een hogere prioriteit dan datadelen ten behoeve van operationele efficiëntie van bedrijven.

^{af} De Data Sharing Coalition is sinds 2020 met dit proces gestart. Politieke sturing en prioritering kan hier voor versnelling zorgen.

^{ag} Er kan ook een prioritering gemaakt worden op basis van het technisch niveau. In feite is de syntaxis van de data een simplificering van o.a. communication protocols, messaging protocols, data format protocols, data model protocols, service method protocols en semantics protocols. Hierbij geldt dat het belangrijk is om de meer fundamentele syntactische standaarden eerst te definiëren, zodat de andere daarop voort kunnen bouwen.

Vergezicht voor datadelen: convergentie naar een publiek datadeelplatform?

In hoofdstuk 4 hebben we gesteld dat verschillende vertrouwensvormen nuttig zijn voor verschillende doeleinden. Vertrouwen via software kent over het algemeen de laagste risico's, maar kent hogere barrières voor adoptie^{ah}. Een datadeeloplossing met minimaal risico via software is dus geschikt voor zeer privacygevoelige informatie, zoals medische en financiële dossiers. Op de lange termijn zullen de hoge barrières een minder groot probleem worden. Naarmate bedrijven verder digitaliseren raakt de datahuishouding beter op orde, waardoor datadelen via software steeds vaker binnen handbereik komt. Data-management tools zoals Snowflake en Palantir kunnen bedrijven hierbij helpen. Wanneer de hoge barrières voor adoptie niet meer in de weg staan, zal datadelen via softwareoplossingen in veel situaties het meest aantrekkelijk zijn.

Wij voorzien op lange termijn dus een grote rol voor datadeelsystemen die vertrouwen organiseren via software. Maar ook voor softwareoplossingen is vertrouwen nodig, zoals bijvoorbeeld het vertrouwen dat de techniek goed functioneert en dat er geen achterdeurtjes bestaan waardoor de data kan weglekken. Hoewel het onzeker is hoe dit zich in de toekomst zal ontwikkelen, denken wij dat de overheid de aangewezen partij zal zijn om een datadeelsysteem gebaseerd op software te beheren. De essentie van deze rol is namelijk garant staan voor de betrouwbaarheid. Het 'bouwen' van de software kan overgelaten worden aan marktpartijen, net zoals het bieden van complementaire dienstverlening^{ai}. De overheid heeft vanwege zijn democratische aard echter de meeste legitimiteit om garant te staan voor betrouwbaarheid en veiligheid van de data, net zoals de centrale bank garant staat voor de waarde van geld.

^{ah} De adoptiebarrières voor vertrouwen in software zijn vaak hoog omdat het een hoge graad van digitalisatie vereist. Blockchain kan bijvoorbeeld gebruikt worden om een logboek bij te houden van wie een dataset inziet en welke modificaties er zijn gedaan, waardoor de integriteit van de dataset wordt bewaakt. Maar bij andere toepassingen van blockchain hebben we gezien dat adoptie vaak zeer moeizaam verloopt. Eén van deze toepassingen is voedseltracering, waar blockchain toegepast kan worden om de afkomst en veiligheid van voedsel aan consumenten te garanderen. De technologie werkt in theorie uitstekend om het maatschappelijke probleem op te lossen, en bedrijven als IBM en Foodchain ID hebben de waarde gedemonstreerd. Toch blijft adoptie door boerderijen en voedselverwerkers ver achter, hoofdzakelijk omdat kennis van blockchain beperkt is en de gepercipieerde complexiteit leidt tot wantrouwen (zie bijvoorbeeld 'Blockchain technology and traceability in the agrifood industry', Patelli, N., Mandrioli, M., Journal of Food Science, 30 september 2020).

^{ai} Door te zorgen dat het systeem interoperabel is kunnen derden hun diensten beschikbaar stellen aan gebruikers van het datadeelsysteem, zoals tools voor data management of data pricing.

Samen staan Nederlandse bedrijven sterk

Standaardisatie van de data-taal en het creëren van vertrouwen zijn enkel de randvoorwaarden voor datadelen. Uiteindelijk is het aan organisaties zelf om het ook echt te gaan doen. Eén onderdeel daarvan is dat bedrijven snelheid maken in de transitie naar een datagedreven bedrijfsvoering. Ook hierbij zijn bedrijven gebaat bij samenwerking. De transitie naar een internet der dingen vergt namelijk grote investeringen, die binnen een sector kunnen worden gedeeld^{aj}.

