

Verkenning van verplichtingen inzake datadeling in de technologiesector

Onderzoek in opdracht van het ministerie van
Economische Zaken en Klimaat

Rotterdam, 30 november 2020



Verkenning van verplichtingen inzake datadeling in de technologiesector

Onderzoek in opdracht van het ministerie van Economische Zaken en
Klimaat

Nicolai van Gorp (e-Economics)
Paul de Bijl (Radicand Economics)
Inge Graef (Tilburg University)
Gabor Molnar (e-Economics)
Roel Peeters (Ecorys)
David Regeczi (Ecorys)

Rotterdam, 30 november 2020

Inhoudsopgave

Management Samenvatting	i
1 Inleiding	1
2 De rol van data in de digitale economie	6
2.1 De rol van data in digitale bedrijfsmodellen	6
2.2 De rol van data in digitale concurrentie	7
2.3 Data en innovatie	10
3 Waarde van verschillende soorten data en drijfveren voor uitwisseling	13
3.1 Verschillende datatypologieën	13
3.1.1 Een op oorsprong gebaseerde typologie	13
3.1.2 Persoonsdata en niet-persoonsgebonden data	16
3.1.3 Op inhoud gebaseerde typologie	17
3.2 De waarde van data en drijfveren om (niet) uit te wisselen	20
3.2.1 Sociale data jagen netwerkeffecten aan	20
3.2.2 Gedragsdata jagen leereffecten aan	21
3.2.3 Zelf-gegenereerde inhoud en operationele data jagen lock-in-effecten aan	23
3.2.4 Contextuele data verschaffen betekenis	24
3.2.5 Profieldata jagen netwerkeffecten aan en vormen een bron van inkomsten	24
3.2.6 De kosten van dataproductie en de impact van verplichte datadeling	25
3.3 Samenvattende infographic	28
4 Wetgeving inzake datadeling	31
4.1 Intellectueel eigendomsrecht en bedrijfsgeheimen	31
4.2 EU-wetgeving inzake bescherming van data	36
4.3 EU-mededingingsrecht	40
5 Beoordeling van mogelijke verplichtingen tot datadeling in vijf casussen	47
5.1 Inleiding	47
5.2 Casus: zoekmachines	48
5.2.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling	48
5.2.2 Huidige wettelijke bepalingen voor verplichte datadeling	51
5.2.3 Belemmeringen/voorwaarden voor verplichte datadeling	52
5.3 Casus: sociale netwerken	53
5.3.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling	53
5.3.2 Huidige wettelijke bepalingen voor verplichte datadeling	56
5.3.3 Belemmeringen/voorwaarden voor verplichte datadeling	57
5.4 Casus: Banken	57
5.4.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling	57
5.4.2 Huidige wettelijke bepalingen voor datadeling	60
5.4.3 Belemmeringen/voorwaarden voor verplichte datadeling	61
5.5 Casus: Administratieve bedrijfssoftware	61
5.5.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling	62
5.5.2 Huidige wettelijke bepalingen voor verplichte datadeling	64
5.5.3 Belemmeringen/voorwaarden voor verplichte datadeling	64

Inhoudsopgave

5.6	Casus: Verticale uitsluiting	65
5.6.1	Factoren voor verplichtingen in de Nederlandse visie op datadeling	65
5.6.2	Huidige wettelijke bepalingen voor verplichte datadeling	67
5.6.3	Belemmeringen/voorwaarden voor verplichte datadeling	68
5.7	Samenvatting	69
6	Conclusie en beleidsimplicaties	70
6.1	Belangrijkste kansen geboden door datadeling	70
6.2	Beleidsopties	73
6.3	Organisatorische en technische aspecten	80
7	Bibliografie	84
8	Bijlagen: casestudies	88

Management Samenvatting

Doelstelling en onderzoeksvraag

Dit rapport gaat in op de vraag: **Wat is de rol van data in de technologiesector en wat zijn de kansen en risico's op verplichte datadeling?** Het antwoord verschaft inzicht in kosten en baten van verschillende vormen van verplichte datadeling met en tussen technologiebedrijven.

De technologiesector behelst de ontwikkeling en levering van technologische producten en diensten. Deze definitie is zowel **te ruim** als **te beperkt** voor deze studie. Te ruim omdat we geen leveranciers van computeronderdelen analyseren die niet zijn betrokken bij data-analyse. Te beperkt omdat wij grote bedrijven analyseren die in hoge mate afhankelijk zijn van data in een context van sterke netwerkeffecten; zoals Amazon, dat over het algemeen wordt gerekend tot de detailhandel. Het meest van belang zijn bedrijven die interacties binnen en tussen waardeketens organiseren met behulp van (data)technologie. Daartoe behoren uiteraard bedrijven zoals Amazon, Apple, Facebook, Google en Microsoft, maar ook bedrijven zoals John Deere, Nike en Honeywell, die hun traditionele producten hebben omgezet in instrumenten voor het verzamelen van data die platforms voeden ten behoeve van data-analyse en innovatie. In feite gaat het om elk bedrijf dat datatechnologie gebruikt voor orkestratie en innovatie.

De rol van data voor technologiebedrijven

Datatechnologie beheert en interpreteert data. Dit wordt effectiever naarmate er meer data en meer soorten data zijn: de kwaliteit van diensten die datatechnologie gebruiken gaat omhoog en dit trekt meer gebruikers aan, waardoor meer data kunnen worden verzameld. Dit circulaire effect kan een "leereffect" worden genoemd.

Door leereffecten te combineren met netwerkeffecten (meer gebruikers trekken nog meer gebruikers aan) en breedtevoordelen (meer gebruikers trekken meer diensten aan, die weer meer gebruikers aantrekken) ontstaat een krachtige groeimotor. Een kleine voorsprong op het gebied van gebruikers en data kan er dan toe leiden dat een markt kantelt naar een monopolie. Zonder deze circulaire effecten is de kans op een dergelijke kanteling kleiner.

Ondanks de risico's dat een markt kantelt, hoeven digitale markten geen afgeschermd monopolie te worden, omdat ze zijn blootgesteld aan "moligopolistische concurrentie". Dit is een dynamische vorm van marktoverschrijdende concurrentie tussen enkele grote bedrijven en veel kleine innovators. Dergelijke markten kunnen aanvechtbaar en dynamisch zijn. Het risico bestaat echter dat moligopolisten concurrentieverstorende strategieën gebruiken die hen beschermen tegen uitdagers, of waarmee ze hun gevestigde positie overhevelen naar andere markten.

Bedrijven die datatechnologieën gebruiken ontwikkelen vaak een sterke honger naar data. Concurrentie kan dit versterken. Bij gericht adverteren, bijvoorbeeld, streven alle aanbieders van advertentieruimte naar meer data om beter te kunnen concurreren. Evenzo zorgt de dreiging van marktoverschrijdende concurrentie ervoor dat moligopolisten meer data verzamelen om betere te kunnen innoveren.

Technologiebedrijven kunnen hun concurrentiepositie versterken door digitale platforms te bouwen die verticale uitwisseling van data faciliteren met downstream ontwikkelaars, die hun ecosystemen verrijken. De mogelijkheid bestaat echter dat zij minder bereid zijn tot datadeling wanneer een downstream speler een concurrent dreigt te worden. Het kan dus zijn dat technologiebedrijven niet bereid zijn tot (horizontale) datadeling.

De rol van data in andere sectoren

Traditionele waardeketens vertonen vaak marktfricties, zoals informatie-asymmetrie, marktmacht en agency-problemen. Dankzij de (data)technologie die wordt gebruikt om interacties binnen de waardeketen te orkestreren, kunnen belanghebbenden dienstverlening en productieprocessen optimaliseren, marktmacht omzeilen of meer controle nemen over de waardeketen. Voorbeelden daarvan zijn te vinden in de logistiek, de zorg, de bouw, het onderhoud van gebouwen, en in de land- en tuinbouw. Dergelijke voorbeelden tonen aan dat datadeling binnen (en ook tussen) waardeketens essentieel is om het potentieel van data te ontsluiten. Deze data moeten worden beheerd en geïnterpreteerd, wat een taak zou kunnen zijn voor technologiebedrijven. De kans bestaat echter dat belanghebbenden niet bereid zijn om data uit te wisselen met anderen, uit angst de controle over hun bedrijf te verliezen. Er kan dus sprake zijn van een hold-up probleem met verschillende belanghebbenden, in welk geval contractuele zekerheid vereist is. In bovengenoemde bedrijfstakken wordt dit probleem aangepakt door te zorgen voor vertrouwde omgevingen voor datadeling, met regels die transparantie en non-discriminatie waarborgen. Met deze governance-systemen kunnen partijen profiteren van datadeling zonder dat dit negatieve gevolgen heeft. Ze bieden scheppers van data (individuele betrokkenen of bedrijven waarnaar de data kunnen worden herleid) de mogelijkheid om te bepalen wie toegang krijgt tot de data.

Om hun bedrijf te beschermen, zijn partijen die hun concurrentiepositie te danken hebben aan exclusieve toegang tot unieke data (bijvoorbeeld banken en elektriciteitsbedrijven) wellicht minder bereid tot verticale of horizontale datadeling. Het weigeren van datadeling kan concurrentie en innovatie in de hele waardeketen belemmeren.

Verschillende soorten data

Om de rol van data te begrijpen, onderscheiden we soorten data op basis van oorsprong, anonimiteit en inhoud. De **oorsprong** geeft aan hoe data zijn verworven, en informeert of een datadeelpllicht de prikkels om ze te verzamelen of te produceren kan ondermijnen. **Anonimiteit** heeft betrekking op de mate waarin data te herleiden zijn tot hun scheppers (betrokkenen of bedrijven). Dit bepaalt in hoeverre data kunnen worden uitgewisseld zonder de privacy van individuele personen te schaden of strategische risico's voor bedrijven te creëren. **Inhoud** verwijst naar het doel van data en welke rol ze spelen in de bedrijfsmodellen van technologiebedrijven. Binnen deze laatste typologie kunnen de volgende soorten data worden onderscheiden:

- **Sociale data** bevatten informatie om interacties tussen mensen te faciliteren.
- **Gedragsdata** worden gegenereerd door interacties met een interface; ze geven inzicht in voorkeuren van gebruikers en mogelijkheden om de kwaliteit van de dienstverlening te verbeteren.
- **Zelf gegenereerde inhoud** (*user generated content*) verwijst naar digitale bestanden die gebruikers (bewust) creëren met behulp van applicaties of clouddiensten. Het equivalent in een business-to-business (B2B)-context zijn **operationele data**.
- **Contextuele data** beschrijven de omstandigheden van een gebeurtenis, een persoon, een bedrijf of een product.
- Een **profiel** beschrijft wat belangrijk of interessant is aan een persoon in een bepaalde context.
- **Trainingsdata** worden gebruikt voor het trainen van algoritmen. Deze categorie kan alle bovenstaande soorten van data bevatten. Wellicht is een (arbeidsintensieve) bewerking nodig voordat ze kunnen worden gebruikt voor trainingsdoeleinden (denk bijvoorbeeld aan het labelen van afbeeldingen).

Voor elke soort inhoud kan worden aangegeven wat de waarde ervan is voor een bedrijfsmodel, hoe ze ontstaan (verstrekt/waargenomen, verkregen/afgeleid, geproduceerd), en in hoeverre ze aan waarde inboeten na anonimisering. Onderstaande tabel biedt een overzicht.

	Inhouds-categorie	Verstrekt / waargenomen	Afgeleid/ geproduceerd	Waarde voor het bedrijfsmodel
veel ↑ waardeverlies na anonimiseren ↓ weinig	Sociaal	ID, contact, sociale verbanden, communicatie	kracht van sociale verbanden	netwerkeffecten
	Profiel	gezondheid, financieel, demografisch	voorkeuren	netwerkeffecten reclame-inkomsten
	Gedrag	transacties, zoeken, surfen, klikken, bekijken, luisteren		leereffecten
	zelf-gegenereerd	afbeeldingen, clips, berichten, operationele data		lock-in leereffecten
	Contextueel	metadata, marktdata en vorige categorieën	kaarten, weer	helpen om andere data te interpreteren
	Trainingsdata	sociaal / profielen / gedrag / zelf-gegenereerd / contextueel	aanvullende bewerking voor gebruik	leereffecten
Opmerking: de meeste soorten data hebben betrekking op een schepper van data (een persoon of een bedrijf).				

Overheidscriteria voor verplichte datadeling toegepast op vijf casussen

De Nederlandse visie op data formuleert criteria voor verplichte datadeling. Deze studie onderzoekt deze criteria in vijf casussen: **1) zoekmachines**, onderzoekt horizontaal delen van data over zoeken en klikgedrag; **2) sociale netwerken**, onderzoekt verticale datadeling door social-media-platformen; **3) betaaldiensten**, met focus op het effect van PSD2 op de concurrentiepositie van banken en de wenselijkheid van verplichte datadeling door toegangzoekers; **4) administratieve bedrijfssoftware**, onderzoekt de mate waarin dataportabiliteit vendor lock-in kan voorkomen; en **5) verticale uitsluiting**, over situaties waarin een platform downstream-ontwikkelaars uitsluit. Onderstaande tabel geeft een overzicht van hoe in elke casus aan de overheidscriteria wordt voldaan. Daaronder presenteren wij algemene conclusies uit deze casussen.

Criteria	Zoeken	Sociale media	Bancair	B2B software	Verticale uitsluiting
Soort data	gedragsdata	Sociale data	profieldata	operationele data	Alle soorten
Richting	horizontaal	horizontaal	sector-overschrijdend	horizontaal	verticaal
Publiek belang	Kantelen van een markt	Kantelen van een markt	Gelijk speelveld	gebruikers-mobiliteit	misbruik
Repliceerbaar	niet meer	in hoge mate	niet door banken	niet relevant	hangt ervan af
Vrijwillig uitgewisseld	H: nee V: niet relevant	H: soms V: ja	Nee	H: nee V: ja	H: niet relevant V: nee
Effect op investeringen	in data: 0 andere: +/- ^a	in data: 0 andere: 0	in data: 0 andere: + ^b	in data: 0 andere: 0/- ^a	in data: 0 andere: + ^c
Negatieve neveneffecten	risico op reverse engineering.	risico's voor privacy	risico's voor privacy	standaarden kunnen maatwerk beperken	mogelijke objectieve redenen voor discriminatie

^a Afhankelijk van hoe sterk de negatieve neveneffecten (zoals vermeld in de tabel) zijn.

^b Als banken toegang hebben tot profieldata, kunnen ook zij (en niet alleen technologiebedrijven) de correlaties tussen bankdata en profieldata onderzoeken. Dit betekent dat meerdere entiteiten kunnen innoveren met dezelfde data.

^c Non-discriminatoire toegang tot data verhoogt de beschikbaarheid van de datapool voor een groter aantal innovators.

Impact van datadeling

Impact van datadeling op innovatie

Verticale datadeling is gericht op operationele efficiëntie en de ontwikkeling van nieuwe diensten. Veel platforms bieden een vertrouwde omgeving voor datadeling en ondersteunen verticale datadeling om het de diversiteit van diensten binnen hun ecosystemen te vergroten. Scheppers van data en/of verwerkingsverantwoordelijken zouden anders niet snel bereid zijn om data uit te wisselen. **Een beleidsvoorstel is om de ontwikkeling van vertrouwde omgevingen voor datadeling te stimuleren.**

Door verticale datadeling kunnen partijen van buiten de bedrijfstak deelnemen aan de waardeketen met data uit andere sectoren. Op die manier kunnen sector-overschrijdende data-analyses worden gefaciliteerd, met als gevolg meer mogelijkheden om innovaties te ontdekken. Wanneer data echter slechts in één richting stromen, kan dat ten koste gaan van een gelijk speelveld. **Een beleidsvoorstel is om te zorgen voor wederkerigheid van datastromen: "u krijgt toegang tot (al) mijn data als ik toegang krijg tot (al) uw data".** Een dergelijke wederkerigheid trekt het innovatiepotentieel voor alle deelnemers aan sectoroverschrijdende datadeling gelijk en maximaliseert het. Bovendien genereert het aan alle kanten netwerkeffecten, omdat het innovatiepotentieel van alle deelnemers groeit met elke extra deelnemer. Dit betekent dat naarmate het aantal deelnemers groeit, het voor andere spelers (uit verschillende sectoren) steeds aantrekkelijker wordt om ook mee te doen aan de datadeelovereenkomst. Merk op dat deze voordelen er niet zouden zijn in het geval van financiële compensatiemechanismen: het innovatiepotentieel van een bedrijf groeit niet als het wordt betaald voor gegevenstoegang. Financiële compensatie leidt bovendien niet tot netwerkeffecten.

Horizontale datadeling draagt bij aan een gelijkwaardiger toegang tot trainingsdata. Als gevolg daarvan zullen concurrenten waarschijnlijk intensiever concurreren door te investeren in intelligentie, in plaats van zoveel mogelijk data te verzamelen in een poging de markt te laten kantelen. Datadeling kan echter ook schadelijk zijn voor innovatie. Uit de zoekcasus blijkt dat uitwisseling van gedragsdata reverse engineering teweeg kan brengen, wat innovatieprikkel negatief zou kunnen beïnvloeden. Een verplichting om gedragsdata uit te wisselen kan in strijd zijn met intellectuele eigendomsrechten, die is gericht op de bescherming van die prikkels.

Impact van datadeling op concurrentie (en innovatie)

Het is mogelijk dat een bedrijf niet zonder meer bereid is tot uitwisseling van data die onmisbaar, uniek en exclusief zijn, en wanneer die exclusiviteit bescherming biedt tegen concurrentiedruk. Deze druk moet hier vanuit een hoek van 360 graden worden gezien: binnen en tussen markten en ook waar het verticale onderhandelingsposities betreft.

Exclusiviteit kan voortkomen uit het verleden (bijvoorbeeld banken vóór PSD2), of het gevolg zijn van een strategie om een concurrentiepositie te versterken (bijvoorbeeld een browser bundelen met een besturingssysteem). **Een beleidsvoorstel is om verticale datadeling verplicht te stellen wanneer afgeschermd concurrentieposities verticale datastromen belemmeren,** om datagedreven innovaties in de waardeketen te ontsluiten en toegang tot de bedrijfstak te bevorderen door middel van dergelijke innovaties.

Exclusiviteit kan ook het gevolg zijn van netwerk- en leereffecten, mogelijk in combinatie met (concurrentiebeperkende) strategieën die dataposities versterken. Het gaat daarbij met name om sociale data, profieldata, gedragsdata en (mogelijk) operationele data. Als er geen sprake is van concurrentieverstorend gedrag, **kan een verplichting om deze gegevens horizontaal te delen het risico van een kanteling van de markt verkleinen:** het stimuleert de netwerk- en leereffecten

van rivalen. In de aanwezigheid van concurrentieverstorende strategieën is een verplichting om gegevens te delen mogelijk niet effectief, noch noodzakelijk. Het dwingen van een dominant zoekbedrijf bijvoorbeeld om zoekgegevens te delen, helpt misschien niet om rivalen te laten groeien als het dominante bedrijf zijn browser en zoekmachine aan zijn besturingssysteem koppelt. Tegelijkertijd is een verplichting om gegevens te delen misschien niet nodig als deze diensten eenmaal zijn losgekoppeld, wanneer koppelverkoop de gegevens in de eerste plaats exclusief heeft gemaakt. In het laatste geval moet **concurrentieverstorend gedrag rechtstreeks worden aangepakt**, voordat een horizontale verplichting tot gezamenlijk gebruik van gegevens wordt overwogen. De reden is dat een remedie die het probleem direct aanpakt, gericht en dus effectiever zal zijn dan een indirecte oplossing.