^{aj} Het vergt bijvoorbeeld investeringen in maatwerk sensortechnologie, industrie-specifieke software voor databeheer, integraties op maat met andere softwaresystemen zoals ERP, en cybersecurity.

De voornaamste stap is dat bedrijven actief deel gaan nemen aan datadeelsystemen. Het gaat hier dus niet om datadeelsystemen die gericht zijn op het delen van operationele data met ketenpartners, maar datadeelsystemen die gericht zijn op het delen van data die gegenereerd wordt voor algoritme-gestuurde toepassingen in het internet der dingen. Dát is de data waarover Nederlandse bedrijven in de nabije toekomst in grote hoeveelheden moeten kunnen beschikken. Bedrijven kunnen drie rollen spelen in datadeelsystemen^{ak}:

^{ak} Zie verschillende artikelen van het BCG Henderson Institute, waaronder 'Orchestrating the value in IoT platform based business models', voor een uitgebreidere beschouwing hoe ieder van de drie rollen centraal kan staan in de bedrijfsstrategie.

Figuur 11. De weg naar datadelen op schaal



- **Initiëren.** Bedrijven kunnen datadeel-systemen opzetten. Deze datadeelsystemen zijn een belangrijke stap richting de standaardisatie van de data-taal en het creëren van vertrouwen. Met name grote bedrijven hebben slagkracht om datadeelsystemen te initiëren voor hun sector. Ze kunnen hierbij een voorbeeld nemen aan verschillende succesverhalen, zoals van John Deere in precisielandbouw, Siemens in industriële machines en Airbus in de luchtvaart¹⁷.
- **Deelnemen.** Nederlandse bedrijven kunnen win-win situaties creëren door data beschikbaar te stellen aan andere, óf door nieuwe toepassingen te vinden voor de data van andere organisaties binnen hun eigen bedrijf. Beide dragen bij aan het stromen van data naar waardevolle toepassingen.
- **Faciliteren.** Voor Nederlandse technologie- en softwarebedrijven bestaat een kans om per sector innovatieve maatoplossingen te ontwikkelen voor de transitie naar

het internet der dingen. Microsoft werkt bijvoorbeeld al samen met Schindler en ThyssenKrupp om software voor intelligente liften te ontwikkelen, en met Honeywell, Schneider Electric en ABB voor industriële automatisering¹⁸. Decentrale datadeelsystemen zijn gebaseerd op interoperabiliteit. Dat betekent dat ze toegankelijk zijn voor andere bedrijven om er complementaire diensten aan te bieden, in tegenstelling tot de gesloten ecosystemen van dataconglomeraten. Deze dataconglomeraten investeren jaarlijks tientallen miljarden in complementaire diensten voor databeheer. Om de transitie naar de datagedreven economie in Nederland en in Europa te bespoedigen is het dus essentieel dat bedrijven ook proposities ontwikkelen voor databeheer, zodat Nederlandse en Europese datadeelsystemen kunnen concurreren met datadeelsystemen van dataconglomeraten op functionaliteiten en gebruiksgemak.



7. DATADELEN BOVENAAN DE AGENDA

Het vonkje van datadelen slaat over in Nederland en Europa...

Data kan op dit moment vaak zijn weg niet vinden naar zijn waardevolle toepassingen. Dit komt door een gebrek aan standaarden en doordat er veel onzekerheid en onduidelijkheid is over van datadelen. Het is dus nodig om via standaarden te zorgen dat datasystemen van verschillende organisaties interoperabel worden, zodat data vrij kan stromen. Daarnaast is nodig dat organisaties en burgers datasoeverein worden, door betekenisvolle controle te krijgen over wie toegang heeft tot hun data en voor welke doeleinden. Datasoevereiniteit moet zo het vertrouwen geven dat organisaties nodig hebben om data te delen. Er is publiek-private en intersectorale samenwerking nodig om standaarden en vertrouwen voor datadelen te organiseren.

In selecte kringen wordt reeds hard gewerkt aan het realiseren van standaarden voor en vertrouwen in datadelen

In selecte kringen wordt reeds hard gewerkt aan het realiseren van standaarden voor en vertrouwen in datadelen. Nederland loopt daarbij voor op veel andere landen. Naast een verscheidenheid aan sectorale initiatieven zijn er in Nederland ook cross-sectorale initiatieven zoals de Data Sharing Coalition en de Nederlandse AI Coalitie^{al}. De Data Sharing Coalition is een groeiende groep bedrijven, kennisinstellingen en IT leveranciers

^{al} Een cross-sectoraal datadeelsysteem wordt ook wel een 'generiek' datadeelsysteem genoemd.

die sinds januari 2020 samenwerkt aan het harmoniseren van standaarden en datadeelsystemen over sectoren heen. De Nederlandse AI Coalitie richt zich specifiek op het faciliteren van artificiële intelligentie, waar datadelen een essentiële voorwaarde voor is. Door deze initiatieven en het lanceren van afsprakenstelsels als iDEAL en SBR in het verleden heeft Nederland veel relevante ervaring opgebouwd. Tegelijkertijd zijn de Data Sharing Coalition en de Nederlandse AI Coalitie nog kleinschalig en bewegen ze traag. Hoewel Nederland voorop loopt met het realiseren van datadelen, is het cruciaal om te beseffen dat de 'race' niet gelopen wordt tegen andere landen die gaan datadelen – maar tegen dataconglomeraten als GAFAM die binnen hun eigen ecosysteem schaal- en synergievoordelen behalen.

De race om datadelen te realiseren loopt Nederland niet tegen andere landen maar vooral tegen dataconglomeraten die binnen hun eigen ecosysteem enorme hoeveelheden en diverse data verzamelen

In februari 2019 publiceerde het Ministerie van Economische Zaken ook een Nederlandse visie op datadelen voor bedrijven. Het feit dat Nederland een dergelijke visie publiceert als één van de eerste landen in Europa getuigt van de leidende rol die Nederland kan spelen om datadelen te realiseren. Tegelijkertijd ligt het ambitieniveau van de visie te laag, gegeven de uitdaging waar we voor staan. Er wordt gesteld dat 'datadeling meer dan nu gebruikelijk moet worden'. In dergelijke formuleringen herkennen wij onvoldoende urgentie.

Het standpunt dat de Nederlandse overheid vervolgens inneemt over zijn eigen rol getuigt ook van een afwachtende houding. Over de eigen rol op middellange termijn wordt gesteld dat 'als partijen data willen delen, dan kan de overheid dit overwegen te faciliteren in de vorm van financiële of organisatorische ondersteuning. In dat geval gelden wel voorwaarden die voor de overheid gebruikelijk zijn, zoals een duidelijk economisch of maatschappelijk belang'. De overheid kiest hier dus voor een reactieve rol, waar wij geloven dat de overheid proactief zou moeten bijdragen aan een versnelde vormgeving van de meerszijdige markt voor datadelen. Daarbij geldt ook dat in onze optiek altijd sprake is van een economisch of maatschappelijk belang. Wanneer bedrijven namelijk geen data kunnen delen door een gebrek aan een gemeenschappelijke data-taal of vertrouwen, is dat een vorm van marktfalen die economisch verkeer en innovatie in de weg staat.

In Europa worden datadelen en datasoevereiniteit inmiddels gezien als 'heilige graal' van onze digitale toekomst. In februari 2020 bracht de Europese Commissie de Europese Data Strategie uit, waarin datadelen centraal staat¹⁹. De Data Governance Act (DGA) van november 2020 is de eerste concrete invulling van deze strategie²⁰. De DGA kent vier pilaren: spelregels voor hergebruik van data uit de publieke sector, eisen aan data-intermediairs, introductie van data-altruïsme^{am} en de oprichting van een 'European Data Innovation Board'. Er bestaat naast dit wetsvoorstel ook een verscheidenheid aan concrete initiatieven en programma's om datadelen in Europa te faciliteren. Zo werd in juni 2020 het Frans-Duitse initiatief GAIA-X gelanceerd: een decentrale cloud met het doel om burgers en organisaties controle over data te geven en daarmee ook datadelen te vergemakkelijken. Datadelen speelt ook een grote rol in Next Generation Internet: een programma waarbij de Europese Commissie €250 miljoen investeert in onderzoek naar de toekomst van het internet²¹. En de International Data Spaces Association brengt Europese bedrijven en onderzoekers bij elkaar om cross-sectoraal datadelen te realiseren op Europese schaal²².

^{am} Data-altruïsme verwijst naar het vrijwillig beschikbaar maken van data door personen of organisaties voor het algemeen belang.