Bij het opleggen van verplichtingen tot horizontale datadeling moet rekening worden gehouden met wederkerigheid in relatie tot de risico's op het kantelen van een markt.

Wederkerigheid betreft hier de onderlinge uitwisseling van specifieke en vergelijkbare soorten data tussen twee horizontale concurrenten. Deze vorm van wederkerigheid voorkomt dat het op een na grootste bedrijf de nieuwe partij met een machtspositie wordt. (Wederom, een financiële compensatie zal dit effect niet hebben). We wijzen erop dat deze vorm van reciprociteit beperkter is dan bovengenoemde vorm van wederkerigheid met betrekking tot sector-overschrijdende datadeling.

Impact op gebruikersmobiliteit

Horizontale overdracht van sociale data en door gebruikers zelf gegenereerde content (inclusief operationele data) beperkt lock-in en vergemakkelijkt overstappen en multi-homing (gelijktijdig gebruik van vergelijkbare diensten). Dit draagt bij aan meer horizontale concurrentie en minder risico op het kantelen van een markt. Bij administratieve bedrijfssoftware wordt vendor lock-in echter ook veroorzaakt door de operationele integratie van software, die moeilijk te ontwarren is. Voor socialemediadiensten leidt migratie van sociale data niet direct tot een verlaging van de overstapdrempels, omdat er geen migratie van connecties mee is gemoeid.

Om overstappen en multi-homing te vergemakkelijken, **is het een beleidsvoorstel om het datamobiliteitsbeleid aan te vullen met beleid dat is gericht op een lichte vorm van interoperabiliteit op dienstniveau**. Het kan bijvoorbeeld helpen om een functionaliteit te hebben waarmee gebruikers bestaande contacten kunnen uitnodigen om zich aan te sluiten bij een nieuw netwerk. Een functionaliteit om berichten tegelijkertijd op meerdere platformen te plaatsen zou ook multi-homing kunnen ondersteunen, zonder de strengere standaardisatie die nodig zou zijn om via elke willekeurige interface berichten op meerdere platformen in te kunnen zien en erop te reageren.

Impact van datadeling op privacy en beveiliging

Sociale data en gebruikersprofielen verliezen veel waarde na anonimisering. Uitwisseling van dergelijke data brengt privacyrisico's met zich mee en vereist een juridische basis, zoals een wettelijke verplichting of toestemming van de gebruiker. Een dergelijke legitimiteit sluit echter niet alle privacyrisico's uit. **Datadeling op grotere schaal vereist dus passende maatregelen om privacy en beveiliging te waarborgen**. Dit geldt in mindere mate voor gedrags- en operationele data, die hun waarde behouden na anonimisering.

Beleidsopties

Hieronder bespreken we deze beleidsvoorstellen in samenhang met en gestructureerd volgens twee beleidsdoelstellingen. De **eerste doelstelling** is om te profiteren van kansen die datadeling biedt, door te zorgen voor de noodzakelijke randvoorwaarden, eventueel in combinatie met verplichtingen. De **tweede doelstelling** is gericht op de aanpak van concurrentieproblemen.

Beleidsdoelstelling I: data mobiliseren en innovatie stimuleren

Op basis van bovenstaande discussie onderscheiden we drie kansen voor datadeling:

- I. waardeketens optimaliseren en ecosystemen verrijken;
- II. (onvoorziene) innovaties in verschillende sectoren ontsluiten; en
- III. klanten mobiliseren.

Deze kansen kunnen elkaar aanvullen. Terwijl de PSD2-richtlijn bijvoorbeeld ten doel heeft om kans I te realiseren, staat deze toe dat de grootste technologiebedrijven (Big Techs) de banksector betreden met data uit andere sectoren, wat bijdraagt aan kans II. Dit kan leiden tot een ongelijk speelveld als gevolg van asymmetrische dataposities: de Big Techs hebben bijvoorbeeld een enorme hoeveelheid data over zoekactiviteiten, locaties en sociale contacten. Daarom zou toegang tot de data van banken afhankelijk kunnen worden gesteld van het verlenen van datatoegang aan banken. Het loont, met andere woorden, om de uitwisselingsovereenkomst wederkerig te maken: *"u krijgt toegang tot (al) mijn data als ik toegang krijg tot (al) uw data"*. Meer in het algemeen zorgt het opleggen van dit soort wederkerigheid bij datadeling ervoor dat technologiebedrijven die actief willen worden in andere sectoren meer van hun data uitwisselen (wat bijdraagt aan kans II). In andere gevallen creëren de hierboven genoemde kansen op zich al een winstsituatie. Zo draagt horizontale datadeling bijvoorbeeld bij aan het overstappen van klanten, zoals in de markt voor administratieve bedrijfssoftware (kans III).

Alvorens verplichte datadeling te overwegen moet worden ingegaan op de randvoorwaarden die een vruchtbare voedingsbodem vormen voor vrijwillige datadeling. Een aandachtspunt is dat er voor datadeling een infrastructuur en een governancestelsel nodig zijn om datatoegang en -overdrachten te beheren en te orkestreren. Deze taken kunnen op verschillende manieren worden uitgevoerd. In een verticale omgeving kan een bedrijf bijvoorbeeld een gecentraliseerde architectuur gebruiken voor datadeling, zoals een sociaal netwerk dat data uitwisselt met downstream ontwikkelaars. In andere situaties worden deze functionaliteiten gedecentraliseerd uitgevoerd via een peer-to-peer architectuur, bijvoorbeeld door horizontale concurrenten die data uitwisselen met elkaar, of door banken die data uitwisselen met toetreders. Deelnemers vanuit de bedrijfstak kunnen ook een aparte entiteit creëren voor datadeling.

In al deze situaties brengt het beheren en orkestreren van datadeling verantwoordelijkheden met zich mee, zoals ondersteuning van innovatie, gebruikersmobiliteit en concurrentie, waarbij de rechten en belangen van scheppers van data in acht moeten worden genomen. Het dragen van dergelijke verantwoordelijkheden dient het publieke belang, maar sluit ook aan bij bedrijfsmodellen van zelfstandige platformen die datadelen met ecosysteemdeelnemers om zo veel mogelijk waarde voor en door deelnemers te (laten) creëren. Deze verantwoordelijkheden worden niet automatisch ondersteund door concurrenten. Daarom is het een no-regret optie om datadelingspraktijken te onderwerpen aan duidelijke principes, waardoor vertrouwen wordt gewekt en een gelijk speelveld wordt gegarandeerd. Door dergelijke principes over de gehele linie toe te passen, worden mogelijke belemmeringen voor vrijwillige uitwisseling weggenomen en wordt de noodzaak van potentiële verplichtingen minder urgent.

Hiertoe stellen wij voor dat **principes van 'good governance' van datadeling** (*Good Data Sharing Governance* - GDSG) worden vastgelegd in algemene wetten:

1. **Bescherming van de rechten van scheppers van data.** Scheppers van data — zowel personen als bedrijven — moeten de mogelijkheid hebben om (i) te bepalen welke data anderen over hen verzamelen, en (ii) het gebruik van hun data te controleren. Dus waar de AVG individuele personen al in staat stelt om de toegang tot persoonsdata te controleren, moeten bedrijven en organisaties ook in staat worden gesteld om toegang tot data te

controleren die naar hen kan worden herleid (in het bijzonder informatie met strategische waarde). Het doel is in de eerste plaats om bij scheppers van data vertrouwen te wekken om data uit te wisselen en/of anderen in staat te stellen data over hen te verzamelen, en ten tweede om klanten te ondersteunen bij overstappen en multi-homing.

2. **Ondersteuning van een gelijk speelveld.** Om randvoorwaarden te creëren voor de vele mogelijkheden met betrekking tot concurrentie en innovatie in bedrijfsmodellen en waardeproposities, moet er gelijke toegang tot data zijn. Daarvoor gelden de volgende voorwaarden:
 - a. *Non-discriminatie op basis van transparantie:* het bedrijf dat datadeling orkestreert moet gelijke voorwaarden toepassen op downstreambedrijven die de data ontvangen, ongeacht of het downstreambedrijf een eigen dochteronderneming is of een derde partij. Door deze voorwaarde neemt het risico op marktafscherming door verticaal geïntegreerde bedrijven af. De voorwaarde moet gelden voor alle relevante aspecten, zoals de kosten van datatoegang, technische haalbaarheid en continue real-time toegang (aangezien een bedrijf dat data bezit of beheert continue real-time toegang heeft, mogen anderen niet met een achterstand beginnen).
 - b. *Wederkerigheid van toegang tot datadeling:* een bedrijf dat om (toegang tot) data verzoekt moet transparant zijn over de eigen data en openstaan voor uitwisseling daarvan met de andere partij of, afhankelijk van de implementatie van datadeling, met een centrale architectuur voor datadeling. Hiertoe zijn mogelijk data-audits benodigd. Wederkerigheid voorkomt dat bedrijven die om datatoegang vragen bestaande datavoordelen te gelde kunnen maken in andere markten. Het vermindert ook het risico op het kantelen van een markt door leereffecten en genereert netwerkeffecten ten voordele van ieders innovatievermogen. (Zoals hierboven uitgelegd, leiden financiële compensatiemechanismen voor gegevenstoegang niet tot dergelijke voordelen).

Principe 1 kan worden geïmplementeerd via algemene wetgeving die van toepassing is op elke entiteit die data verzamelt en verwerkt. Met betrekking tot persoonsgegevens (verstrek/waargenomen) verplicht de AVG feitelijk al de toepassing van principe 1. **Aangezien de AVG niet van toepassing is op bedrijfsdata, is het nuttig om een zakelijk equivalent van de AVG in te voeren.**

Principe 2 kan ook worden geïmplementeerd via algemene wetgeving die geldt voor datadeling tussen partijen. Het lijkt wellicht praktisch om principe 2 op te nemen in een herziene EU-verordening met betrekking tot platform-to-business (P2B)-relaties, maar deze verordening beperkt zich tot online platforms. Een andere kandidaat is de EU-verordening betreffende vrij dataverkeer, maar die verordening heeft uitsluitend betrekking op niet-persoonsgebonden data. Daarom **is een nieuwe en meer algemene verordening noodzakelijk**. Het is nuttig om de mogelijkheden te verkennen voor opname in de op handen zijnde wetgevingsinitiatieven van de EU, die zijn aangekondigd in de Europese datastrategie (gepubliceerd door de Europese Commissie in 2020). De aanstaande initiatieven (waaronder een mogelijke Datawet) zijn bedoeld om B2B-datadeling te vergemakkelijken en te ondersteunen.

Beide principes moeten primair **van toepassing zijn op verstrekte/waargenomen data**, aangezien deze soorten data minimale risico's op reverse engineering en mogelijke schendingen van bepaalde intellectuele eigendomsrechten met zich meebrengen. Mochten dergelijke risico's zich toch voordoen, dan moeten uitzonderingen op principe 2 mogelijk zijn. De oorspronkelijke verwerkingsverantwoordelijke dient dan de bewijslast te dragen en objectieve redenen tegen uitwisseling te onderbouwen (bijv. efficiëntie, privacybescherming, mogelijke schendingen van bepaalde intellectuele eigendomsrechten, risico's op reverse engineering). We verwachten dat

deze gevallen zeldzaam zullen zijn bij verstrekte/waargenomen data, die geen verband houden met of niet herleidbaar zijn tot scheppers van data. Bovendien is principe 1 van toepassing wanneer ze wel herleid kunnen worden tot scheppers van data.

De GDSG-principes richten zich voornamelijk op het hold-up probleem, dat een belemmering kan zijn voor belanghebbenden om data uit te wisselen binnen en tussen waardeketens. Ze geven vorm aan de ontwikkeling van vertrouwde omgevingen voor datadeling, maar verplichten belanghebbenden in eerste instantie niet om data uit te wisselen. Een brede acceptatie van deze principes stimuleert alle belanghebbenden in de dataeconomie echter wel om deel te nemen aan uitwisseling van ecosystemen, met name vanwege de wederkerigheidsclausule, met netwerkeffecten als resultaat. Wanneer bedrijven ondanks deze principes nog steeds niet bereid zijn tot verticale datadeling, kan het nodig zijn om verplichtingen op te leggen, vooral in die gevallen **waar bedrijven hun concurrentiepositie te danken hebben aan exclusieve toegang tot unieke data** en (om defensieve redenen) niet bereid zijn om data uit te wisselen met andere belanghebbenden binnen en tussen waardeketens. Een dergelijke terughoudendheid belemmert concurrentie en innovatie. **Sectorspecifieke verplichtingen tot verticale datadeling** zijn dan misschien noodzakelijk.

Beleidsdoelstelling II: aanpak van concurrentieproblemen die voortkomen uit netwerk- en leereffecten

Naast het stimuleren van innovaties en het mobiliseren van gebruikers (kansen I, II en III hierboven), kan verplichte datadeling helpen om winner-takes-all-effecten te voorkomen. Meer in het bijzonder kan een verplichting tot horizontale datadeling per geval bijdragen tot concurrerende marktresultaten wanneer:

1. data leiden tot leer- en/of netwerkeffecten (d.w.z. gedragsdata, sociale data, profieldata), en
2. deze data uniek zijn en exclusief ter beschikking staan van een marktleider (d.w.z. niet replicerbaar zijn door een concurrent).

Een beslissing per geval om horizontale uitwisseling te verplichten zou rekening moeten houden met de volgende criteria:

3. Kunnen concurrenten met een klein (zeg 5 tot 10%) marktaandeel de dataschaal die benodigd is om gelijke tred te houden met de leider niet realiseren? Als ze dat niet kunnen (het rapport geeft aan wanneer dit risico kan optreden), dan kan een datadeelplicht nodig zijn om een level playing field te herstellen.
4. Past de marktleider (andere) concurrentiebepurende strategieën toe om zijn positie af te schermen? In dat geval moet dit gedrag geadresseerd worden om een datadeelplicht effectief te laten zijn. Echter, wanneer de exclusiviteit van data het gevolg is van dit concurrentiebepurend gedrag, kan het voldoende zijn dit gedrag direct via de mededingingswet aan te pakken, waardoor een datadeelplicht (wat een minder direct en minder doelgericht middel is) niet meer nodig is.
5. Brengt datadeling risico's op reverse engineering met zich mee? Dit is met name relevant voor gedragsdata. In dergelijke gevallen moet men terughoudend zijn met het opleggen van een datadeelplicht. Zoals hierboven gesteld, draagt de oorspronkelijke verwerkingsverantwoordelijke de bewijslast om een objectieve redenen voor het weigeren van datadelen te onderbouwen, waaronder het mogelijk overtreden van bepaalde intellectuele eigendomsrechten en het risico op reverse engineering.

Nadat bovengenoemde omstandigheden in overweging zijn genomen, moet een verplichting om specifieke soorten data horizontaal uit te wisselen (bijvoorbeeld geanonimiseerde zoekopdrachten en klikgedrag bij zoekmachines) in de eerste plaats worden opgelegd aan de marktleider. Om een

haasje-over-effect met betrekking tot het marktleiderschap echter te voorkomen, **moeten alle concurrenten de specifieke soorten data die het probleem veroorzaken onderling uitwisselen, opdat geen enkele speler een datavoordiel kan ontwikkelen**. De bevoegde autoriteit dient te bepalen welke data moeten worden uitgewisseld. Dit is een beperktere definitie van wederkerigheid dan vermeld in relatie tot de eerste van de GDSG-principes, die van toepassing moet zijn op alle en verschillende soorten data tussen verschillende belanghebbenden.

Een dergelijke datadelingsregeling kan wellicht worden gerealiseerd onder het mededingingsrecht, op voorwaarde dat wederkerigheid als een ‘redelijke toegangsprijs’ wordt beschouwd, en dat een groepsvrijstelling onder Artikel 101 van het Verdrag betreffende de werking van de Europese Unie (VWEU) het mogelijk maakt voor concurrenten om data met elkaar uit te wisselen. Bovendien moeten de verplichtingen in overeenstemming zijn met de bepalingen in de AVG wanneer het gaat om data die niet kunnen worden geanonimiseerd zonder waardeverlies (met name sociale en profieldata). Op grond van de AVG kan een wettelijke verplichting die wordt opgelegd op basis van andere wetgeving (zoals het mededingingsrecht) de rechtsgrond vormen voor de uitwisseling van persoonsdata zonder toestemming.

Aanvullend beleid

Om de beleidsdoelstellingen te realiseren is het niet altijd voldoende om datadeling te stimuleren en/of verplicht te stellen. Aanvullende beleidsopties kunnen nodig zijn. Een algemene beleidsaanvulling is het voorkomen van concurrentiebeperkende strategieën, gericht op het monopoliseren van verwante of parallelle markten nadat een markt is gekanteld. Soms is sectorspecifiek beleid noodzakelijk als aanvulling op het datadelingsbeleid, zoals bijvoorbeeld een lichte vorm van een interoperabiliteitsverplichting voor socialemediadiensten.

Samenvattende infographic

We sluiten het hoofdstuk af met een figuur waarin de hierboven besproken beleidsopties worden samengevat.



1 Inleiding

Doelstelling en onderzoeksvraag

De belangrijkste vraag die door dit rapport wordt beantwoord luidt: **wat is de rol van data in de technologiesector en wat zijn de kansen en risico's op verplichte datadeling?** Het antwoord op deze vraag moet leiden tot inzichten in mogelijke varianten van verplichtingen tot datadeling tussen bedrijven op verschillende internetmarkten ten behoeve van het algemeen belang.

Wat is de technologiesector?

Er is geen duidelijke definitie van de technologiesector vergeleken met sectoren als elektriciteit, financiële diensten, logistiek enzovoort. Volgens Wikipedia en Investopedia heeft de technologiesector betrekking op de ontwikkeling en levering van op elektronica gebaseerde technologieproducten (bijvoorbeeld smartphones, zelfrijdende voertuigen) en -diensten (bijvoorbeeld navigatie, communicatie, cloudcomputing). Naar onze mening is deze definitie in het kader van dit onderzoek zowel te ruim als te beperkt. De definitie is te ruim in de zin dat bijvoorbeeld een bedrijf als Foxconn (qua omzet het op twee na grootste technologiebedrijf)¹, tot de sector zou behoren, terwijl dat bedrijf voornamelijk leverancier is van computeronderdelen en zich niet bezighoudt met data-analyse. De definitie is te beperkt voor onze doeleinden aangezien deze een bedrijf als Amazon lijkt uit te sluiten (Fortune 500 classificeert Amazon als detailhandelsbedrijf in plaats van als technologiebedrijf)².

De focus van dit onderzoek ligt op bedrijven die in hoge mate afhankelijk zijn van data in een context van sterke netwerkeffecten. Met (data)technologie orkestreren ze interacties binnen en tussen waardeketens. Met andere woorden, dit onderzoek richt zich primair op digitale platforms die (potentieel) op mondiale schaal opereren, met data als (zeer) belangrijke productiefactor. Daaronder vallen uiteraard bedrijven als Amazon, Apple, Facebook, Google en Microsoft, maar ook bedrijven als John Deere, Nike en Honeywell, die hun traditionele producten hebben omgezet in apparatuur om data te verzamelen ten behoeve van platforms voor data-analyse en innovatie. Het omvat in feite elk bedrijf dat datatechnologie gebruikt voor orkestratie en innovatie.