Hoewel de Europese Data Strategie en de verschillende Europese initiatieven gegrond zijn in een goed begrip van de problematiek en aansluiten bij onze aanbevelingen, vrezen wij dat daadkrachtige implementatie op Europese schaal een uitdaging zal zijn. De Data Governance Act moet nog worden goedgekeurd door het Europese Parlement en door alle lidstaten. Het risico hierbij is dat de wetgeving ten prooi valt aan fragmentatie doordat ieder land zijn eigen belangen beschermt. Eerder is bijvoorbeeld een investering van €500 miljoen als onderdeel van het Europese Digital Innovation Hubs programma verspreid over 450 lokale hubs, waardoor het effect grotendeels is verwaterd^{an}. Ook bij GAIA-X rommelt het. Zo kijken veel landen de kat uit de boom en klaagden Nederlandse deelnemers dat Frankrijk agressief probeert om eigen standaarden door te drukken²³. Juist voor artificiële intelligentie is het essentieel om op Europese schaal samen te werken, zodat bedrijven toegang hebben tot dezelfde schaalvoordelen als Amerikaanse en Chinese concurrenten. Maar multilaterale samenwerking blijkt een complex en tijdrovend proces^{ao}. Nederland kan het zich dus niet veroorloven om op Europa te wachten, maar moet zélf en in samenwerking met andere Europese voorlopers toewerken naar datadelen op schaal.

Nederland kan het zich niet veroorloven om op Europa te wachten, maar moet zélf en in samenwerking met andere Europese voorlopers toewerken naar datadelen op schaal

^{an} Ter vergelijking: China heeft meer dan twee keer zoveel inwoners en minder dan half zoveel AI hubs (zie 'Europe is missing out on the AI revolution – but it isn't too late to catch up' van François Candelon en Rodolphe Charme Di Carlo, 4-12-2020).

^{ao} Een alternatieve route naar multilaterale samenwerking is via bedrijven. Consortia als HERE technologies laten zien dat grote bedrijven die de urgentie van datadelen inzien onderling in staat zijn om in hoog tempo te bouwen aan datadeelsystemen. Zo kunnen bedrijven de route naar datadelen op Europese schaal drastisch bespoedigen.

...nu moeten we het vuur leven inblazen

We weten welke kant we op moeten. Nu moeten we de stappen zetten om er te komen – en snel. Datadelen verdient een plek bovenaan de agenda van overheid en bedrijfsleven. Wij roepen de Nederlandse overheid op om direct na de verkiezingen aan de slag te gaan met het afdwingen van standaardisatie van de data-taal, het stimuleren en ondersteunen van afsprakenstelsels voor datadelen en het bespoedigen van de implementatie van de Europese Data Strategie (gelijktijdig met de aanbevelingen uit het rapport 'De online wereld.nl'). Het nog op te richten Ministerie van Digitale Zaken en

Dataverkeer moet hierin de leiding nemen. Nederlandse bedrijven roepen wij op om data centraal te stellen in de bedrijfsvoering en om actief deel te nemen aan datadeelsystemen. Concurrentiële afwegingen en coördinatieproblemen zouden dit niet in de weg moeten staan; immers geldt voor de transitie naar de datagedreven economie dat Nederlandse bedrijven samen zullen winnen of samen zullen verliezen.

Voor de transitie naar de datagedreven economie geldt dat Nederlandse bedrijven samen zullen winnen of samen zullen verliezen

Door datadelen op schaal te realiseren in Europa krijgen Europese bedrijven toegang tot schaal- en synergievoordelen van data. Hierdoor zijn zij in staat om innovatieve AI-toepassingen te verwerken in hun diensten en producten. Zo kunnen ze concurrerend blijven in hun eigen markt, zonder afhankelijk te worden van Amerikaanse of Chi-

nese dataconglomeraten. Datadelen zorgt er daarmee voor dat Nederlandse bedrijven aan het roer staan bij de vormgeving van de algoritmen die onze intelligente leefwereld zullen besturen. Bovendien beschermt deze ontwikkeling onze economische concurrentiepositie en dus onze welvaart in de 21e eeuw.

François Cadelon vat de Europese transitie naar de datagedreven economie samen met een uitspraak van de Italiaanse dichter Lampedusa: "Als we willen dat de dingen hetzelfde blijven, zullen we alles moeten veranderen."

DANKWOORD

Als DenkWerk hebben we bij de totstandkoming van dit rapport veel waardevolle bijdragen en ondersteuning ontvangen. Wij willen in het bijzonder de volgende personen bedanken voor hun waardevolle inzichten:

Axel Arnbak	De Brauw Blackstone Westbroek
Cecile Schut	Autoriteit Persoonsgegevens
Constantijn van Oranje	TechLeap.nl
Douwe Lycklama	INNOPAY
François Candelon	BCG Henderson Institute
Haye Hazenberg	Ministerie van Binnenlandse Zaken
Jeroen van den Hoven	Technische Universiteit Delft
José van Dijck	Universiteit Utrecht
Lokke Moerel	Morrison & Foerster
Marian Grubben	Europese Commissie
Marleen Stikker	Waag
Massimo Russo	BCG Henderson Institute
Nico van Eijk	CTIVD
Remy Gieling	MT/Sprout
Valerie Frissen	SIDN Fonds
Viktor Mayer-Schönberger	Universiteit van Oxford
Yvonne van der Brugge	Logius

Daarnaast willen we Robert van der Veeke bedanken voor zijn uitvoerende rol in het onderzoeks- en schrijfproces. Anna Wijers bedanken wij voor haar hulp bij het onderzoek en voor haar kritische blik. Dieuwertje ten Feld en Mieke Stoop bedanken wij voor hun ondersteuning. Douwe Lycklama zijn we dankbaar voor het delen van zijn expertise tijdens ons onderzoek en het aandachtig meelesen en verscherpen van de tekst. Tot slot willen wij BCG Design Studios Amsterdam bedanken voor het verzorgen van de visuele vormgeving van het rapport.

DenkWerk blijft, uiteraard, verantwoordelijk voor de conclusies en opvattingen in dit rapport.

We stimuleren het te allen tijde om de inzichten uit dit rapport te gebruiken. We zouden het op prijs stellen als u het rapport daarbij als bron vermeldt:

DenkWerk – De datagedreven toekomst.nl. Hoe we vormgeven aan onze toekomst in de datagedreven wereld (2021)

APPENDIX: SAMENWERKING EN CONCURRENTIE IN DE DATAGEDREVEN ECONOMIE

De belangrijkste conclusie in dit rapport is dat het een absolute noodzaak is om op korte termijn de condities te creëren voor datadelen op schaal. Tot slot neemt deze appendix de vraag in beschouwing hoe bedrijven data zullen gaan delen wanneer de condities hiervoor zijn gecreëerd en welke impact dit heeft op de competitieve dynamiek in markten.

De paringsdans van datadelen

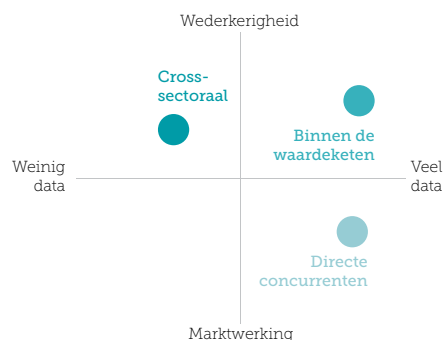
De komende jaren zullen de condities voor datadeling – hopelijk in hoog tempo – verbeteren. Naarmate de condities verbeteren, zal er een paringsdans ontstaan voor datadelen: wie gaat met wie en onder welke voorwaarden? Wij verwachten drie archetypische relaties. Daarbij verschilt steeds hoeveel data zal worden gedeeld en of dat naar verwachting zal gebeuren volgens een model van marktwerking of wederkerigheid. Figuur 12 classificeert datadeling binnen de verschillende archetypische relaties.

- **Binnen de waardeketen.** Partners in een waardeketen zullen naar verwachting veel data delen en bovendien vaak volgens een wederkerig model. Partners in een waardeketen hebben namelijk een gevestigde relatie en een patroon van samenwerking. Bovendien weten ze van elkaar welke data ze hebben en hoe ze deze data kunnen toepassen. Ook is het vaak in het belang van bedrijven om partners in de waardeketen te helpen. Door data te delen met leveranciers kunnen deze hun processen verbeteren en dus goedkopere en/of hoogwaardigere diensten en producten leveren. Voor afnemers geldt dat een verbetering van hun processen mogelijk leidt tot meer succes in de markt en dus indirect tot meer vraag naar jouw producten. Datadelen binnen de waardeketen kent dus het grootste coöperatieve en het kleinste concurrentiële domein.