De focus ligt op digitale platforms omdat netwerk- en lock-in-effecten een sterke factor zijn in dit type bedrijfsmodel, wat grote risico's met zich mee kan brengen op negatieve effecten van marktmacht op concurrentie en innovatie. Digitale platforms kenmerken zich door hun intermediaire functie en bijbehorende netwerkeffecten. Vanwege die effecten neemt de totale waarde die gebruikers aan een platform ontlene exponentieel toe met het aantal gebruikers. Met behulp van data kan een platform zijn intermediaire functie beter vervullen, waardoor de netwerkeffecten sterker zijn, het aantal deelnemers sneller groeit, en het platform weer meer data kan verzamelen (waarmee de cirkel rond is). Data maken het ook gemakkelijker om de diversiteit van diensten, aangeboden door derden of door het platform zelf, te verbreden. Meer diversiteit genereert meer variatie aan data. Dit zorgt voor meer differentiatie, waardoor het platform nog interessanter wordt voor gebruikers en waardoor nog meer data kunnen worden verzameld (de cirkel is wederom rond). Deze circulaire groeidrivers zorgen ervoor dat digitale markten vaak door de aanwezigheid (zeer) grote spelers worden gekenmerkt, die mogelijk als poortwachters optreden. Dit leidt tot zorgen over concurrentie, innovatie en keuzevrijheid.

Aan de hand van een aantal casussen kijken we specifiek naar sociale media (zoals Facebook en LinkedIn), zoekdiensten (zoals Google en Bing) en bedrijfsproces- en administratiesystemen (zoals

¹ https://en.wikipedia.org/wiki/Lijst_van_qua_omzet_grootste_techologiebedrijven

² <https://fortune.com/global500/2019/amazon-com/>

Oracle, SAP en Microsoft). Daarnaast besteedt het onderzoek specifieke aandacht aan de banksector. Banken zijn (nog) geen digitale platforms, maar de recente verplichting tot datadeling (PSD2) maakt het nu gemakkelijker voor anderen (waaronder big techs) om hun markt te betreden. De lessen die uit deze casussen kunnen worden getrokken leveren een bijdrage aan de algemene conclusies.

Ten tweede moeten we duidelijkheid scheppen over het publieke belang waarvan hier sprake is. We richten ons op de impact van datadeling op innovatie, productiviteit, keuzevrijheid en concurrentie. In alle gevallen hebben we oog voor negatieve neveneffecten en afwegingen. Zo kan er sprake zijn van een afweging tussen productiviteitswinst en concurrentie (en keuzevrijheid) versus verminderde beveiliging en privacy. Innovatie kan aan beide kanten van de schaal terecht komen. Enerzijds stimuleert datadeling innovatie (omdat meer partijen data kunnen analyseren). Anderzijds kan verplichte datadeling leiden tot meeliften op en reverse engineering van innovatieactiviteiten. Dit schaadt innovatiestimulansen.

Ten slotte onderscheiden we horizontale datadeling (tussen aanbieders van vergelijkbare diensten), verticale datadeling (tussen partijen met een verticale relatie in de waardeketen), en sector-overschrijdende datadeling (tussen partijen die geen vergelijkbare diensten aanbieden en niet actief zijn in dezelfde waardeketen). Deze classificatie geeft de indruk dat internetmarkten duidelijk kunnen worden gedefinieerd, wat uiteraard niet het geval is. Zo strijden Google en Facebook enerzijds om de "ogen en oren" van eindgebruikers, en anderzijds om advertenties. Advertentiediensten lijken ook steeds meer op e-commercediensten naarmate bedrijven minder bereid zijn om te betalen voor "klikken" en meer bereid om te betalen voor aankopen. Ondanks vervagende marktgrenzen is het onderscheid tussen horizontaal, verticaal en sector-overschrijdend goed voor het inzicht in de impact van datadeel- en portabiliteitsverplichtingen. Conclusies over marktdefinities kunnen op basis van het onderscheid echter niet worden getrokken.

Specifieke onderzoeksvragen

De subvragen die in deze studie worden beantwoord zijn als volgt:

1. Wat is de rol van data in de technologiesector? Welke soorten data zijn belangrijk voor innovatie en concurrentie?
2. Welke mogelijkheden biedt bredere datadeling (horizontaal, verticaal en sector-overschrijdend) voor bedrijven op specifieke internetmarkten?
3. Vindt deze datadeling ook plaats zonder dat de overheid een rol speelt in markten of gemeenschappen? En zou datadeling plaats kunnen vinden met gerichte financiële of organisatorische ondersteuning vanuit de overheid?
4. Welke (wettelijke) rechten / verplichtingen op het gebied van datadeling zijn al van toepassing? Welke mogelijkheden hebben partijen, bijvoorbeeld via het recht op dataportabiliteit krachtens de AVG, om datadeling of toegang tot data af te dwingen?
5. In hoeverre zijn – per markt - de criteria van toepassing die de overheid in haar visie op datadeling tussen bedrijven heeft aangemerkt als criteria voor verplichte datadeling?
 - a. Draagt potentiële datadeling bij aan het publieke belang?
 - b. Zijn de data niet eenvoudig te (re)produceren of te verzamelen door andere partijen dan de partij(en) waarvoor verplichte datadeling wordt overwogen?
 - c. Lukt het belanghebbenden niet of niet voldoende om op vrijwillige basis passende afspraken te maken over toegang tot de data?
 - d. Worden stimulansen voor investeringen in data niet substantieel verminderd door de invoering van verplichte datadeling?
 - e. Zijn de implicaties voor het publieke belang zorgvuldig afgewogen?
6. Welke specifieke varianten van verplichte datadeling in de technologiesector zijn te overwegen?

- a. Voor welke data zou een uitwisselingsverplichting zinvol zijn en waarom? Hoe moet worden bepaald om welke data het precies moet gaan?
 - b. Voor welke partijen zou verplichte datadeling zinvol zijn? En met welke partijen kunnen de data worden uitgewisseld? Hoe kunnen deze partijen worden onderscheiden?
 - c. Onder welke voorwaarden kunnen de data worden uitgewisseld? Hoe moet een redelijke prijs worden bepaald?
7. Welke juridische, economische en technische consequenties zijn te verwachten van de verschillende varianten van verplichtingen tot datadeling?

Opbouw van het rapport

In hoofdstukken 2 en 3 wordt vraag 1 in detail besproken. Hoofdstuk 2 schetst de rol van data in de digitale economie. Op basis van bestaande literatuur wordt in dit hoofdstuk een algemene beschrijving gegeven van bedrijfsmodellen van digitale platforms en hoe deze concurreren. Het wordt duidelijk dat data een belangrijke input zijn voor in elkaar grijpende circulaire groeiders van digitale platforms: netwerkeffecten (meer gebruikers trekken meer gebruikers aan), leereffecten (meer gebruikers genereren meer data, meer informatie, hogere kwaliteit, meer gebruikers en — terug naar het begin — meer data) en breedtevoordelen (meer gebruikers nodigen uit tot meer variatie in diensten, wat meer gebruikers aantrekt). Als gevolg hiervan kan een kleine voorsprong in het aantal gebruikers en data een kanteling van de markt teweegbrengen, met een monopolie tot gevolg. Sommige auteurs benadrukken het risico dat een datavoorsprong in één markt leidt tot monopolisering van andere markten als gevolg van een sneeuwbaaleffect. Anderen daarentegen betogen dat digitale markten worden blootgesteld aan "mologopolistische" concurrentie, een vorm van dynamische marktoverschrijdende concurrentie tussen een aantal grote bedrijven en een lange rits kleine innovators.

In hoofdstuk 3 wordt dieper ingegaan op de eerste vraag door middel van een bespreking van de verschillende soorten data, hoe digitale platforms die data verwerven en wat ze ermee doen. In het hoofdstuk wordt per soort data besproken of het bedrijfsmodel al dan niet aanspoort tot de uitwisseling van deze data (horizontaal, verticaal of sector-overschrijdend) en welke belemmeringen er zijn voor datadeling. In de regel zijn platforms niet gemotiveerd om horizontaal data uit te wisselen: dit maakt gebruikers mobieler, maakt concurrenten sterker, en nodigt (soms) uit tot reverse engineering. Platforms en platformgebruikers willen meestal wel deelnemen aan verticale datadeling, omdat dit leidt tot efficiëntieverhoging en innovaties.

In hoofdstuk 4 wordt vraag 4 beantwoord door middel van een overzicht van bestaande wettelijke kaders die datadeling faciliteren, voorschriften leveren voor het verplicht stellen van datadeling en grenzen vaststellen voor datadeling.

In hoofdstuk 5 worden vragen 2, 3, 4 en 5 integraal beantwoord in relatie tot vijf casussen (zoekmachines, sociale media, betaaldiensten, administratieve bedrijfssoftware en verticale uitsluiting). In dit hoofdstuk worden de criteria van de Nederlandse overheid toegepast om de wenselijkheid van verplichte datadeling te beoordelen. Ook wordt ingegaan op de kansen die datadeling biedt voor bedrijven en het publieke belang (vraag 2), mogelijkheden voor vrijwillige datadeling (vraag 3), en mogelijke negatieve gevolgen voor investeringen en andere publieke belangen. Per casus bespreken we bestaande wettelijke rechten en plichten ten aanzien van datadeling (vraag 4). De casestudy's in Hoofdstuk 5 zijn gebaseerd op vijf uitgebreide achtergrondstudies, die zijn opgenomen in een bijlage. We merken op dat de achtergrondstudies input hebben geleverd voor de casestudy's die in dit rapport zijn opgenomen, maar dat ze geen deel uitmaken van dit rapport. De enige reden om deze achtergrondstudies als bijlage op te nemen,

is om de geïnteresseerde lezer de inhoud van deze achtergrondstudies niet te ontzeggen (het delen van kennis is immers als het delen van data: daar hebben we allemaal baat bij)³.

In hoofdstuk 6 worden de voorgaande hoofdstukken samengebracht ter beantwoording van vragen 6 en 7. In dit hoofdstuk worden de economische kansen besproken die datadeling biedt ten aanzien van innovatie, concurrentie en keuzevrijheid, gevolgd door een bespreking van beleidsopties en hoe deze juridisch kunnen worden geïmplementeerd. Het hoofdstuk wordt besloten met een kort overzicht van de technische implicaties van datadeling.

Onderzoeksmethode

De algemene onderzoeksmethodiek bestaat uit de volgende onderdelen:

- Onderzoek naar algemene inzichten uit de literatuur (voornamelijk economisch en juridisch) en deskundigenrapporten.
- Onderzoek naar empirische bevindingen uit de wetenschappelijke literatuur (empirische economische studies, jurisprudentie met name op het gebied van het mededingingsrecht) en openbare bronnen (nieuws, blogs).
- Een verkenning van vijf casussen, gebaseerd op vijf achtergrondstudies.

De achtergrondstudies (en casussen) zijn gebaseerd op:

- Bestudering van theoretische academische literatuur en van economische en juridische deskundigenrapporten met betrekking tot de casussen.
- Bestudering van empirische economische onderzoeken, jurisprudentie (met name op het gebied van het mededingingsrecht), en anekdotisch bewijs aangetroffen in openbare bronnen (nieuws, blogs) met betrekking tot de casussen.
- 13 semigestructureerde diepte-interviews met marktpartijen, bedrijfsverenigingen en toezichhouders. Semigestructureerd betekent dat in grote lijnen soortgelijke vragen zijn gebruikt, aangevuld met casus-specifieke vervolgvragen.

Met het oog op de wenselijkheid van zo groot mogelijke openheid van de geïnterviewde partijen, hebben we interviewpartners ervan verzekerd dat hun identiteit of individuele achtergrond niet bekend zou worden gemaakt. Drie partijen behoren tot de groep GAFAM⁴. Daarnaast hebben we kleinere platforms en gebruikers van zakelijke platforms geïnterviewd. We hebben dus gesproken met zowel gevestigde als nieuwe spelers, en met partijen met mogelijk verschillende belangen.

Als gevolg van de semigestructureerde aanpak van de interviews werden dezelfde vragen meerdere keren gesteld: in verschillende sectoren en aan partijen met verschillende visies en belangen. In deze interviews zijn we begonnen met een set veelvoorkomende vragen, die soms werden gevolgd door casus-specifieke discussies. Waar nodig vroegen we interviewpartners om materiaal te leveren in de vorm van empirisch en/of anekdotisch bewijs, ter ondersteuning en ter illustratie van uitspraken in de interviews.

Nader omschreven was de onderzoeks aanpak voor de achtergrondstudies en casussen als volgt:

1. Op basis van een analyse van algemene economische en juridische inzichten, alsmede van bestaande empirische waarnemingen, hebben we de reikwijdte van de achtergrondstudies omschreven en hypothesen en vragen geformuleerd. De achtergrondstudies bestrijken het spectrum van horizontale, verticale en sector-overschrijdende B2B-datadeling.

³ Voorts merken we op dat de achtergrondstudies in de bijlage enkel in het Engels beschikbaar zijn aangezien deze door een internationaal team van experts is uitgevoerd.

⁴ Google, Apple, Facebook, Amazon, Microsoft

2. Aan de hand van de interviews en beschrijvende bronnen (nieuws, blogs) hebben we onderzocht wat datadeling betekent in de praktijk van de casussen en welke relevante ontwikkelingen zich de afgelopen jaren hebben voorgedaan.
3. Uitgaande van de structuur van de interviewvragen, gingen de achtergrondstudies vervolgens dieper in op de specifieke casussen.
4. Op basis van deze achtergrondstudies konden we vraag 2, 3, 4 en 5 beantwoorden met betrekking tot de vijf casussen.
5. De casussen in dit rapport verschaffen inzicht in zowel de sectoren in kwestie (bijvoorbeeld met betrekking tot bedrijfsmodellen, relevante soorten data en het belang van data voor innovatie) als in overkoepelende vragen over datadeling. Op basis van overeenkomsten en verschillen tussen de casussen hebben we vijf situaties onderscheiden waarin datadeling gerechtvaardigd is (horizontaal, verticaal of sector-overschrijdend). We hebben daarnaast de beleidsimplicaties voor elk van deze vijf situaties onderzocht.

2 De rol van data in de digitale economie

2.1 De rol van data in digitale bedrijfsmodellen

Steeds meer bedrijven passen platform-gebaseerde bedrijfsmodellen toe of bouwen hun bedrijfsmodel op het concept van een platform. Een (meerzijdig) platform kan worden omschreven als een bedrijf met meerdere gebruikersgroepen dat interactie tussen die groepen faciliteert en aan diverse soorten gebruikersgroepen verschillende prijzen in rekening brengt, terwijl er platform-overschrijdende netwerkeffecten zijn (zie hieronder)⁵. Een sociaal netwerk is een prominent voorbeeld van een dergelijk platform.

Data vormen een essentieel onderdeel van de 'groeimotor' van een digitaal platform, omdat ze een belangrijke input zijn voor leereffecten. Leereffecten zijn een van de drie aanjagers van de groei van een platform. Andere groeifactoren zijn netwerkeffecten (schaalvoordelen aan de vraagzijde) en breedtevoordelen:⁶

- Netwerkeffecten (direct en indirect of platform-overschrijdend): meer gebruikers maken het platform aantrekkelijker voor nieuwe gebruikers.
- Breedtevoordelen: een groter aantal gebruikers ondersteunt de invoering van nieuwe diensten, waardoor het platform aantrekkelijker wordt voor nieuwe gebruikers.
- Leereffecten: een groter aantal gebruikers genereert meer nuttige data en verbetert functionaliteiten. Dit maakt het platform aantrekkelijker voor nieuwe gebruikers.

Netwerkeffecten verwijzen naar het exponentiële verband tussen het aantal gebruikers en hun interactiemogelijkheden.⁷ Het aantal interactiemogelijkheden is een graadmeter voor de waarde die een individuele gebruiker aan een platform ontleent. Uitgaande van positieve netwerkeffecten, groeit de waarde die een individu aan een lidmaatschap hecht met elke nieuwe gebruiker met wie interactie mogelijk is. Op een gegeven moment begint de relevantie van meer gebruikers terug te lopen en begint deze relatie af te vlakken. Door investeringen in infrastructuur en governance-algoritmen (zoals zoekfilters) kan een platform dit keerpunt uitstellen, aangezien die individuele personen helpen om relevante interactiemogelijkheden te identificeren.

Breedtevoordelen hebben betrekking op verwante functionaliteiten en diensten. Als een platform eenmaal veel gebruikers heeft, kan het interacties tussen zijn gebruikers onderling en met dienstenaanbieders faciliteren. Dit vergroot de reikwijdte van interactiemogelijkheden wat weer meer gebruikers aantrekt. Meer gebruikers genereren meer data. Deze zijn belangrijk voor *profileringstechnieken*⁸ en filters voor de relevantie van 'matches' tussen gebruikers. Meer data verbeteren de kwaliteit van de gebruikerservaring, ondersteund door kunstmatige intelligentie. De verbeterde gebruikerservaring trekt vervolgens nieuwe gebruikers aan, waardoor algoritmen sneller kunnen leren, waardoor de kwaliteit van de gebruikerservaring wederom wordt verbeterd en meer gebruikers worden aangetrokken. De circulaire waardeestroom als gevolg van leereffecten ziet er daarom uit als in Figuur 2.1 hieronder.

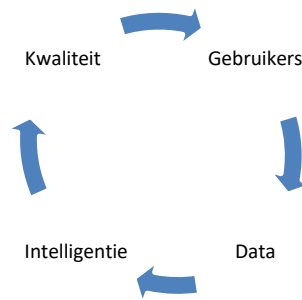
⁵ (Katz & Sallet, 2018).

⁶ (Van Gorp & De Bijl, 2019).

⁷ Indien elke gebruiker even relevant is en $f(n)$ het aantal interactiemogelijkheden bij n gebruikers representeert, dan is $f(1) = 0$ en $f(n) = f(n-1) + (n-1)$, een exponentieel toenemende functie. (De Bijl & Van Gorp, 2016-2019)

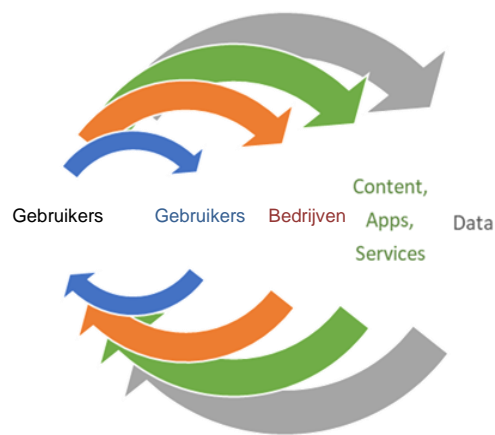
⁸ *Profiling* behelst het analyseren en verbinden van iemands interacties om op profielen gebaseerde trends te ontdekken, marktstatistieken te produceren en gedragsanalyses uit te voeren.

Figuur 2.1 Leereffecten als circulaire waardeestroom (De Bijl & Van Gorp, 2016-2019)



Netwerkeffecten, breedtevoordelen en leereffecten zijn aanjagers van groei met één gemeenschappelijke deler: de gebruikers. Als gevolg hiervan versterken de circulaire waardestromen elkaar en vormen ze samen een krachtige groeimotor met een vliegwieleffect - zie Figuur 2.2.

Figuur 2.2 Groeimotor van een platform (De Bijl & Van Gorp, 2016-2019)



Deze circulariteit zorgt voor een snelle marktgroei en een mogelijke kanteling van de markt. Bovendien stimuleert de groeimotor de ontwikkeling van digitale platforms tot ecosystemen die een breed scala aan diensten leveren, en mogelijk de ene na de andere markt laten kantelen (Prüfer & Schottmüller, *Competing with Big Data*, 2017).