- **Directe concurrenten.** Delen met directe concurrenten zal het minst vaak gebeuren, omdat het concurrentiële domein hier erg groot is. Er bestaat een significant risico dat een concurrent de gedeelde data gebruikt om zijn product te verbeteren of processen te optimaliseren. Toch kan het zijn dat concurrenten onderling datadelen om zo hun positie in de markt te verstevigen versus andere concurrenten. Dit is wat Mercedes-Benz en BMW poogden met zelfrijdende auto's, en wat boerencoöperaties doen via JoinData of John Deere. In dit geval gaat het vaker om datadeling volgens een model van wederkerigheid. Het is zelden in het strategisch belang van een bedrijf om data aan concurrenten te verkopen wanneer deze de data kunnen gebruiken voor het behalen van concurrentievoordeel. Wederkerige deling helpt beide partijen om leidend te zijn in de volgende innovatieslag.

- **Cross-sectoraal.** Bij cross-sectoraal datadelen is het concurrentiële domein beperkt. Er is vaak sprake van asymmetrisch datadelen, waarbij één partij een toepassing heeft voor data van een ander, maar niet andersom. Het voornaamste samenwerkingsmodel hiervoor is datahandel.

Figuur 12. Paringsdans van datadelen



Naast de archetypische relatie bepaalt ook de relatieve omvang van partijen of en hoe ze datadelen. Een bekend verschijnsel in de bedrijfskunde is een natuurlijke tendens naar 'soort zoekt soort'. Dit kan worden verklaard vanuit speltheorie.

Als grote bedrijven alleen data delen met elkaar, drukken ze kleinere partijen uit de markt

Grote bedrijven zijn gewilde datadeelpartners, simpelweg omdat ze veel data hebben. Zij krijgen dus de eerste keuze met welke partijen ze willen samenwerken. Dit zullen andere grote partijen zijn. Zo ontstaat de situatie waarin de grootste partijen uit verschillende sectoren met elkaar gaan samenwerken. Bedrijven die net een maatje kleiner zijn en dus de tweede keuze hebben zullen elkaar opzoeken, enzovoort. Dit is bijvoorbeeld te zien in de wereld van chemicaliën, waar producenten vaak samenwerken met groothandels van vergelijkbare omvang. Het nadeel hiervan is dat kleinere bedrijven minder toegang krijgen tot de nodige data om te concurreren en innoveren in een datagedreven economie. Als grote bedrijven alleen data delen met elkaar, drukken ze dus in feite kleinere partijen uit de markt. Dit is niet noodzakelijk: als grote partijen data delen met elkaar, kost het ze weinig extra's om de data ook te delen met kleinere partijen (mits een functionerende meerzijdige markt voor datadelen aanwezig is). De motivatie om alleen data te delen binnen een selecte alliantie van grote bedrijven zou alleen zijn om elkaar in het zadel te houden: ik deel geen data met jouw concurrenten, als jij geen data deelt met mijn concurrenten. Deze strategie is te omschrijven als datacollusie.

Datacollusie en het pleidooi voor een databelasting

Omdat datacollusie een natuurlijke maar suboptimale uitkomst is in een datagedreven economie, pleit Viktor Mayer-Schönberger, professor aan Oxford, voor een databelasting. Bedrijven worden hierdoor gedwongen een deel van hun data in het publieke domein beschikbaar te stellen (er wordt belasting betaald door data vrij te geven). Er zijn drie belangrijke argumenten voor een databelasting.

- Ten eerste vermindert dit het probleem van datacollusie.
- Ten tweede is het een oplossing voor het probleem dat datagedreven bedrijven steeds minder belasting betalen. Naarmate meer economisch verkeer plaatsvindt in de vorm van datastromen, zoals nu al het geval is bij B2C internetbedrijven als Google en Facebook, werken onze belastingmechanismen niet goed. Dit uit zich in een gemiddelde belastingdruk van 9,5% voor digitale bedrijven vergeleken met een gemiddelde van 23,2% voor traditionele bedrijven²⁴. Door het invoeren van een databelasting, worden bedrijven gedwongen om privaat bezit in dienst te stellen van het publiek belang.
- Tot slot kan een databelasting ook helpen bij het versnellen van de transformatie naar een datagedreven economie. Het beschikbaar stellen van data in het publieke domein creëert namelijk een competitieve druk om de data te verzilveren. Waar bedrijven voorheen op hun gemak op zoek konden gaan naar innovatieve toepassingen voor hun data, moeten ze nu op hun hoede zijn voor uitdagers die de data gebruiken voor disruptieve toepassingen. Een vergelijkbare logica kennen wij ook voor intellectueel eigendom. Patenten hebben slechts een beperkte houdbaarheidsdatum zodat na verloop van tijd competitieve druk ontstaat om door te innoveren.