Opgemerkt wordt dat er ook sprake kan zijn van leereffecten bij bedrijven met traditionele bedrijfsmodellen (bijvoorbeeld Tesla), maar het risico op het kantelen van de markt is dan aanzienlijk kleiner omdat ze niet inhaken op netwerkeffecten die niet aanwezig zijn bij lineaire bedrijfsmodellen.

2.2 De rol van data in digitale concurrentie

De rol van data in concurrentie

Het is duidelijk dat data een sleutelrol spelen in digitale concurrentie. Voor platforms met advertenties zijn data bijvoorbeeld een essentieel onderdeel van het verdienmodel. Daarnaast vormen data een essentiële input voor de intelligentie van algoritmen.

Veel platforms bieden online advertentieruimte aan. Deze aanbieders van online advertentieruimte onderscheiden zich wat betreft schaal (d.w.z. omvang van de deelnemersgroep) en hun vermogen

om zich te richten op individuen in de deelnemersgroep, overeenkomstig de criteria van de adverteerders. Hoewel deze criteria van adverteerder tot adverteerder verschillen, kunnen aanbieders van online advertentieruimte zich onderscheiden door zoveel mogelijk profieldata te verzamelen. De Bijl & Van Gorp (2018) leggen uit dat hoe intenser de concurrentie tussen aanbieders van advertentieruimte is, hoe meer zij geneigd zijn om profieldata te verzamelen. Deze datahonger stimuleert zowel de ontwikkeling van nieuwe diensten als innovatie in 'tracking technology'. Het eerste kan als positief worden beschouwd; het tweede is twijfelachtiger, omdat het de mogelijkheden beperkt voor privacygevoelige gebruikers om onder de radar te blijven.

De mogelijkheden om door middel van algoritmen de kwaliteit van intermediaire diensten te verhogen neemt toe naarmate er meer data beschikbaar zijn over gebruikers en hun contacten en interacties met andere gebruikers. Platforms gebruiken data ook om nieuwe functionaliteiten en diensten te introduceren. Met meer relevante data zet dit proces van data-gestuurd leren de groeimotor van het platform in werking (zie hierboven). Theoretische literatuur bevestigt het bestaan van leereffecten en het belang van schaal in het leerproces van algoritmen⁹. De empirische literatuur is echter zeer beperkt als het gaat om het relatieve belang van dit effect (vergeleken met andere factoren). Zo wordt de markt voor zoekmachines vaak genoemd als voorbeeld van een markt waarin datavoordelen (in combinatie met netwerkeffecten) hebben geresulteerd in het kantelen van de markt. De casus inzake zoekmachines toont aan dat Google's leidende marktpositie wellicht ook door andere factoren kan worden verklaard (zie hoofdstuk 5).

Terwijl sommige auteurs het risico benadrukken dat meerdere markten worden gedomineerd door één enkel bedrijf, merken anderen (Petit, 2018) op dat digitale markten niet worden getransformeerd tot afgeschermd monopolies. Ze omschrijven het concurrentieproces op digitale markten als 'moligopolistische concurrentie'. Dit is een vorm van dynamische concurrentie over marktgrenzen heen tussen enkele grote bedrijven en een lange rits kleine innovators. Concurrentie kenmerkt zich dan door digitale "monopolies" die blootstaan aan concurrentie van up- of downstream spelers, of door actoren uit totaal verschillende markten.

Bij concurrerende platforms is sprake van onderlinge '*intermediation*' en '*disintermediation*'¹⁰. Dit is de essentie van digitale concurrentie. Bedrijven streven ernaar om controle te krijgen over essentiële inhoud, data of gebruikersgroepen. *Intermediation* kan leiden tot poortwachterposities die door gebruikers moeilijk zijn te omzeilen als ze andere gebruikers of inhoud willen bereiken. Concurrenten van parallelle markten kunnen dergelijke posities echter bestrijden met *disintermediation*. Daarom wordt concurrentie niet zozeer gekenmerkt door rivaliteit tussen aanbieders van vergelijkbare diensten, maar veeleer door innovatie in toegangswegen tot essentiële inhoud, data of gebruikersgroepen. Marktgrenzen vervagen in dit concurrentieproces. Internetbrowsers kunnen bijvoorbeeld besturingssystemen disintermediëren door een alternatieve route te bieden voor toegang tot software en het bekijken van inhoud. Moligopolistische concurrentie impliceert dus dat monopolisten in parallelle markten (zoals de Big Tech-spelers) elkaar steeds vaker tegenkomen en elkaar bestrijden over traditionele marktgrenzen heen.

Startende en snelgroeiende bedrijven spelen een belangrijke rol in de concurrentie over marktgrenzen heen, aangezien ze vaak technologieën of bedrijfsmodellen introduceren die gevestigde posities betwisten. Na een overname door de Big Techs kan een startende of snelgroeiend bedrijf toegang krijgen tot een grote dataset die anders buiten bereik zou blijven. Hiermee kan het een groei-impuls krijgen. In ruil daarvoor krijgt het acquirerende bedrijf voet aan

⁹ Agrawal, Gans, & Goldfarb (2019), Prüffer & Schottmüller (2017), Mayer-Schönberg & Ramge (2018), Anderson (2020).

¹⁰ **Intermediatie:** Een platform plaatst zichzelf tussen gebruikers of gebruikersgroepen, ten behoeve van de intermediatie (en mogelijk het faciliteren) van matches of interacties. **Desintermediatie:** Een platform stelt gebruikers of gebruikersgroepen in staat om een bestaande intermediair te omzeilen, zodat zij directe interactie met elkaar kunnen hebben.

de grond in een nieuwe markt. Het vooruitzicht van een overname kan een innovatiestimulans zijn voor startende en snelgroeiende bedrijven. Prüfer & Schottmüller (2017) wijzen er echter op dat de Big Techs door middel van hun datavoordeel en klantenbestand de technologie van startende en snelgroeiende bedrijven zonder een overname kunnen imiteren en verbeteren. Dit vermindert de investeringsprikkel en geeft Big Techs een betere onderhandelingspositie bij het bepalen van de overnameprijs. Een data-voordeel door een gevestigd bedrijf, kan zo dus de prikkel voor onafhankelijke innovatie uithollen (Prüfer & Schottmüller, 2017). Bovendien kan een bedrijf op overnamepad gaan wanneer het zich bedreigd voelt door een nieuwkomer, om vervolgens de innovatie te “liquideren”. Dit laatste heeft uiteindelijk hetzelfde effect als het uithollen van de investeringsprikkel, maar is nog erger omdat er nu middelen worden besteed aan technologieën die hun weg naar de markt uiteindelijk toch niet vinden.

Data als belemmering voor concurrentie?

Als data, of preciezer gezegd data-gestuurd leren, een concurrentievoordeel oplevert, is het de vraag of exclusieve toegang tot data een moeilijk te overbruggen voordeel verschaft aan een gevestigd bedrijf. Van Til, Van Gorp en Price (2017) en ook Hagiu en Wright (2020) bieden een lijst vragen waarmee kan worden beoordeeld of data-gestuurd leren, naast andere bepalende factoren, bijdraagt aan duurzame concurrentievoordelen. Het is niet concurrentiebeperkend om een dergelijk voordeel na te streven, maar de concurrentie wordt mogelijk wel belemmerd wanneer bedrijven op hun lauweren kunnen rusten nadat ze een niet meer in te halen voorsprong hebben genomen. De belangrijkste vragen, die gelden voor alle soorten data, luiden als volgt:

1. Zijn de data **exclusief** en **uniek**? Dat wil zeggen, kunnen ze elders worden verkregen, gemakkelijk worden geïmiteerd, of leiden ze tot reverse engineering?
2. Zijn er **netwerkeffecten** gerelateerd aan het gebruik van de data? Helpen data van de ene gebruiker om het product te verbeteren voor anderen (wat een belangrijk voordeel is in de strijd om nieuwe klanten) of voor diezelfde gebruiker (wat leidt tot customer lock-in)?
3. Hoe snel nemen **data-gerelateerde schaalopbrengsten af**? Dat wil zeggen, is er een minimum efficiënte schaal en kunnen concurrenten die schaal gemakkelijk realiseren?
4. Hoe snel **devalueert** de relevantie van data **in de loop van de tijd**? Met andere woorden, is er behoefte aan constante (bijna) real-time updates?
5. Hoe kort zijn **leercycli** en hoe snel is de **adoptiegraad**? Dat wil zeggen, kunnen uit de data verkregen inzichten snel in producten/diensten worden geïntegreerd, of kost die integratie veel tijd, zodat concurrenten hun eigen R&D kunnen uitvoeren om vergelijkbare functionaliteiten te verkrijgen?
6. Hoeveel waarde voegt data toe aan de stand-alone waarde van het aanbod? Of, in ruimere zin, zijn er **andere bepalende factoren voor de concurrentiepositie van een bedrijf** en hoe belangrijk is de rol van data ten opzichte daarvan?
7. Hoe moeilijk is het om **innovaties te imiteren** op basis van de data, en zijn er bedrijven die **andere bedrijfsmodellen** gebruiken om te concurreren met het bedrijf dat over de data beschikt?

In hoofdstuk 5 analyseren we in hoeverre de criteria van de Nederlandse overheid voor het verplicht stellen van datadeling van toepassing zijn op specifieke gevallen, en behandelen we deze vragen in de casussen met betrekking tot zoekmachines en sociale netwerken.

2.3 Data en innovatie

Innovatie-ecosystemen

In voorgaande discussies is de term platform gebruikt in relatie met de functie ervan als intermediair. Een platform kan echter ook verwijzen naar een (technologische) basis voor het leveren van meerdere diensten. Van Gorp en Batura (2015) trekken de analogie met een chassis in de auto-industrie dat als basis dient voor meerdere automodellen. Er kan dus een onderscheid worden gemaakt tussen innovatieplatforms en interactieplatforms (Evans & Gawer, 2016). Evans en Gawer stellen vast dat het combineren van interactie- en innovatieplatforms online bedrijfsmodellen een impuls geven. Voorbeelden zijn het innovatieplatform *Apple iOS* in combinatie met het interactieplatform *App Store*, het interactieplatform van *Facebook* in combinatie met het innovatieplatform *Facebook for developers*, en besturingssysteem *Android* in combinatie met de *Google Play* app store. Het idee is dat het netwerkeffect van het interactieplatform de schaal- en breedtevoordelen van het innovatieplatform stimuleert (De Bijl & Van Gorp, 2016-2019).

Het begrip innovatieplatform heeft betrekking op de kenmerken van data als infrastructuur voor innovatie¹¹. Dit geldt des te meer voor 'big data', aangezien de diversiteit aan mogelijke toepassingen steeds groter wordt naarmate het volume, de variëteit en de snelheid van data toenemen. Vooral wat betreft variëteit is sprake van een exponentiële toename in het aanbod van mogelijke goederen en diensten,¹² die in omvang niet door mensen kan worden voorzien (zoals blijkt uit geautomatiseerd leren en kunstmatige intelligentie).

Datadeling om innovaties binnen ecosystemen te stimuleren

Uit bovenstaande discussie wordt duidelijk waarom digitale spelers hun besturingssystemen of datasets zouden willen openstellen voor derden, bijvoorbeeld door app-ontwikkelaars uit te nodigen om 'erbovenop' te innoveren. De daaruit voorkomende innovaties vergroten en verrijken hun ecosystemen. Platforms als Facebook en Strava¹³ wisselen bijvoorbeeld een aanzienlijke hoeveelheid data uit met downstream app-ontwikkelaars, om de gebruikerservaring binnen hun ecosystemen te verrijken en de betalingsbereidheid van eindgebruikers of adverteerders te bevorderen. Een ander voorbeeld is John Deere, dat landbouwmachines heeft uitgerust met sensoren om data te verzamelen (en een 'big dataset' te bouwen). Het bedrijf biedt klanten en app-ontwikkelaars toegang tot deze database¹⁴. Deze strategie stimuleert niet alleen de beschikbaarheid van innovatieve landbouwondersteunende diensten (wat de productiviteit van boeren ten goede komt), maar het stimuleert ook de verkoop van John Deere tractoren, net zoals de schaal en de reikwijdte van apps in de App Store de verkoop van Apple-apparaten stimuleren.

Het gaat in deze context om "verticale datadeling". In verticale relaties tussen platforms en hun partners (met name app-ontwikkelaars, makers van inhoud, en adverteerders) resulteren aanvullende voordelen in krachtige impulsen voor datadeling. Een platform biedt (deels) toegang data omdat het succes van een partner bijdraagt aan de waarde van het ecosysteem van het platform. De platformpartners hebben baat bij de toegang tot de data van een platform omdat dit de

¹¹ Frischmann (2012) definieert infrastructuur aan de hand van de volgende criteria:

1. Het middel kan worden gebruikt op een niet-concurrerende manier om tegemoet te komen aan een scala aan behoeften.
2. De maatschappelijke vraag naar het middel is voornamelijk het gevolg van downstream-productieactiviteiten die het middel nodig hebben als input.
3. Het middel kan worden gebruikt als input voor allerlei producten en diensten, waaronder particuliere, openbare en sociale producten.

¹² De overeenkomst tussen netwerkeffecten en leereffecten wordt snel duidelijk wanneer we de parameters in de formule in voetnoot 5 herdefiniëren als n = het aantal soorten data en $f(n)$ is het aantal correlatiemogelijkheden.

¹³ Strava is een sport-app met sociaalnetwerk-kenmerken.

¹⁴ <https://www.precisionag.com/digital-farming/data-management/john-deere-opens-data-platform-to-other-software-suppliers/>

ontwikkeling en distributie van diensten, inhoud of advertenties faciliteert. Toegang tot data versterkt netwerkeffecten en ondersteunt een groeistrategie voor het platform en zijn partners.

Het bedrijfsmodel van een verticale (downstream) partner kan in de loop van de tijd veranderen, bijvoorbeeld wanneer functies worden geïntroduceerd die overeenkomsten vertonen met kernfunctionaliteiten van het platform, of deze in geheel kopiëren. Als dat gebeurt wordt het upstream bedrijf bedreigd door horizontale concurrentie, wat uiteraard niet bijdraagt aan de groei van zijn ecosysteem; integendeel. In een situatie van horizontale concurrentie komt datadeling met aanbieders van een vergelijkbare dienst neer op het ondersteunen van concurrenten. Bedrijven zullen dit niet zo snel doen.¹⁵ Dus een platform dat wordt geconfronteerd met (of anticipeert op) een overgang van een verticale naar een horizontale relatie met een ander bedrijf zal veel minder geneigd zijn tot datadeling met dit bedrijf. Platformeigenaren zullen de substitutie- en complementariteitseffecten tussen hun eigen functionaliteiten en die van apps van derden afwegen. Daarom zijn bestaande (vrijwillige) datadeel-overeenkomsten hoofdzakelijk verticaal, aangedreven door wederzijdse belangen, tegen commerciële voorwaarden en voor beperkt gebruik (bijvoorbeeld niet-concurrerend met activiteiten van het platform zelf).

Gezien de sterke prikkels voor een platform om wel aan verticale datadeling te doen, is het verdacht wanneer het dit weigert. Een dergelijke weigering kan legitiem zijn, bijvoorbeeld wanneer datadeling in strijd is met de privacy. Bij gebrek aan een legitieme reden werkt een weigering om deel te nemen aan verticale datadeling meestal in het nadeel van concurrentie en innovatie, en kan dit worden gezien als een concurrentiebeperkende uiting van een machtspositie (het gedrag op zich duidt al op een machtspositie)¹⁶.

Datadeling om innovaties in de bredere economie te stimuleren

Impuls voor efficiëntie en innovaties door datadeling binnen sectoren

Bovengenoemd voorbeeld van John Deere is uitzonderlijk, omdat de meeste traditionele spelers niet gewend zijn aan datadeling en daartoe wellicht ook niet bereid zijn. Denk aan elektriciteitsbedrijven of banken die, via sectorspecifieke regulering, zijn gedwongen om data van slimme meters en bankgegevens uit te wisselen. Het doel van datadeling is om toegang tot de markt te verschaffen of om anderen innovatieve ideeën te laten ontwikkelen. Regelgeving is noodzakelijk, omdat de bedrijfsmodellen van deze bedrijven niet zijn gebaseerd op ecosystemen met downstream-innovators en positieve neveneffecten van datadeling. Traditionele bedrijven neigen ertoe data te zien als een belangrijke input voor hun bestaande diensten en een mogelijke bron voor lock-in van eindgebruikers, in plaats van als een bron voor innovatie door derden waarvan zij kunnen meeprofiteren.

Het verplicht stellen van toegang tot data is één manier om concurrentie en innovatie (binnen en tussen sectoren) te stimuleren. Een alternatief is om een ecosysteem te faciliteren waarin traditionele bedrijven kunnen profiteren van vrijwillige datadeling. De basis hiervoor is te vinden in netwerkeffecten, direct of indirect. Directe netwerkeffecten bij datadeling komen voort uit de complementaire aard van data wanneer deze worden uitgewisseld tussen gebruikers die data verzamelen en aanmaken, en tevens data gebruiken voor innovaties. Van Gorp (2018) geeft aan dat deelnemers een exponentiële groei ervaren van het aantal mogelijke correlaties dat ze kunnen verkennen met elke nieuwe deelnemer aan een datadeel-ecosysteem.¹⁷ Hierdoor wordt het, net als bij interactieplatforms, aantrekkelijker om deel te nemen aan een datadelingsplatform naarmate het

¹⁵ Een uitzondering is wellicht de (beperkte) datadeling voor specifieke doeleinden, zoals gezamenlijke O&O. In het algemeen helpt datadeling concurrenten echter om hun groeimotoren van brandstof te voorzien. Indien een gevestigd bedrijf een (bijna) monopoliepositie heeft als gevolg van netwerkeffecten, kan datadeling nieuwkomers helpen om het voordeel van de first-mover van het gevestigde bedrijf aan te pakken.

¹⁶ Zie Van Gorp en De Bijl (2019) voor uitgebreidere discussie.

¹⁷ Zie voetnoot 12

aantal deelnemers groter wordt. Wederkerigheid (d.w.z. ik wissel mijn data uit als u de uwe uitwisselt) is dan van cruciaal belang om directe netwerkeffecten in gang te zetten. Indirecte netwerkeffecten kunnen zich voordoen als bedrijven data verzamelen of aanmaken maar deze niet zelf analyseren. Indirecte netwerkeffecten kunnen dan in gang worden gezet door (zoals een app store) ontwikkelaars toegang te geven tot de data en door interacties mogelijk te maken tussen ontwikkelaars en de bedrijven die data verstrekken.

Een voorbeeld van dergelijke datadelingsplatformen is (wederom) te vinden in de landbouwindustrie (Join Data), maar ook in diverse andere sectoren, zoals de logistiek (iShare) en de zorg (MedMij). Belanghebbenden begrijpen dat een open datadelingsplatform (dat voor iedereen toegankelijk is) waardevoller is dan een besloten platform (zoals beheerd door John Deere). Dit inzicht heeft geleid tot overeenkomsten of gedragscodes voor eerlijke en transparante datadeling tussen verschillende partners in de waardeketen¹⁸. Een gedragscode is niet voldoende om netwerkeffecten in gang te zetten, maar het is wel een essentieel ingrediënt, naast standaarden en API's. Het platform Join Data¹⁹ komt dicht in de buurt van een algemeen toegankelijk datadelingsplatform waaraan een ontwikkelaarsplatform is gekoppeld,²⁰ en waar duidelijke regels gelden voor wie toegang heeft tot welke data.²¹

Aanboren van nieuwe markten door datadeling tussen bedrijfstakken

De in de landbouwsector waargenomen ontwikkelingen sluiten aan bij wat Nederlandse en Europese beleidsmakers wenselijk achten voor andere sectoren. Datadeling tussen bedrijfstakken en ecosystemen is echter minstens zo belangrijk voor innovatie als uitwisseling binnen sectoren. Data vormen een waardevolle input voor een breed scala aan bedrijfstakken (vrijwel alle bedrijfstakken) en het combineren van sector-overschrijdende data kan innovaties opleveren die niet te voorspellen zijn. Zo wordt het mogelijk om *big datasets* te creëren en grootschalige analyses van mogelijke correlaties uit te voeren. Sector-overschrijdende datadeling kan bijgevolg leiden tot onvoorziene ontdekkingen, waardoor weer nieuwe markten ontstaan. De Europese strategie voor datadeling bouwt voort op dit idee en beoogt een omgeving te creëren waar data worden uitgewisseld binnen en tussen bedrijfstakken. Het idee is dat er een innovatieplatform ontstaat waarop ondernemingen van uiteenlopende aard (klein, middelgroot, groot en supergroot) kunnen floreren door middel van datadeling, net zoals app-ontwikkelaars floreren in de ecosystemen van de Big Techs.

¹⁸ <https://www.medmij.nl/afsprakenstelsel/>

<https://www.ishareworks.org/>

<http://www.fao.org/family-farming/detail/en/c/1127623/>

¹⁹ <https://www.join-data.nl/>

²⁰ <https://www.join-data.nl/developers/>

²¹ Het is interessant dat het ecosysteem landbouwdata behandelt als persoonsdata krachtens de AVG: als de data kunnen worden herleid naar een bepaalde boerderij, moet de boer toestemming geven alvorens een partij toegang krijgt tot de data. Als de data niet kunnen worden herleid naar een bepaalde boerderij, dan kunnen de data vrijelijk worden gebruikt door ontwikkelaars.

3 Waarde van verschillende soorten data en drijfveren voor uitwisseling

In dit hoofdstuk wordt dieper ingegaan op de rol van data, door de verschillende soorten data aan de orde te stellen, de wijze waarop digitale platforms deze data verwerven en wat ze ermee doen. In het hoofdstuk wordt per soort data besproken of het bedrijfsmodel van een platform al dan niet een stimulans biedt voor datadeling (horizontaal, verticaal of sector-overschrijdend), en wat de belemmeringen zijn voor de bereidheid en het vermogen om uit te wisselen. De datatypologieën komen terug in de daaropvolgende bespreking van de waarde van data en de drijfveren voor uitwisseling.

3.1 Verschillende datatypologieën

3.1.1 Een op oorsprong gebaseerde typologie

In het kader van een discussie over privacy en beveiliging verschaft de OESO (2014)²² een datataxonomie gebaseerd op oorsprong. In haar richtlijnen met betrekking tot het recht op dataportabiliteit baseert de Groep Databescherming Artikel 29 zich ook op deze taxonomie om de reikwijdte te bepalen van het recht op dataportabiliteit krachtens Artikel 20 van de AVG. Deze taxonomie is van nut bij de beoordeling van datadelingsbeleid. Het onderscheid tussen datacategorieën leent zich er niet direct voor om het in een juridische vorm te gieten, omdat ze niet altijd nauwkeurig zijn omschreven, of elkaar uitsluiten.

Categorieën

De taxonomie onderscheidt vier categorieën: **verstrekte data**; **waargenomen data**; **verkregen data**; en **afgeleide data**²³. Het verschil tussen verstrekte en waargenomen data is dat 'verstrekte data' *actief* en *willens en wetens* zijn onthuld, gecreëerd of uitgewisseld door een gebruiker bij het gebruik van een dienst, en dat 'waargenomen data' *passief* zijn onthuld, gecreëerd of uitgewisseld door een gebruiker die zich *niet volledig bewust of geheel onbewust* is hiervan. Of dit verschil essentieel is voor beleidsdoeleinden hangt af van het geval in kwestie²⁴. Het verschil tussen verkregen en afgeleide data is dat verkregen data niet zijn gebaseerd op analyses van waarschijnlijkheid en dus kunnen worden beschouwd als '*harde feiten*', terwijl afgeleide data het product zijn van op waarschijnlijkheid gebaseerde redeneringen en kunnen worden beschouwd als '*mogelijke feiten*'. Ook dit verschil is wellicht niet altijd van belang voor beleidsdoeleinden, behalve wanneer ten foutief afgeleide data worden uitgewisseld waardoor de betrokkene schade ondervindt²⁵.

²² OESO (2014) "Samenvatting van de rondetafelconferentie van privacydeskundigen van de OESO, Bescherming van Privacy in een Data-gedreven Economie: Inventarisatie van het Huidige Gedachtegoed", Werkgroep voor Beveiliging en Privacy in de Digitale Economie, DSTI/ICCP/REG(2014)3, 21 maart 2014 (OECD, 2014)

²³ Zie (OESO. 2014. p. 5)

²⁴ Bijvoorbeeld in het geval van een eindgebruiker lock-in met een bepaalde wearable uitgerust met sensoren die hartslag, bloeddruk, aantal stappen enz. verzamelen, is het niet relevant of de data worden beschouwd als verstrekt of waargenomen. Het zijn de ruwe data die moeten worden uitgewisseld (of overgedragen) om de weg vrij te maken voor de keuzevrijheid van de consument. Bij een eindgebruiker lock-in op een clouddienstmarkt kan portabiliteit van verstrekte data, en in het bijzonder door gebruikers aangemaakte data, echter relevanter zijn dan portabiliteit van waargenomen data, aangezien het daarbij gaat om "*tekst-, geluid-, software- of beeldbestanden [...] die de gebruiker opzettelijk aanmaakt met gebruik van de applicatie of clouddienst, zoals documenten, verwerkte datasets, bewerkt beeldmateriaal, opgenomen geluiden enz.*" (ISO/IEC, 2017).

²⁵ Bijv. wanneer een algoritme uit transactiedata afleidt dat iemand een roker is, maar dat diegene in werkelijkheid wekelijks sigaretten koopt voor haar moeder. Wanneer deze data worden uitgewisseld met de zorgverzekeraar, kan de verzekeringspremie onterecht omhooggaan.

Hieraan zou nog de categorie **geproduceerde** data kunnen worden toegevoegd. Deze betreft data waarvoor specifieke investeringen waren vereist om deze te verzamelen, zoals het opzetten van een enquête, het organiseren van interviews, of het verzamelen van informatie via focusgroepen. Deze data verschillen van de andere categorieën omdat ze niet zijn verkregen of afgeleid en ze niet zijn verstrekt of waargenomen bij het gebruik van een dienst. We beschouwen deze categorie niet als een kandidaat voor verplichte uitwisseling, omdat er over het algemeen geen belemmeringen zijn voor het verzamelen van dergelijke data, en omdat het verplicht uitwisselen ervan de stimulans om ze te produceren ernstig zou ondermijnen.

Het deskundigenrapport '*Competition policy in the digital era*' uit 2019 (Crémer, de Montjoye, & Schweitzer, 2019) stelt dat een onderscheid tussen op oorsprong gebaseerde soorten data relevant kan zijn in het kader van de "essentiële faciliteiten-doctrine" in het mededingingsrecht, bijvoorbeeld om te beoordelen of data al dan niet onmisbaar en substitueerbaar zijn. Het rapport ziet dit onderscheid als conceptueel nuttig, maar stelt niet voor om het tot een wettelijk onderscheid te maken omdat het niet altijd duidelijk is en er grijze zones zijn. Deze grijze zones hebben niet zozeer betrekking op de definities van datacategorieën, maar veeleer op de mate waarin data substitueerbaar zijn en op de impact van uitwisseling op de stimulans om te investeren in data-analyse en dataverzameling.

Het is mogelijk dat er sprake is van overlap tussen verstrekte en waargenomen data, evenals tussen verkregen en afgeleide data²⁶, maar er bestaat een duidelijke afbakening tussen verstrekte en waargenomen data enerzijds en verkregen en afgeleide data anderzijds. Dit duidelijk onderscheid kan dus wel opgenomen worden in juridische teksten; bijvoorbeeld in privacywetgeving. Maar, het laatstgenoemd onderscheid zegt verder niets over de (voor de mededingingswet relevante) vraag of data substitueerbaar of repliceerbaar zijn.²⁷

Datadeling en prikkels om te investeren

Prikkels om te investeren in verkregen en afgeleide data kunnen afnemen door verplichtingen om dergelijke data uit te wisselen (met name als de databeheerder hier geen adequate vergoeding voor ontvangt). Dit geldt in mindere mate voor verstrekte of waargenomen data, omdat deze data in veel gevallen nevenproducten zijn van dienstverlening.

Niet alle verstrekte of waargenomen data zijn echter slechts nevenproducten. Data over de interacties van gebruikers met een interface (bijvoorbeeld klikgedrag) zijn afhankelijk van wat aan die gebruikers is getoond en op welke manier. Met andere woorden, ze worden bepaald door het algoritme en door het ontwerp van de gebruikersinterface. Het verplicht stellen van de uitwisseling van dergelijke data kan investeringsprikkels ondermijnen. Dit lichten wij toe met behulp van onderstaande figuren.

In Figuur 3.1 wordt een visualisatie gepresenteerd van het gebruik en de productie van data door platform A. Datastukjes 1, 2 en 3, aangegeven aan de linkerkant van de figuur, zijn verstrekt of

²⁶ Crémer et al (2019) lijken dit te erkennen, aangezien zij stellen dat het onwaarschijnlijk is dat een verplichting tot uitwisseling van afgeleide data tot stand komt, en dat verzoeken om toegang naar verwachting doorgaans betrekking hebben op waargenomen data, die veelal niet kunnen worden gerepliceerd, en op vrijwillig aangeboden (d.w.z. verstrekte) data, waarbij het veel inspanning zou vergen om ze opnieuw vrijwillig aan te bieden. Ook de Groep Databescherming Artikel 29 merkt op bij de interpretatie van de reikwijdte van de AVG dat, hoewel verstrekte en waargenomen data overdraagbaar moeten worden gemaakt door verwerkingsverantwoordelijken naar aanleiding van verzoeken om dataportabiliteit van betrokkenen, dit niet het geval is voor verkregen en afgeleide data volgens de Groep Databescherming Artikel 29 (Groep Databescherming Artikel 29, 2017). Dit komt doordat de Groep Databescherming Artikel 29 verkregen en afgeleide data niet beschouwt als persoonsdata 'verstrekt door de betrokkene', volgens de terminologie in Artikel 20 van de AVG.

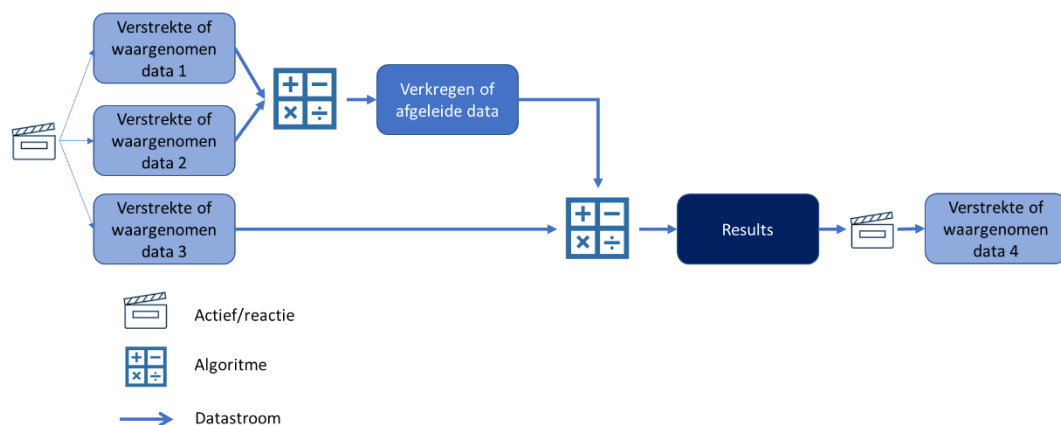
²⁷ Uit de waarneming dat iemand voornamelijk rode kleding koopt kan bijvoorbeeld worden afgeleid dat de favoriete kleur van die persoon rood is, maar die informatie kan ook door de persoon in kwestie zijn verstrekt in antwoord op een vraag die werd gesteld bij het aanmelden voor een dienst.

waargenomen als gevolg van een actie van een platformgebruiker (bijvoorbeeld het invullen van een enquête bij aanmelding). Vervolgens worden deze data ingevoerd in algoritmen die uiteindelijk leiden tot een output van de dienst (bijvoorbeeld een rangschikking). Datastukjes 1 en 2 worden eerst verwerkt door een algoritme om informatie te verkrijgen of af te leiden, die vervolgens wordt ingevoerd in het volgende algoritme en daar wordt gecombineerd met datastukje 3. De resultaten worden getoond aan de gebruiker, waarna het platform de gebruikersreactie registreert. Dit leidt tot verstrekt/waargenomen datastukje 4 (aan de rechterkant van de figuur).

Wil een concurrerend platform even relevante resultaten opleveren als platform A, dan moet kunnen beschikken over de verstrekte en waargenomen datastukjes 1, 2 en 3, zoals aangegeven aan de linkerkant van de figuur (of substituten voor die data). Vervolgens moet de concurrent innovatieve activiteiten ondernemen om de verkregen data te produceren, en vervolgens de gewenste output van de dienst (bijvoorbeeld een rangschikking). Zelf als een concurrerend platform toegang heeft tot dezelfde waargenomen/verstrekte data, is het hoogst onwaarschijnlijk dat de resultaten van beide platforms hetzelfde zijn, omdat ze verschillende algoritmen gebruiken. Eindgebruikers zullen voor het platform met de beste resultaten kiezen. Dit verschaft beide platforms een prikkel om hun algoritmen te verbeteren.

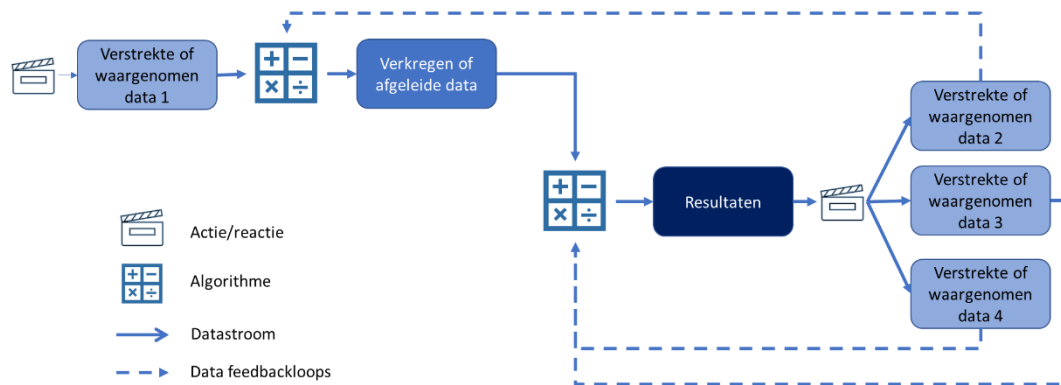
Als de platforms echter (ook) toegang hebben tot elkaars waargenomen datastukje 4 aan de rechterkant van de figuur, is het mogelijk dat ze elkaars resultaten onderwerpen aan reverse engineering. Wanneer het gemakkelijker is om de kwaliteit van een toonaangevend platform te evenaren door middel van reverse engineering gebaseerd op datastukje 4 (vergeleken met pogingen om dit te bewerkstelligen met datastukjes 1, 2 en 3), kunnen concurrenten in de verleiding komen om dit te doen en daarmee elkaars drijfveren ondermijnen om de algoritmen te verbeteren.

Figuur 3.1 Visualisatie van gebruik en productie van data (zonder feedbackloops / leereffecten)



Reverse engineering wordt ingewikkelder wanneer de waargenomen data aan de rechterkant van de figuur opnieuw in het algoritme worden ingevoerd waardoor leereffecten optreden. Zie Figuur 3.2, hoe meer een platform leereffecten (of datafeedbackloops) activeert, hoe moeilijker het is voor concurrenten om resultaten te onderwerpen aan reverse engineering.

Figuur 3.2 Gestileerde visualisatie van gebruik en productie van data (met feedbackloops / leereffecten)



Datafeedbackloops sluiten reverse engineering echter niet uit. Dit blijkt uit de casus inzake zoekmachines (hoofdstuk 5, sectie 5.2), die de mogelijkheid tot reverse engineering van een dienst met sterke leereffecten illustreert. De casus bespreekt een test waarmee Google zijn vermoedens bevestigde dat Microsoft zoekresultaten had reverse-engineered op basis van gegevens die het via Internet Explorer had opgehaald. Dit bewijst overigens niet nog geen bewuste intentie van Microsoft, want de uitkomsten van de test kunnen ook een inherent gevolg zijn van het leerproces waarop zoekalgoritmen zijn gebaseerd (zie hoofdstuk 5 voor een meer gedetailleerde bespreking).

3.1.2 Persoonsdata en niet-persoonsgebonden data

Binnen alle voorafgaande categorieën kan een onderscheid worden gemaakt tussen **persoonsgegevens** en **niet-persoonsgebonden** data, waarbij laatstgenoemde omschrijving data betreft die niet kunnen worden gekoppeld aan of herleid naar een individuele persoon. Data kunnen naar individuele personen worden herleid als de data in kwestie *identifiers* bevatten of daarmee zijn gelabeld.²⁸ Er is een grote verscheidenheid aan mogelijke identifiers, zoals een gebruikersnaam of een unieke code gegenereerd door de dienstverlener (bijvoorbeeld Gebruiker X123); adressen kunnen ook dienen als identifier (IP-adres, e-mailadres en bezoekadres), evenals creditcardnummers, telefoonnummers, computer- of browservingerafdrukken, cookies enzovoort.²⁹ Data kunnen worden **geanonimiseerd** (waarmee ze worden omgezet in niet-persoonsgebonden data) door identifiers ervan te verwijderen.

Op identificeerbaarheid gebaseerde taxonomie

Er zijn verschillende gradaties van anonimiseren en de mogelijkheden om dergelijke inspanningen ongedaan te maken. OESO (2019) bevat een datatypologie die ISO/IEC 19441 wordt genoemd. Dit is een standaard die is ontwikkeld om data-interoperabiliteit en -portabiliteit in cloudcomputingdiensten te realiseren. Deze standaard onderscheidt vijf categorieën van identificeerbaarheid van data. ISO/IEC 19441 is met name relevant voor datadeling, omdat deze typologie classificeert naarmate data kunnen worden gekoppeld aan een identiteit (d.w.z. een individuele persoon of een organisatie).

Op identificeerbaarheid gebaseerde datatonomie (OESO, 2019, p. 26-27)

- **Geïdentificeerde data** kunnen ondubbelzinnig in verband worden gebracht met een individuele persoon, omdat er persoonlijk identificeerbare informatie kan worden waargenomen.