De discussie rondom een databelasting is nog in een vroeg stadium. Het roept enkele praktische vragen op, zoals hoeveel data moet worden afgedragen. Wij stellen voor om dit een afgeleide te maken van hoeveel data het bedrijf zelf toepast en deelt. Een complexer vraagstuk is wélke data geïnd moet worden. Het ene datapunt is niet gelijk aan het andere, zoals bij geld wel het geval is. Bovendien heeft een datapunt voor verschillende bedrijven een andere waarde, waardoor er voor derden een commercieel belang ontstaat bij de keuze van de belastingdienst voor wélke data ze innen.

Een andere relevante discussie is of Europa zichzelf met een databelasting niet in de voet schiet. Als Europese bedrijven namelijk verplicht data moeten delen in het publieke domein, komt deze data onvermijdelijk ook bij Chinese en Amerikaanse bedrijven terecht. Deze zouden hiermee een voordeel kunnen behalen, omdat Europese bedrijven alleen tot Europese data toegang hebben terwijl zij tot Europese én hun eigen data toegang hebben. Wij houden rekening met een toekomst waar Europa een databelasting moet afdwingen bij zijn handelspartners. Europa neemt dan het initiatief om een groot digitaal handelsblok te bouwen.

Wij denken dat een databelasting op den duur een noodzakelijke maatregel zal zijn. Maar de vormgeving van de maatregel kent vele complexiteiten, zoals bepalen welke data wordt geïnd en hoe we omgaan met niet-Europese handelspartners. Hier zijn nu nog geen goede antwoorden op, maar het behoeft wel onze aandacht. Voor nu is het echter urgenter om datadelen op schaal mogelijk te maken door de juiste condities te scheppen. Hiervoor dragen overheid en bedrijfsleven een gedeelde verantwoordelijkheid. Als we deze condities niet tijdig creëren, trekken dataconglomeraten uit de Verenigde Staten en China de datagedreven economie naar zich toe. Dat is onomkeerbaar. De precieze vormgeving van de Europese datagedreven economie kan ook in een later stadium worden bijgestuurd, bijvoorbeeld middels een databelasting.

BEGRIPPENLIJST

Algoritme – Een beslisregel waarmee een systeem automatisch beslissingen kan nemen. Vroeger werden deze beslisregels expliciet door de programmeur gecodeerd, maar tegenwoordig leren steeds meer algoritmen zichzelf de beste beslisregel aan gegeven een doelstelling en inputdata. Dit fenomeen heet machine learning.

Blockchain – Een technologie om informatie decentraal op te slaan waardoor alle deelnemende partijen de echtheid van informatie kunnen controleren.

Databelasting – Een maatregel die bedrijven verplicht om een deel van hun data beschikbaar te stellen in het publieke domein, met name om datacollusie te voorkomen en om te zorgen dat er competitieve druk blijft bestaan op het toepassen van data voor innovatie.

Data common / lake – Een economisch samenwerkingsprincipe voor het delen van data waarbij deelnemende partijen data storten in een gezamenlijk 'meer' waar iedereen vervolgens vrij uit kan vissen. Deze vorm van datadelen is gebaseerd op wederkerigheid en niet op datahandel.

Datacollusie – Een strategie voor grote bedrijven om enkel data te delen in een selecte alliantie om elkaar voordelen te gunnen versus andere concurrenten.

Datadeelsysteem – Een set aan afspraken en mechanismen waarmee data gedeeld kan worden. Een datadeelsysteem kan centraal zijn volgens een centraal model (platform) of decentraal model (open ecosysteem met standaarden). Datadeelsystemen verschillen onderling wat betreft de economische samenwerkingsprincipes, technische samenwerkingsprincipes en complementaire diensten.

Datasoevereiniteit – Het principe dat een persoon of organisatie te allen tijde controle heeft over zijn data in zoverre hij daar juridisch recht op heeft.

Federatief leren – Een manier van machine learning waarbij algoritmen simultaan getraind worden op verschillende datasets die hierdoor niet gebundeld hoeven te worden. Gevoelige data hoeft hierdoor zijn omgeving niet te verlaten waardoor privacy niet in het geding komt.

Interoperabiliteit – De mogelijkheid om verschillende systemen of apparaten met elkaar te kunnen laten communiceren. Een voorbeeld hiervan zijn telefoons. Zo kan een mobiele telefoon bij KPN communiceren (bellen) met een mobiele telefoon van Vodafone, maar ook met vaste telefoonlijnen.