²⁸ <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-are-identifiers-and-related-factors/>

²⁹ Zie (Senellart, 2016)

- **Gepseudonimiseerde data** zijn data waarbij alle identifiers zijn vervangen door aliassen die niet kunnen worden teruggedraaid door middel van redelijke inspanningen van derden.
- **Niet-gekoppelde gepseudonimiseerde data** zijn data waarbij alle identifiers zijn gewist of vervangen door aliassen die geen verband houden met de individuele personen, terwijl derden dit verband niet door middel van redelijke inspanningen kunnen herstellen.
- **Geanonimiseerde data** zijn niet gekoppeld, terwijl de kenmerken dusdanig zijn gewijzigd (bijvoorbeeld via randomisatie) dat er redelijkerwijs op kan worden vertrouwd dat individuele personen niet kunnen worden geïdentificeerd door de afzonderlijke data of in combinatie met andere data.
- **Geaggregeerde data** zijn statistische data zonder vermeldingen op individueel niveau, die gecombineerde informatie bevatten van een dermate groot aantal individuele personen dat persoonlijke kenmerken niet identificeerbaar zijn.

Het conceptuele verschil tussen persoonsdata en niet-persoonsgebonden data kan ook gelden voor informatie die naar bedrijven kan worden herleid. Zo behandelt Join Data, een platform voor uitwisseling van landbouwdata, data over boerenbedrijven als persoonsdata krachtens de AVG: als de data herleidbaar zijn naar een individuele boerderij, dan moet de boer toestemming geven alvorens derden toegang kunnen krijgen tot de data. Als de data niet herleidbaar zijn naar een individuele boerderij, dan kunnen ze vrijelijk worden gebruikt door ontwikkelaars.

Praktische implicaties

Vanuit juridisch oogpunt is een belangrijk onderscheid tussen persoonsdata en niet-persoonsgebonden data het feit dat persoonsdata vallen onder de databeschermingswetgeving, met name de AVG, die plichten oplegt aan verwerkingsverantwoordelijken en rechten verleent aan betrokkenen met betrekking tot de verwerking van persoonsdata. De ISO/IEC 19441 typologie kan helpen om privacyrisico's te beoordelen en de gewenste of vereiste mate van juridische en technische bescherming ondersteunen die nodig is in datadelingssituaties.

Anonimisering kan ook privacy- en beveiligingsrisico's die samenhangen met datadelingsbeleid verminderen. Hoe meer data echter worden geanonimiseerd, hoe minder bruikbaar ze worden³⁰ en hoe minder effectief dergelijk beleid wordt.³¹ Adar (2007) biedt oplossingen om privacy en onderzoeksbehoeften in evenwicht te brengen. Na verloop van tijd wordt dit evenwicht echter moeilijk te handhaven, aangezien de mogelijkheden voor heridentificatie lijken toe te nemen (Groep Databescherming Artikel 29, 2014). Rocher et al. (2019) tonen aan dat, zelfs met een minimum aan informatie over persoonlijke kenmerken, heridentificatie mogelijk is.

3.1.3 Op inhoud gebaseerde typologie

Hieronder beschrijven we verschillende datacategorieën, op basis van het soort informatie dat de data bevatten. In de literatuur bestaat geen consensus over een op inhoud gebaseerde datataxonomie. Een reden hiervoor kan zijn dat veel bedrijven hun eigen taxonomieën ontwikkelen, aangepast aan hun behoeften³². Diverse beleidsdocumenten, presentaties en artikelen gaan in op datacategorieën die zijn gebaseerd op inhoud, maar niet altijd op een manier die tegemoetkomt aan onze behoeften³³. Op basis van deze documenten en onze eigen inzichten, stellen we een op inhoud gebaseerde datataxonomie voor, waarbij onderscheid wordt gemaakt tussen **sociale data**,

³⁰ <https://medium.com/swlh/big-data-is-hard-to-anonymize-here-is-what-it-means-for-development-39e309f3b136>

³¹ Het volgende punt toont aan dat de mate waarin data waarde verliezen na anonimisering varieert voor verschillende soorten data. Het interpreteren van data over zoek- en klikgedrag kan bijvoorbeeld gemakkelijker zijn met contextuele persoonsdata zoals locatie, het gebruikte apparaat of iemands zoekgeschiedenis, maar de data kunnen zonder dergelijke contextuele persoonsdata nog steeds zeer relevante informatie bevatten. Geanonimiseerde social graph data zijn echter praktisch nutteloos.

³² Zie <https://www.talend.com/resources/gdpr-compliance-create-data-taxonomy/> en <https://digitalguardian.com/blog/how-should-you-classify-your-data-guide-using-context-content-and-user-based-data-classification>

³³ (ISO/IEC, 2017), (Senellart, 2016), (Van Til, Van Gorp, & Price, 2017)

gedragsdata, door gebruikers **zelf-gegenereerde** inhoud, **contextuele** data, **profieldata**, en **trainingsdata**. De grenzen tussen deze categorieën zijn niet altijd scherp te trekken: sommige data kunnen in meerdere categorieën voorkomen en sommige data zijn verkregen uit data in een andere categorie. Hoewel deze categorieën zijn ontwikkeld met persoonsdata in het achterhoofd, kunnen ze ook op bedrijven worden toegepast.

Sociale (netwerk)data

Sociale data (of sociale netwerkdata) bevatten informatie die **nodig is om sociale interacties** tussen mensen **te faciliteren**, alsmede de inhoud van dergelijke interacties. We maken onderscheid tussen

- **contactinformatie**, die wordt gebruikt voor het identificeren van en contact leggen met een individueel persoon³⁴,
- **social graph data**, die beschrijven wie met wie is verbonden³⁵, en
- **communicatiedata**, die beschrijven wie met wie communiceert en eventueel de inhoud en de aard van de communicatie.

In een zakelijke context zou het equivalent van sociale graph data bestaan uit CRM en SRM data (*Customer and Supplier Relationship Management data*). Het equivalent van communicatiedata kan ruimer worden geïnterpreteerd en omvat ook transactiedata (beschrijving van wie wat koopt van wie, hoe vaak en voor welk bedrag).

Sociale data of *social network data* worden grotendeels verstrekt of waargenomen. Door social graph data echter te combineren met communicatiedata, kan een platform informatie verkrijgen of afleiden over sterkte of rangschikking van persoonlijke contacten. Opgemerkt wordt dat sociale data grotendeels hun waarde verliezen als ze worden geanonimiseerd.

Gedragsdata

Veel onlinediensten verzamelen gedragsdata die worden **gegenereerd door klantbetrokkenheid met een gebruikersinterface**, aangezien deze data inzichten kunnen opleveren in de voorkeuren van gebruikers, alsook mogelijkheden om de kwaliteit van de dienstverlening te verbeteren.

Gedragsdata kunnen de volgende data bevatten:

- **transactiedata**, die beschrijven wat iemand heeft gekocht en voor welk bedrag;
- **zoekdata**, die beschrijven welke zoekopdrachten iemand heeft ingevoerd;
- **browserdata** of **klikgedrag**, waarbij wordt beschreven welke websites iemand heeft bezocht, het traject naar de website (aantal doorkliks en pogo-sticking³⁶), en bezoekduur (verblijftijd);
- **verbruiksdata inhoud**, die de inhoud beschrijven (bijv. muziek, films) waarvoor iemand heeft gekozen;
- enzovoort, afhankelijk van de functionaliteiten van de dienst en de gebruikersinterface.

Gedragsdata kunnen worden verstrekt of waargenomen, afhankelijk van de vraag of het verzamelen ervan bewust of onopgemerkt plaatsvindt. Het is eenvoudig om gedragsdata te verzamelen, omdat die data voortkomen uit interacties tussen de algoritmen/interface van een platform en gebruikers (zoals uiteengezet in sectie 3.1.1). Dit betekent echter dat gedragsdata informatie bevatten over de algoritmen en hun output, die aan het licht kan komen bij het analyseren van deze data. Het interpreteren van gedragsdata is effectiever in combinatie met contextuele persoonsdata, maar zonder dergelijke persoonlijke details bevatten ze nog steeds waardevolle informatie.

³⁴ Inclusief de naam van de eigenaar, het adres, telefoonnummer, e-mailadres, of de identiteit van de gebruiker op elektronische netwerkdiensten.

³⁵ Via vriendschap, familiebanden, vrienden op een sociaal netwerk, volgers op een sociaal netwerk enzovoort.

³⁶ Pogo-sticking verwijst naar de mate waarin gebruikers heen en weer blijven gaan tot ze vinden wat ze zoeken.

Door gebruikers gegenereerde inhoud en operationele data

Door gebruikers gegenereerde inhoud verwijst naar tekst, geluid, software, video's of beeldbestanden die gebruikers (bewust) aanmaken met behulp van applicaties of clouddiensten (ISO/IEC, 2017). Deze data zijn meestal het resultaat van een creatief proces van de gebruiker en ze hebben zowel intrinsieke (artistieke) waarde als persoonlijke (gevoels)waarde.

Het equivalent van door gebruikers gegenereerde inhoud in een B2B-context zijn operationele data, waaronder logistieke prestatie-indicatoren, input-/outputtabellen en andere data die informatie verschaffen over de efficiëntie van de activiteiten van een bedrijf. Het verschil met door gebruikers gegenereerde persoonsdata is dat operationele data doorgaans niet het resultaat zijn van een creatief proces van de gebruiker, maar meestal machine-to-machine data betreft die zijn verzameld via sensoren. Het "creatieve" proces vindt plaats in algoritmen die een kas, een boerderij, een centrum voor logistiek of een ziekenhuis aansturen.

Net als bij gedragsdata is het interpreteren van door gebruikers gegenereerde inhoud en operationele data effectiever in combinatie met contextuele persoons-/ bedrijfsspecifieke data. Zonder dergelijke details zijn deze data echter nog steeds waardevol.

Contextuele data

Contextuele data bieden **context (omstandigheden) aan een gebeurtenis, een persoon, een bedrijf of een product**. Deze data zijn van belang voor een breder begrip van specifieke stukjes informatie en het plaatsen daarvan in een groter geheel. In theorie kunnen alle soorten data contextuele informatie verschaffen. Welke data worden gebruikt om de context te beschrijven hangt af van de vraag die moet worden beantwoord. Zo kunnen **beschrijvende data** (bijvoorbeeld beschrijvingen van de inhoud van een webpagina of productkenmerken) belangrijke contextuele informatie zijn met het oog op advertentie- of opslagdiensten. Zowel **locatiedata** als **zoekgeschiedenis** kunnen belangrijke contextuele informatie zijn voor zoekdiensten (UK CMA, 2019). **Marktdata** over verhandelde volumes en prijzen kunnen van belang zijn voor het leveren van B2B- en B2C-bemiddelingsdiensten. **Kaartdata** met informatie over wegen, bergen, winkelloccaties enzovoort zijn relevant voor navigatie en logistieke diensten, en voor andere op locatie gebaseerde zoekdiensten. Andere soorten contextuele data kunnen **weerinformatie** en **tijdsaanwijzingen** zijn, of zelfs de op handen zijnde verjaardag van een familielid. Voor andere doeleinden kunnen **financiële data** (die iemands inkomen, uitgaven en spaargeld beschrijven) of **gezondheidsdata** (die de fysieke en mentale gezondheid van personen beschrijven) belangrijke contextuele data zijn – bijvoorbeeld bij het aanvragen van een hypotheek. We wijzen erop dat we hier niet volledig kunnen zijn.

Profieldata

Een profiel bevat data die beschrijven **wat belangrijk of interessant is aan een persoon op een bepaald moment en een bepaalde locatie, of in een bepaalde context**^{37 38}. Uit welke data het profiel bestaat verschilt van dienst tot dienst en van adverteerder tot adverteerder. Daarom is het onmogelijk om volledig te zijn. Een aandachtspunt is dat profieldata (per definitie) persoonlijk zijn

³⁷ "Een profiel is een soort achtergrondverhaal, dat zich meestal richt op een persoon en wat er op dit moment belangrijk of interessant is aan die persoon" (Deahl, 2019). Een profiel kan heel uitgebreid zijn en zelfs als boek worden gepubliceerd (met name in het geval van beroemdheden). Een profiel kan kort zijn, als inleiding tot iemands CV. Een profiel kan ook in de vorm van sleutelwoorden voorkomen in een databank (bijvoorbeeld ten behoeve van gerichte onlineadvertenties).

³⁸ Het proces van het beschrijven van iemands profiel, of profilering, wordt wettelijk gedefinieerd in Artikel 4, lid 4 van de AVG, waarin profilering wordt gedefinieerd als "elke vorm van geautomatiseerde verwerking van persoonsdata waarbij aan de hand van persoonsdata bepaalde persoonlijke aspecten van een natuurlijke persoon worden geëvalueerd, met name met de bedoeling zijn beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsingen te analyseren of te voorspellen."

en het grootste deel van hun waarde, zo niet al hun waarde, verliezen als ze worden geanonimiseerd.

Over het algemeen bevatten profieldata **voorkeuren** (met betrekking tot inhoud, personen, diensten, organisaties of objecten). Deze informatie kan actief worden *verstrek*t door gebruikers wanneer ernaar wordt gevraagd, maar wordt vaak *afgeleid* uit gedrags- of operationele data. Profielen kunnen ook **demografische informatie** bevatten over iemands geslacht, leeftijd, nationaliteit, familiebanden enzovoort; **financiële data** die iemands inkomen, uitgaven en spaargeld beschrijven; of **gezondheidsdata** die de fysieke en mentale gezondheid van personen beschrijven. Veel van deze data zijn verstrekt of waargenomen, maar ze kunnen ook afgeleid zijn³⁹. Afhankelijk van de gebruikssituatie kan een profiel ook **sociale data**⁴⁰ bevatten.

Trainingsdata

Alle voorgaande soorten data kunnen worden gebruikt als input voor trainingsalgoritmen. Uit de voorgaande discussie blijkt dat deze soorten data direct kunnen worden gebruikt voor training, maar dat ze in sommige gevallen handmatig moeten worden verwerkt. Een voorbeeld is het handmatig toevoegen van labels aan verstrekte of waargenomen data, zoals van het internet gescrapete foto's⁴¹. Het verzamelen (of produceren) van dergelijke *geproduceerde* trainingsdata vergt specifieke uitgaven en inspanningen, zoals het bepalen van de aard van de gewenste data en het maken van een opzet om ze te genereren.

3.2 De waarde van data en drijfveren om (niet) uit te wisselen

Hieronder worden de op inhoud gebaseerde datacategorieën nader toegelicht en wordt per soort data aangegeven waarin de waarde van de data is gelegen (d.w.z. voor wat voor soort analyses ze kunnen worden gebruikt) en waarom platforms wel of niet bereid zouden zijn deze data uit te wisselen. Waar relevant wordt algemene discussie gevolgd door specifieke opmerkingen voor zakelijke gebruikers van platforms.

3.2.1 Sociale data jagen netwerkeffecten aan

Algemeen

Social graph data zijn gerelateerd aan netwerkeffecten. De essentie van een platform zit hem in de functionaliteit om contacten te leggen en te communiceren met andere gebruikers. Daarom helpen sociale data over de ene gebruiker direct om de diensten voor anderen te verbeteren. Toegang krijgen tot iemands sociale data op een telefoon of op een sociaal netwerk kan van waarde zijn voor de groeistrategie van een platform.

Apps kunnen nieuwe gebruikers vragen om toegang tot lijsten met contactpersonen in telefoons of e-mailprogramma's. Het verlenen van toegang tot iemands sociale persoonsdata is echter geen kleinigheid, omdat deze data doorgaans ook betrekking hebben op de contacten van de individuele persoon. Anonimisering lost dit probleem niet op, omdat het leidt tot verlies van de waarde van de data. In de praktijk gebeurt het echter wel dat individuele personen aan derden toegang verlenen tot hun social graph data (bijvoorbeeld opgeslagen in apps en op apparaten), en zelfs tot

³⁹ Als een vrouw bijvoorbeeld een profielfoto van zichzelf met een klein kind naar Facebook uploadt en nog veel meer foto's van datzelfde kind op haar Facebookpagina plaatst, dan zou Facebook daaruit kunnen afleiden dat deze persoon een moeder is. Evenzo kunnen gezondheidsdata worden afgeleid uit transactiedata (bijvoorbeeld koopt iemand regelmatig sigaretten?) of locatiedata (bijvoorbeeld maakt iemand regelmatig een fietstocht?).

⁴⁰ Het opnemen van sociale data in profielen kan bijvoorbeeld interessant zijn voor bedrijven die zich bij wijze van strategie richten op individuele personen met veel 'volgers' (zogenaamde social media influencers).

⁴¹ *The Economist*, "For AI, data are harder to come by than you think", 11 juni 2020.

communicatiedata (een Gmailgebruiker kan bijvoorbeeld derden toestemming geven om e-mails te lezen)⁴².

Door apps toegang te verlenen tot sociale data, worden concurrentie en innovatie binnen de economie als geheel ondersteund: startende bedrijven kunnen gemakkelijker netwerkeffecten initiëren en eindgebruikers staan minder snel bloot aan lock-in. Om dezelfde redenen kan een platform zich niet bereid tonen tot horizontale uitwisseling van sociale data. Platforms zijn echter wel gemotiveerd om sociale data verticaal uit te wisselen met app-ontwikkelaars binnen hun ecosystemen. Net als in de economie als geheel, bevordert datadeling binnen ecosystemen concurrentie en innovatie binnen die ecosystemen. Datadeling optimaliseert de waarde die op het platform gecreëerd wordt, en daarmee het inkomstenpotentieel van het platform. Het verschil tussen verstrekt/waargenomen en verkregen/afgeleid heeft in deze geen invloed op de motivatie om verticaal data uit te wisselen. Als het proces van waardecreatie bijvoorbeeld baat heeft bij de uitwisseling van afgeleide informatie over de sterkte van sociale verbanden in iemands social graph, dan is het waarschijnlijk dat het platform deze data uitwisselt.

De standaard gebruiksvoorwaarden van een platform fungeren vaak als overkoepelende overeenkomst, waaronder data worden uitgewisseld met een groot aantal andere dienstverleners.

Zakelijke gebruikers

Voor platforms die bemiddelen in een B2B-context, dienen sociale (netwerk)data hetzelfde doel: het faciliteren van interacties in de vorm van communicatie en transacties. Zakelijke gebruikers kunnen soortgelijke motieven hebben om hun sociaalnetwerkdata als consumenten mee te willen nemen naar andere platforms: het faciliteren van overstappen en multi-homing. Bovendien hebben B2B-bemiddelingsplatforms dezelfde motieven als sociale-mediaplatformen om deze data uit te wisselen met applicaties die hun ecosystemen verrijken, en om ze niet uit te wisselen met concurrenten die vergelijkbare functionaliteiten aanbieden.

3.2.2 Gedragsdata jagen leereffecten aan

Gedragsdata worden doorgaans gebruikt om de gebruikerservaring te verbeteren en om gebruikers aan te sporen (*nudging*) tot vervolgacties, zoals een advertentie bekijken of een aankoop doen. De aansporing bestaat hier uit het zodanig ontwerpen van een digitale omgeving, dat mensen subtiel worden gestimuleerd om zich op een bepaalde manier te gedragen. Rangschikking is een eenvoudig voorbeeld, aangezien hoger gepositioneerde onderwerpen gemakkelijker worden opgemerkt. Rangschikking wordt toegepast door zoekmachines,⁴³ sociale netwerken⁴⁴ en e-commercesites.⁴⁵

Zoals toegelicht in sectie 3.1.1, zijn bepaalde leereffecten gerelateerd aan analyse van gedragsdata: door de interacties met een gebruikersinterface te analyseren, kan een platform de interface en onderliggende algoritmen verbeteren. Om dergelijke experimenten uit te voeren zijn mogelijk data nodig die de context van het waargenomen gedrag beschrijven (zie 'Contextuele '

⁴² <https://support.google.com/accounts/answer/3466521?hl=nl>

⁴³ Een zoekmachine kan verschillende gedragsdata gebruiken (zoals zoekopdrachten en klikgedrag) om de positionering van zoekresultaten zoals gemeten door andere gedragsdata (doorklikfrequentie) te verbeteren.