Meerzijdige markt – Een markt waarin niet alleen een grote groep afnemers maar ook een groep aanbieders actief is. Een meerzijdige markt wordt vormgegeven volgens een centraal model (platform) of decentraal model (open platforms met standaarden).

Metadata – Data die de kenmerken van een dataset beschrijft, zoals de omvang en het moment waarop de set gecreëerd of gedeeld is.

Poortwachter – Een bedrijf dat vanwege zijn dominante positie in een markt de toegang tot aangrenzende markten kan beheren, bijvoorbeeld door exclusieve toegang tot de consument of door een monopolie op essentiële data. Een voorbeeld is de App Store van Apple, die als poortwachter geldt voor mobiele apps.

Taxonomie – Een overzicht van de semantische interpretaties van verschillende datapunten in een data-taal.

BRONVERMELDING

- 1** Newzoo via Statista, number of smartphone users worldwide 2016 – 2021, september 2019
- 2** Gartner, public cloud services market size 2009-2022, november 2020
- 3** 'The Dawn of Digital Medicine', The Economist, 5 december 2020
- 4** Zie artikel 'Bankieren dankzij Google', Financieel Dagblad, 07-12-2020
- 5** Zie 'Eurobarometer survey on data protection', Europese Commissie, juni 2019
- 6** Gebaseerd op projectervaring van BCG
- 7** Een gemodificeerde variant van de vier verschillende vertrouwensarchitecturen in 'Blockchain and the New Architecture of Trust' van Kevin Warbach, november 2018
- 8** Zie bijvoorbeeld 'Blockchain revolution' van Alex Tapscott en Don Tapscott, 2014, voor nadere toelichting
- 9** Zie voor verdere toelichting op het snel opschalen van platforms 'Alles Transactie' van Shikko Nijland, Chiel Liezenberg en Douwe Lycklama
- 10** 'Data? It's all about access', Mariane ter Veen, Innopay, 20-11-2019
- 11** Zie voor meer uitleg van het Solid project onder andere de blog 'One small step for the web...' van Sir Tim Berners-Lee van 22 oktober 2018
- 12** 'What would it take for Tim Berners-Lee's new Web replacement to succeed?', Josh Bernoff, 2 oktober 2018
- 13** Zie de websites van eHerkenning en de Data Sharing Coalition voor meer toelichting (<https://www.eherkenning.nl/> & <https://datasharingcoalition.eu/nl/>)
- 14** Zie de website van e-Estonia: <https://e-estonia.com/solutions/e-identity/id-card/>
- 15** 'Taiwan to roll out electronic ID card in July 2021', Taiwan News, 05-10-2020
- 16** Voor het governance model van SBR, zie https://nl.wikipedia.org/wiki/Standard_Business_Reporting; Voor het governance model van GSM, zie [https://nl.wikipedia.org/wiki/Gsm_\(communicatie\)#Geschiedenis](https://nl.wikipedia.org/wiki/Gsm_(communicatie)#Geschiedenis)
- 17** Zie 'Orchestrating the Value in IoT Platform-Based Business Models' van het BCG Henderson Institute
- 18** Zie noot 17
- 19** Zie <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy>
- 20** Zie <https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-data-governance-data-governance-act>
- 21** Zie de website van Next Generation Internet voor meer informatie (<https://www.ngi.eu/>)
- 22** Zie de website van de International Data Spaces Association (<https://www.internationaldataspaces.org/>)
- 23** Zie 'Hoe cloudproject GAIA-X spartelt in het Europese moeras', Financieel Dagblad, 23-11-2020
- 24** Getallen afkomstig uit 'Questions and Answers on a Fair and Efficient Tax System in the EU for the Digital Single Market' van de Europese Commissie

Denkend aan Holland zie ik breede rivieren traag door oneindig laagland gaan, rijen ondenkbaar ijle populieren als hoge pluimen aan den einder staan; en in de geweldige ruimte verzonken de boerderijen verspreid door het land, boomgroepen, dorpen, geknotte torens, kerken en olmen in een grootsch verband. de lucht hangt er laag en de zon wordt er langzaam in grijze veelkleurige dampen gesmoord, en in alle gewesten wordt de stem van het water met zijn eeuwige rampen gevreesd en gehoord.

HENDRIK MARSMAN, 1936



DENKWERK