⁴⁴ Zoals de 'vriendenfeeds' op Facebook, die door Facebook worden gebaseerd op de positionering van iemands vrienden, uitgaande van communicatiedata.

⁴⁵ Buy Box van Amazon kan worden beschouwd als een positioneringsmechanisme, maar een nogal extreme vorm daarvan, aangezien het gebruikers 'verleidt' om geen andere leveranciers te zoeken dan de leverancier getoond in de Buy Box. Het systeem is erop gericht om winkeliers aan te moedigen beter te presteren om zo de Buy Box te winnen. Behalve zoekopdrachten en prijzen, gebruikt Amazon (onder andere) beoordelingsdata van consumenten en 'prestatiedata' met betrekking tot winkeliers (dit zijn in wezen gedragsdata die beschrijven hoe een winkelier zich bij eerdere logistieke processen heeft gedragen) om te bepalen welke winkelier 'de Buy Box wint'. (Retail 90, 2017)

hieronder)⁴⁶. Gedragsdata kunnen ook worden gebruikt om individuele voorkeuren af te leiden (zie 'Profiel' hieronder), waardoor een platform gerichte advertenties kan weergeven of relevante inhoud kan voorstellen.

Deze leereffecten versterken netwerkeffecten en vergroten het risico op het kantelen van een markt. Er is gesteld dat schaafeffecten bij het leren deze uitkomsten veroorzaken. Volgens sommigen kan een klein datavoordel in de loop van de tijd uitgroeien tot een leidende marktpositie, een kanteling van de markt naar een monopolie veroorzaken, en resulteren in mogelijkheden om vervolgens ook andere markten te monopoliseren⁴⁷. Anderen hebben argumenten aangevoerd die dit effect verzwakken⁴⁸, zoals het gegeven dat het leerproces wordt gekenmerkt door (snel) afnemende schaalopbrengsten en het feit dat gedragsdata vrij snel over de datum kunnen zijn, afhankelijk van de gebruikssituatie. Deze kenmerken zouden betekenen dat first-movers niet het soort voordeel behalen waardoor ze op hun lauweren kunnen gaan rusten. Op hun beurt zijn deze argumenten weerlegd, zoals wordt besproken in onderstaande tekst.

Schaafeffecten in data-analyse

Er is gesteld dat de relatie tussen data en de kwaliteit van dienstverlening wordt gekenmerkt door **afnemende schaalopbrengsten**⁴⁹, wat het voordeel van de *first-mover* zou beperken (He, et al, 2017). Dit argument heeft doorgaans betrekking op het **volume** van de dataset en niet op de **variëteit** ervan. Data hebben de eigenschap dat ze elkaar aanvullen, omdat de mogelijkheden om correlaties te onderzoeken toenemen door een grotere verscheidenheid aan data. Om die reden verbetert het opnemen van contextuele en profieldata in de analyse de interpretaties van gedragsdata. De relatie tussen het aantal verschillende soorten data en het aantal correlaties dat onderzocht kan worden volgt een vergelijkbaar exponentieel groeipatroon als de relatie tussen het aantal gebruikers en interactiemogelijkheden. Daardoor is er sprake van **toenemende variëteitsopbrengsten**. Bovendien leert de statistiek ons dat wanneer verscheidenheid toeneemt, het volume moet toenemen om te compenseren voor de afnemende vrijheidsgraden. **Toenemende opbrengsten**, niet als gevolg van verscheidenheid maar door volume in het algemeen, kunnen ook voortkomen uit leereffecten: meer data kunnen een (lichte) kwaliteitsverhoging teweegbrengen, waardoor nieuwe gebruikers worden aangetrokken, die op hun beurt meer data genereren die in het leerproces kunnen worden ingevoerd om de kwaliteit te verhogen. Zo kan een klein datavoordel na verloop van tijd een leidende marktpositie tot gevolg hebben. Dus hoewel data afnemende opbrengsten kunnen laten zien in technische zin (nauwkeurigheid), kan er sprake zijn van toenemende opbrengsten in economische zin (marktaandeel).⁵⁰

De complementaire aard van data en de mogelijkheid om data af te leiden uit andere data kunnen echter in potentie het voordeel van de first-mover beperken, omdat deze leiden tot een mate van **substituëerbaarheid** van data. Zo kan een mobiel besturingssysteem via de technische instellingen van de smartphone waarnemen dat iemand een gsm-abonnement heeft bij Vodafone. Maar een social-mediaplatform kan dezelfde informatie afleiden uit het feit dat deze persoon Vodafone volgt op dat platform en uit de waarneming dat deze persoon via dat platform communiceert met de helpdesk van Vodafone.

Een ander kenmerk van data waarvan wordt aangevoerd dat het het voordeel van de first-mover verkleint is dat, afhankelijk van de gebruikssituatie, de **relevantie van data vrij snel verloren kan gaan**. Daarom

⁴⁶ In het geval van testen op kleurgebruik in de gebruikersinterface, bevat contextuele informatie zowel andere ontwerpkenmerken van de gebruikersinterface als locatiedata waaruit kan worden afgeleid of de persoon op het moment van gebruik een auto bestuurt, waardoor hij waarschijnlijk niet goed oplet en het experiment dus minder zin heeft (of juist meer zin, afhankelijk van de geteste functionaliteit).

⁴⁷ Agrawal, Gans, & Goldfarb (2019), Prüffer en Schottmüller (2017), Mayer-Schönberg & Ramge (2018), Anderson (2020)

⁴⁸ Waaronder Hal Varian van Google en Pat Bajari van Amazon, volgens Anderson (2020)

⁴⁹ Dit is geen verrassing, aangezien algemene statistieken ons leren dat relevantie (R²) toeneemt met de steekproefomvang (n) in een afnemend tempo; op zeker moment is de omvang van de steekproef groot genoeg om duidelijke resultaten op te leveren en leidt een toename van de steekproefomvang niet tot meer relevantie.

⁵⁰ Agrawal, Gans, & Goldfarb (2019)

kan het belangrijk zijn om toegang te hebben tot realltime data, om de houders van data een "snelheidsvoordeel" te ontzeggen. Het internet wordt voortdurend geactualiseerd met nieuwe informatie, waardoor het concept 'relevantie' een bewegend doelwit wordt. Als zodanig hebben zoekalgoritmen constante updates nodig met nieuwe (real-time) data combinaties. Dit geldt met name voor zoekalgoritmen die zoekopdrachten gebruiken als input data (He, et al, 2017), maar hetzelfde geldt voor data en algoritmen die worden gebruikt door sociale netwerken, omdat relaties en de sterkte van relaties binnen een social graph voortdurend in ontwikkeling zijn. Dit vereist een hoge **snelheid** waarmee data worden geactualiseerd zodat schaal, als in het aantal gebruikers, nog belangrijker wordt: hoe meer gebruikers, hoe waarschijnlijker het is dat het benodigde aantal gerelateerde zoekopdrachten om relevantie voort te brengen (of het benodigde aantal sociale interacties in het geval van sociale netwerken) wordt gehaald voordat de houdbaarheidsdatum van de data is verstreken.

Een laatste kenmerk van data wordt gevormd door de enorme breedtevoordelen die aan data gerelateerd zijn, aangezien (dezelfde) data de input kunnen vormen voor een grote verscheidenheid aan diensten. Gedragsdata voeden bijvoorbeeld persoonsprofielen die de relevantie van zoekresultaten of winkelsuggesties helpen vergroten, ze helpen een sociaal netwerk om zinvolle interacties mogelijk te maken en ze stellen een financiële dienstverlener in staat om betere adviezen te geven. Overigens kan een input als olie even veelzijdig zijn, maar in tegenstelling tot data is olie een rivaliserende input en kan deze dus niet worden hergebruikt om verschillende diensten of producten te genereren.

Horizontale uitwisseling van gedragsdata kan een effectief instrument zijn om het risico op kanteling van de markt bij sterke leereffecten te beperken. Echter, omdat gedragsdata het resultaat zijn van interacties met de gebruikersinterface (en de algoritmen die op de achtergrond worden uitgevoerd), kan uitwisseling ervan concurrenten de mogelijkheid bieden voor reverse engineering (afhankelijk van het aantal en het belang van data feedback loops). In het geval van reverse engineering verdwijnt voor alle partijen de stimulans om te investeren en innoveren. Dit risico is een legitiem argument tegen uitwisseling van gedragsdata. Een besluit om uitwisseling van gedragsdata verplicht te stellen moet daarom de winst voor concurrentie en de kosten die reverse engineering met zich meebrengt balanceren. Een aandachtspunt hierbij is dat het risico op reverse engineering een potentieel conflict introduceert tussen verplichte datadeling en bepaalde intellectuele eigendomsrechten; met name de databankrichtlijn (wanneer uitwisseling informatie verschafft over de structuur van databanken, d.w.z. webindexen).

Weerstand tegen uitwisseling van gedragsdata kan ook samenhangen met privacyrisico's (zoals aangegeven door deskundigen in de interviews voor de casussen). Gedragsdata behouden echter behoorlijk veel waarde na anonimisering. Combinaties van zoekopdrachten en klikgedrag kunnen bijvoorbeeld worden uitgewisseld zonder identiteiten bekend te maken, terwijl ze nog steeds kunnen worden geanalyseerd om de kwaliteit van zoekalgoritmen te verbeteren. In pseudogeanonimiseerde vorm neemt de waarde natuurlijk toe, aangezien in dat geval meer profiel- en contextuele data in de analyse kunnen worden opgenomen, maar dit brengt wel verhoogde risico's voor privacy en beveiliging met zich mee.

3.2.3 *Zelf-gegenereerde inhoud en operationele data jagen lock-in-effecten aan*

Algemeen

Door gebruikers zelf-gegenereerde inhoud is veelal het resultaat van een creatief proces van de gebruiker, en heeft zowel intrinsieke (artistieke) waarde als persoonlijke (gevoels)waarde. Dit laatste zorgt ervoor dat gebruikers doorgaans aan de data gehecht zijn en de toegang ertoe niet willen kwijtraken. Dit kan lock-in-effecten veroorzaken, wat de reden is dat een platform wellicht niet snel bereid is om de data uit te wisselen of over te dragen. Behalve vanwege hun gevoels- en artistieke waarde, hebben de data ook waarde als input voor analyse. Een platform kan mogelijk

een grote verscheidenheid aan informatie (zoals sociale data, contextuele data en profieldata) verkrijgen uit deze ongestructureerde databestanden.

Zakelijke gebruikers

Operationele data bieden essentiële input voor de prestaties van een bedrijf en zijn als zodanig een primaire bron voor operationele, tactische en strategische besluitvorming. Onbeperkte toegang tot de data is van cruciaal belang voor de bedrijfsvoering. Onvolledige dataportabiliteit of interoperabiliteit tussen de oude en de nieuwe infrastructuur kan leiden tot verlies van toegang tot (sommige) data. Deze vrees heeft een lock-in van zakelijke gebruikers met de huidige softwareaanbieders en aanbieders van cloud-infrastructuren tot gevolg.

Bij operationele data vindt (als het ware) het 'creatieve proces' plaats in algoritmen die de operationele prestatie van een kas, een boerderij, een logistiek centrum of een ziekenhuis aansturen. In theorie kan uitwisseling van deze data dus ook het risico op reverse engineering inhouden (net als bij gedragsdata). In de praktijk is het aantal variabelen dat in de vergelijking wordt ingevoerd (en er weer in wordt teruggevoerd) waarschijnlijk echter veel hoger, wat het risico op reverse engineering verkleint. Hierdoor zijn bedrijven en platforms sneller bereid (of zouden ze dat moeten zijn) om operationele data uit te wisselen met anderen die ermee willen innoveren. Bezorgdheid over het vrijgeven van commerciële en strategische informatie kan worden weggenomen door de data zodanig te anonimiseren dat deze niet kunnen worden herleid naar een bepaald bedrijf.

3.2.4 Contextuele data verschaffen betekenis

Contextuele data bieden context (omstandigheden) aan een gebeurtenis, een persoon, een bedrijf of een product. Deze data zijn van belang voor een breder begrip van specifieke stukjes informatie door ze in een groter geheel te plaatsen. Wanneer contextuele informatie betrekking heeft op een bedrijf of persoon, is deze persoons-/bedrijfsspecifiek. Dergelijke contextuele informatie verliest aan waarde na anonimisering. Het combineren van geanonimiseerde contextuele informatie met andere geanonimiseerde data houdt echter het risico in dat data worden gedeanonimiseerd. Dit probleem kan worden ingedamd als contextuele data worden gebruikt op een meer geaggregeerd niveau⁵¹. In de regel geldt echter dat hoe meer contextuele data worden uitgewisseld in combinatie met geanonimiseerde data, hoe groter de kans dat geanonimiseerde data kunnen worden gedeanonimiseerd.

3.2.5 Profieldata jagen netwerkeffecten aan en vormen een bron van inkomsten

Algemeen

Profieldata worden in grote lijnen voor twee doeleinden gebruikt: om indirecte netwerkeffecten aan te jagen en om advertentieopbrengsten te genereren. Indirecte netwerkeffecten worden aangejaagd door profieldata, aangezien deze nodig zijn om apps, inhoud en productaanbiedingen af te stemmen op potentiële gebruikers. In dynamische zin kan ook worden gesteld dat profieldata kunnen worden gebruikt om de ontwikkeling van deze apps, inhoud en productaanbiedingen te faciliteren. Daarnaast worden profieldata gebruikt om advertenties af te stemmen op een doelgroep.

⁵¹ Door bijvoorbeeld geanonimiseerde zoekopdrachten te taggen met geanonimiseerde locatiedata, kan een onderzoeker hoogstwaarschijnlijk alle zoekopdrachten op een bepaald adres filteren en ze koppelen aan de persoon die daar woont. De locatiedata moeten dus worden geaggregeerd op bijvoorbeeld stadsniveau om dergelijke deanonimiserering te voorkomen.

Een algemene verplichting om profieldata uit te wisselen is in strijd met de AVG. Profielen kunnen niet worden geanonimiseerd zonder een behoorlijke hoeveelheid (zo niet alle) waarde te verliezen. Krachtens de AVG is toestemming altijd vereist. In een hypothetische situatie waarin mensen massaal toestemming zouden geven of zelfs opdracht zouden geven aan platforms om profieldata uit te wisselen en over te dragen, zou dit de concurrentie tussen aanbieders van advertenties fundamenteel veranderen. De 'rat race' om data wordt waarschijnlijk minder sterk. Wanneer alle advertentieplatforms over dezelfde profieldata beschikken vanwege uitwisselingsverplichtingen, is er niet langer een verschil in hun vermogen om zich op individuen te richten aan de hand van verschillende criteria. Diensten kunnen (moeten) zich dan meer van elkaar onderscheiden door zich in de eerste plaats op een specifieke doelgroep te richten. Strava richt zich bijvoorbeeld op atleten en Lume op vrouwen die zich bezighouden met spiritualiteit. Deze platforms onderscheiden zich als aanbieders van advertentieruimte, niet vanwege hun vermogen om zich op specifieke personen te richten, maar doordat ze adverteerders de mogelijkheid bieden om zich te verbinden aan het thema van de dienst en de bijbehorende gebruikers.

Op voorhand is het niet duidelijk of het nettoresultaat positief is voor eindgebruikers. Minder honger naar data kan welkom zijn als het gaat om privacybescherming, maar dit voordeel kan wegvallen doordat uitgewisselde profielen niet kunnen worden geanonimiseerd zonder dat ze veel (zo niet alle) waarde verliezen. Bovendien stimuleert de huidige datahonger innovatie van diensten. Zonder deze honger verandert de drang tot innovatie en wordt deze wellicht minder (dat weten we niet).

Zakelijke gebruikers

In het geval van bedrijven kunnen profieldata worden gebruikt om investeringsmogelijkheden te signaleren. Banken, investeerders, effectenbeurzen enzovoort zijn geïnteresseerd in de financiële gezondheid van bedrijven en in andere informatie over de levensvatbaarheid van bedrijfsmodellen, waaronder CRM data (zie hierboven) en operationele data (zie hieronder). In dit opzicht zijn de door platforms als Google en Facebook aangeboden sociale inlogfuncties een interessant fenomeen. Door sociale inlogfuncties aan te bieden krijgen platforms toegang tot gedetailleerde gebruiksdata van duizenden apps en websites. Dit verschaft hen met enorm veel kennis, op real-time basis, over marktaandeel en alles wat verder uit de data kan worden afgeleid. Betrokkenen beschouwen deze data mogelijk als strategisch en zijn wellicht geen voorstanders van openbare beschikbaarheid (zo publiceren niet alle apps hun gebruiksdata). Door gebruik te maken van sociale inlogfuncties worden veel van deze strategische data echter uitgewisseld met de Big Techs, waardoor deze een informatievoordeel behalen ten opzichte van andere potentiële investeerders/kopers. De sociale inlogservice maakt zo niet alleen het leven gemakkelijker voor apps en websites van derden (door hen toegang te verschaffen tot sociale en profieldata van gebruikers) en hun gebruikers (waardoor er minder rompslomp is met wachtwoorden), maar komt de service ook ten goede aan de aanbieder ervan in de vorm van waardevolle marktinformatie over winstgevend overnames (mogelijk uit preventieve overwegingen).

3.2.6 De kosten van dataproductie en de impact van verplichte datadeling

Het verzamelen of produceren van data kan inspanningen, investeringen en andere uitgaven met zich meebrengen. Deze kosten zijn bescheiden of zelfs verwaarloosbaar voor data die een bijproduct zijn van bestaande activiteiten. De kosten om andere soorten data te genereren kunnen aanzienlijk zijn. Hieronder lichten we deze kosten en hun implicaties nader toe. In het kader van deze discussie volgen we de classificatie van op oorsprong gebaseerde soorten data: verstrekte/waargenomen data, verkregen/afgeleide data, en geproduceerde data.

Waargenomen en verstrekte data hebben betrekking op het onderscheid tussen ruwe en bewerkte data, waarbij waargenomen/verstrekte data tot de eerste categorie behoren. Deze

soorten data zijn (actief of passief) aangemaakt, bekendgemaakt of uitgewisseld door gebruikers bij het gebruik van een dienst. Ook kunnen ze door sensoren als ruwe data zijn verzameld. De inspanningen en uitgaven die gepaard gaan met het verzamelen van deze data doorgaans relatief beperkt. Dataverzameling door online zoekmachines vereist bijvoorbeeld alleen dat gebruikers zoekopdrachten intypen en klikken op resultaten: het gaat in feite om de levering van de dienst zelf. Evenzo is het verzamelen van sociale data een nevenproduct van het aanbieden van socialemediadiensten. Deze data kunnen vervolgens (tegen verwaarloosbare kosten) worden uitgewisseld met adverteerders, downstream-app-ontwikkelaars en anderen.

Uitzonderingen doen zich voor wanneer verstrekte of waargenomen data moeten voldoen aan hoge kwaliteitsnormen, zoals technische precisie in het geval van cruciale applicaties, of formele legitimiteit voor boekhoudkundige doeleinden. Dergelijke gevallen doen zich doorgaans voor in de B2B-sfeer, waar het verzamelen van data kan vragen om investeringen in complexe systemen (bijvoorbeeld technische, sensor-, ICT-, of administratieve systemen). Een datadeelplicht (zonder adequate vergoeding) zou in die gevallen de stimulans kunnen aantasten om te investeren in het opzetten van de systemen om deze data te verzamelen.

Verkregen en afgeleide data zijn bewerkte data. Allereerst moet worden opgemerkt dat als waargenomen of verstrekte (ruwe) data repliceerbaar zijn, over het algemeen hetzelfde geldt voor bewerkte data die zijn verkregen of afgeleid uit de ruwe data. De reden daarvoor is dat anderen net zo goed methoden (bijv. algoritmen) kunnen bedenken die de bewerkte data genereren. Daarom is het doorgaans niet nodig of wenselijk om de uitwisseling van bewerkte data verplicht te stellen. De vraag is of de onderliggende waargenomen of verstrekte data problemen veroorzaken die uitwisselingsverplichtingen wenselijk maken, bijvoorbeeld wanneer deze data niet repliceerbaar zijn.

Verkregen en afgeleide data zijn meestal gebaseerd op berekeningen die zijn uitgevoerd door algoritmen. Om dit mogelijk te maken moet een bedrijf een systeem installeren dat daarvoor geschikt is (bijvoorbeeld databanksoftware die op bestaande computers draait) en investeren in software die analyses kan uitvoeren. Een voorbeeld zijn kritieke prestatie-indicatoren van bedrijven (bijv. ziekenhuizen). Deze zijn doorgaans gebaseerd op onderliggende data die worden verzameld in het kader van routinematige dienstverlening⁵². Soms zijn eenvoudige berekeningstechnieken voldoende om prestatie-indicatoren te verkrijgen. In andere gevallen brengt het meer kosten met zich mee om verkregen data te produceren. Neem weersvoorspellingen. Voor het registreren van de actuele weersomstandigheden zijn door sensoren en satellieten geleverde data nodig. Het grootste deel van dergelijke data kan worden gekocht. Echter, om het weer te voorspellen moeten geavanceerde algoritmen worden ontwikkeld waarvoor krachtige computers nodig zijn. De kosten voor het produceren van verkregen data kunnen dus sterk uiteenlopen.

Geproduceerde data bevatten data die zijn voortgekomen uit enquêtes, klantenproeven, A/B-testen, experimenten en handmatige bewerkingen (uitgevoerd op waargenomen/verstrekte data). Dat laatste gebeurt bij trainingsdata voor AI-applicaties. Een voorbeeld is het handmatig toevoegen van labels aan waargenomen of verstrekte data, zoals van het internet gescrapete foto's⁵³. Voor het verzamelen (of produceren) van geproduceerde data moeten specifieke kosten worden gemaakt en inspanningen worden geleverd, zoals het bepalen van de aard van de gewenste data en het maken van een opzet om deze te genereren. Als deze kosten substantieel zijn, kunnen uitwisselingsverplichtingen een verstorende werking hebben op concurrentie en innovatie.

⁵² Door Jones et al. (2015) wordt dit besproken in het geval van ziekenhuizen.

⁵³ *The Economist*, "For AI, data are harder to come by than you think", 11 juni 2020.

De kosten voor het produceren van data, en bijgevolg ook de impact van verplichtingen om deze data te delen, kunnen enorm uiteenlopen. Neem, bijvoorbeeld, data die zijn geproduceerd op basis van een uitgebreide en dure enquête: een uitwisselingsverplichting zou gemakkelijk de prikkels om zo'n enquête uit te voeren kunnen ondermijnen. Bovendien zijn er geen belemmeringen voor anderen om soortgelijke enquêtes uit te voeren. Iets dergelijks geldt ook voor trainingsdata, waarvan de (vaak handmatige) bewerking veel kosten met zich mee kan brengen. Een ander voorbeeld zijn geografische kaarten. Deze worden op zeer verschillende manieren geproduceerd met behulp van data afkomstig van uiterst specifieke bronnen, zoals landmetingen, satellietbeelden en institutionele/officialle bronnen.⁵⁴ De mate van detail en de aard van de kaarten zijn bepalend voor de kosten die het produceren ervan met zich meebrengt (ook qua gespecialiseerde expertise). Om deze reden wijden sommige bedrijven zich aan het produceren van kaarten en geven hun producten in licentie aan firma's die kaartdata gebruiken als input (bijvoorbeeld aanbieders van routeplanningsdiensten).

De kosten in kwestie kunnen zowel vast als incrementeel zijn. Het opzetten van een structuur om data te produceren brengt vooraf te maken vaste en verzonken kosten met zich mee. Kosten zijn incrementeel wanneer bijvoorbeeld voor het genereren van data menselijke arbeid nodig is, zoals bij trainingsdata, of wanneer de hoeveelheid data van belang is voor het niveau van de informatie (c.q. kunstmatige intelligentie). In beide gevallen vertalen hogere uitgaven zich direct in hogere kwaliteit van de geproduceerde data⁵⁵. Bedrijven kunnen dergelijke trainingsdata mogelijk zelf produceren of deze kopen bij dataproducenten, zoals bedrijven die trainingsdata produceren. De hoogte van dergelijke incrementele kosten kan enorm uiteenlopen.

Voor alle soorten hierboven besproken data geldt dat de kosten voor het genereren van data kunnen oplopen als gevolg van wettelijke beperkingen. Dergelijke kosten houden mogelijk niet rechtstreeks verband met de dataproductie, maar kunnen niettemin onvermijdelijk zijn. Een voorbeeld is het produceren van data door middel van observatiecamera's in openbare ruimtes of bij particuliere gebouwen. Als gevolg van de privacyregelgeving en aansprakelijkheidsrisico's met betrekking tot beveiligingsinbreuken (bijvoorbeeld onderschepping van data door indringers) worden dan (mogelijk dure) voorzorgsmaatregelen opgelegd.

De conclusie luidt dat de kosten om data te genereren sterk uiteen kunnen lopen, maar zelfs als sprake is van een dure aangelegenheid, dan nog is datadeling niet altijd schadelijk voor de prikkels om data te verzamelen. Een relevante vraag is niet alleen (of niet zozeer) of de prikkels voor het (investeren in) verzamelen van data verdwijnen door een verplichting tot datadeling. Deze prikkels zijn vaak ook afhankelijk van het bedrijfsmodel waarvoor gekozen is, en of dat afhankelijk is van exclusieve toegang tot data of niet. Er zijn ook andere bedrijfsmodellen die hier minder afhankelijk van zijn. Datadeling helpt te voorkomen dat bedrijfsmodellen worden ontwikkeld op basis van exclusieve data en (daaruit voortkomende) customer lock-in. Met andere woorden, datadeling nodigt uit tot het ontwikkelen van bedrijfsmodellen die baat hebben bij datadeling, of die niet afhankelijk zijn van de aan- of afwezigheid van een uitwisselingsverplichting.

We sluiten laatstgenoemde bevindingen af met twee voorbeelden. Het eerste voorbeeld betreft een bedrijfsmodel dat baat heeft bij datadeling: een bouwbedrijf dat sensoren installeert in nieuwe gebouwen. Via een digitaal platform kan het bedrijf de verzamelde data beschikbaar stellen aan een veelheid aan dienstverleners, zoals schoonmaak- en onderhoudsbedrijven. Het tweede voorbeeld betreft een bedrijfsmodel dat mogelijk niet afhankelijk is van (of 'immuun is voor') datadeling: Amazon Go, de in de VS gevestigde keten van buurtsupermarkten. In deze supermarkten wordt het winkelend publiek stap voor stap gevolgd door camera's en andere

⁵⁴ Nagaraj en Stern (2020).

⁵⁵ *The Economist*, "For AI, data are harder to come by than you think", 11 juni 2020.

technologieën. Zo worden data verzameld op basis waarvan een winkel consumenten virtueel kan laten afrekenen en zij automatisch betalen voor artikelen die ze hebben afgehaald. Wanneer die ruwe winkeldata gedeeld moeten worden lijkt de waardepropositie of het bedrijfsmodel, dat is gebaseerd op het gemakkelijker maken van boodschappen doen, niet te ondermijnen. Het kan natuurlijk wel de privacy van de consument aantasten, alsook de wens van het bedrijf om strategische overwegingen met betrekking tot de lay-out van de supermarkten geheim te houden (dat betekent dat er ook argumenten tegen datadeling zijn in dit voorbeeld).

3.3 Samenvattende infographic

Bovenstaande bespreking van typologieën en datacategorieën wordt samengevat in de onderstaande infographic.

De kleurenbalken in de grafiek geven aan in hoeverre er sprake is van ruwe data (d.w.z. verstrekt/waargenomen) of van de output van een analyse (d.w.z. verkregen of afgeleid). Het delen van ruwe data heeft doorgaans geen invloed op de prikkels om te investeren.

De verticale positionering van de datacategorieën geeft aan in hoeverre soorten data waarde verliezen na anonimisering. Sociale en profieldata verliezen veel waarde na anonimisering. Daarom is toestemming van betrokkenen vereist voor het delen van die data. Gedragsdata en contextuele data kunnen ook na anonimisering nog van waarde zijn, al nemen de technische mogelijkheden om data te deanonimiseren toe met de beschikbaarheid van contextuele data. Over het algemeen nemen de mogelijkheden om te deanonimiseren over de tijd toe (met name in een context van Big Data), waardoor het steeds moeilijker wordt om privacybehoeften af te wegen tegen andere beleidsdoelstellingen.

De rechter kolom van de figuur toont:

- I. De **waarde van data** voor platforms vanuit het perspectief van hun bedrijfsmodellen: aanjagen van netwerkeffecten (sociale data en profieldata), aanjagen van leereffecten (gedragsdata en operationele data), gebruikersbetrokkenheid (zelf-gegenereerde inhoud), en ter ondersteuning van data-analyse door te helpen bij het interpreteren van correlaties (contextuele data).
- II. De **repliceerbaarheid van data**; de meeste data zijn (deels) repliceerbaar of beschikbaar uit andere bronnen. Sociale data zijn deels repliceerbaar door gebruikers te vragen om toegang tot hun contactlijsten in telefoons en e-mailprogramma's. Dit geldt niet voor door gebruikers zelf gegenereerde inhoud, aangezien er in dat geval geen simpele één-klik-optie is waarmee gebruikers anderen toegang kunnen verlenen. Evenzo hoeven gedragsdata niet per se repliceerbaar te zijn. Zo kunnen zoekopdrachten en klikgedrag door zowel zoekmachines als webbrowsers worden waargenomen. Tien jaar geleden betekende dit dat de data repliceerbaar waren door het aanbieden van webbrowsers. Het feit dat Google momenteel zowel de zoek- als de browsermarkt domineert, heeft de repliceerbaarheid van de data echter beperkt.
- III. De relatie tot **mobilititeit van gebruikers**; de kolom toont in hoeverre datamobilititeit van invloed is op gebruikersmobilititeit (of, omgekeerd, customer lock-in). Dit is doorgaans van toepassing op data die emotionele of strategische waarde hebben voor gebruikers (zoals door gebruikers zelf gegenereerde inhoud, inclusief operationele data).
- IV. **Prikkels om data te delen**; bij gebrek aan regels of regelgeving hebben platforms een prikkel om data uit te wisselen met verticale partners binnen hun ecosysteem (tenzij dit bedrijfsgeheimen blootlegt, wat het geval kan zijn met gedragsdata), maar niet met horizontale concurrenten of belanghebbenden in andere sectoren die buiten hun ecosysteem opereren. De grens tussen een verticale partner en een horizontale concurrent is niet altijd scherp te trekken. Sommige platforms (bijvoorbeeld Facebook) hanteren een brede definitie van wat een

partnerschap inhoudt en maken het mogelijk voor verticale partners om hun functionaliteiten te dupliceren. Andere platforms (bijvoorbeeld Strava) passen een strikte definitie toe en sluiten de datatoevoer af zodra een partner hun functionaliteiten dupliceert.

- V. **Gevolgen van datadeling**; het breed delen van sociale, gedrags- en profieldata is van invloed op de concurrentie, doordat de risico's op kanteling van de markt en de datahonger erdoor afnemen. De effectiviteit van uitwisseling van sociale en profieldata kan echter beperkt zijn doordat de data praktisch nutteloos zijn na anonimisering, en dus toestemming vereist is voor effectief datadelen. Bovendien brengt de het delen van gedragsdata risico's op reverse engineering met zich mee en kan het opleggen van een datadeelplicht in strijd kan zijn met bepaalde intellectuele eigendomsrechten. Het delen van contextuele data samen met geanonimiseerde data gaat gepaard met het risico op deanonimisering van de data.

In geanonimiseerde vorm, verliezen deze data ...

... meer waarde

... minder waarde

		I Waarde	II repliceerbaar?	III Gerelateerd aan mobility?	IV Vrijwillig gedeeld?	V Gevolgen van delen?
Sociale data	Verstrek of waargenomen					
	Verkregen of afgeleid					
	Contact informatie					
Profieldata	Social Graph (inclusief kracht van connecties)	Directe NWE	✓	✓	Enkel verticaal	<ul style="list-style-type: none"> Toestemming vereist Beperkt het risico op kantelen Privacy risico's
	Voorkeuren		✓*			
	gezondheidsdata	Indirecte NWE	✓	✗		
	financiële data	inkomsten	✓	✗		
	demografische data		✓			
gedragsdata	Transactie data				Enkel verticaal	<ul style="list-style-type: none"> Beperkt het risico op kantelen Risico op reverse engineering In strijd met IE-wetgeving
	zoekopdrachten	leereffecten	?	✗		
	Surf- en klikgedrag					
	Content consumptie					
zelf- gegenereer de inhoud	Foto's, Video's, Audio, tekstdocumenten	betrokkenheid	✓	✓	Enkel verticaal	<ul style="list-style-type: none"> Industrie-inspanningen voor standaardisatie nodig
	Operationele data	leereffecten	✓	✓		
Contextuele data	Sociale data, profieldata, en gedragsdata				Enkel verticaal	<ul style="list-style-type: none"> Privacyrisico's aangezien deze gegevens helpen om gegevens te de-anonimiseren
	Kaartendata, weerdata	interpretatie	✓*	✗		
	marktdata, beschrijvende (meta) data					
	Tijdsdata					

* Afgeleide gegevens zijn even repliceerbaar als de verstrekte en waargenomen gegevens waarop ze zijn gebaseerd

4 Wetgeving inzake datadeling

Bestaande wettelijke rechten en plichten inzake datadeling bevatten, naast algemene en horizontale regels die van toepassing zijn op de economie als geheel, ook sectorspecifieke regels. In dit hoofdstuk worden drie algemeen geldende regimes voor datadeling in de EU behandeld: intellectueel eigendomsrecht, databescherming en mededingingsrecht. Aangezien de regels in deze regimes voortkomen uit EU-kaders, ligt de nadruk op EU-wetgeving. Indien relevant, zijn voorbeelden uit of verwijzingen naar de Nederlandse situatie opgenomen. De analyse toont dat de regimes voor databescherming en mededingingsrecht datadeling mogelijk kunnen maken, maar ook beperkingen kunnen opleggen. Bovendien worden datasets soms beschermd door intellectuele eigendomsrechten en bedrijfsgeheimen. De invoering van wettelijke rechten of plichten inzake datadeling kan daardoor gevolgen hebben voor de desbetreffende bescherming van intellectuele eigendomsrechten of bedrijfsgeheimen. We onderzoeken ook hoe databescherming en het mededingingsrecht zowel mogelijkheden bieden als beperkingen opleggen aan datadeling. We zullen inzichten uit dit hoofdstuk toepassen op het onderzoek naar specifieke casussen (hoofdstuk 5).

4.1 Intellectueel eigendomsrecht en bedrijfsgeheimen

In de EU bestaat er geen specifiek eigendomsrecht voor data als zodanig,⁵⁶ maar kunnen datasets vallen onder de bescherming van bestaande intellectuele eigendomsrechten en bedrijfsgeheimen. In dit gedeelte worden het auteursrecht, het "sui generis" databankenrecht, en bedrijfsgeheimen besproken vanuit het perspectief van de bescherming van data(sets).

Bescherming van oorspronkelijke inhoud onder het auteursrecht

Onder het auteursrecht wordt de expressie van een oorspronkelijk werk beschermd en worden aan auteurs tijdelijke exclusieve rechten verleend, waaronder het reproductierecht en het recht op mededeling aan het publiek.⁵⁷ Hoewel ideeën, feiten of data op zich niet aanmerking komen voor auteursrechtelijke bescherming, kan het auteursrecht wel dienen als bescherming van databanken die zijn aangemaakt door providers op basis van informatie die zij hebben verzameld, en ook van oorspronkelijke inhoud geleverd door gebruikers. De bescherming van oorspronkelijke gebruikersinhoud en de relevantie daarvan voor datadeling worden besproken in deze sub-sectie. De bescherming van databanken komt aan de orde in de volgende sub-sectie.

De bescherming van oorspronkelijke inhoud die door gebruikers is geüpload kan bijvoorbeeld betrekking hebben op oorspronkelijke berichten, foto's of video's geüpload op een sociaal netwerk. Hoewel gebruikers auteursrecht kunnen hebben op de inhoud die ze uploaden, moeten ze onder

⁵⁶ Enkele jaren geleden vond echter een discussie plaats over het creëren van een eigendomsrecht voor niet-persoonsgebonden data (het zogenoemde dataproductentenrecht). Er lijkt nu consensus over te bestaan dat een dergelijk nieuw exclusief recht niet geschikt is voor het beoogde doel, en het debat lijkt te zijn verschoven naar het creëren van mogelijkheden voor toegang tot data in plaats van eigendom van data. Zie H. Zech, 'A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data' (2016) 11 *Journal of Intellectual Property Law & Practice* 460; J. Drexler, "'Designing Competitive Markets for Industrial Data—Between Propertization and Access'" (2017) 8 *JIPITEC* 257. Zie ook de Nederlandse visie op datadeling tussen bedrijven uit 2019, blz. 15, beschikbaar op <https://www.government.nl/documents/reports/2019/02/01/dutch-vision-on-data-sharing-between-businesses>: 'In eigendomsrechtelijke termen is er niemand die data 'bezit' omdat er geen sprake is van fysieke objecten'.

⁵⁷ Artikelen 2 en 3 van Richtlijn 2001/29/EG van het Europees Parlement en de Raad van 22 mei 2001 betreffende de harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij (Richtlijn Informatiemaatschappij) [2001] PB L 167/10 zoals gewijzigd door Richtlijn (EU) 2019/790 van het Europees Parlement en de Raad van 17 april 2019 inzake auteursrechten en naburige rechten in de digitale eengemaakte markt [2019] PB L 130/92.

de algemene voorwaarden van platforms doorgaans de provider een licentie verlenen om al hun inhoud te gebruiken, inclusief inhoud beschermd door intellectuele eigendomsrechten. Het bestaan van verschillende parallelle claims ten aanzien van inhoud kan handhavingsproblemen opleveren, zoals voorkomt in de muzieksector.

Een ander voorbeeld van dergelijke problemen is een geschil uit 2011 tussen databeschermingsactivist Max Schrems en Facebook. In reactie op het verzoek van Schrems om toegang te krijgen tot zijn persoonsdata op grond van de databeschermingswet, weigerde Facebook toegang te verlenen tot alle persoonsdata die het over Schrems bezat, omdat een deel van zijn data kon worden aangemerkt als bedrijfsgeheim of intellectueel eigendom van Facebook.⁵⁸ Rechten of plichten inzake datadeling kunnen ook betrekking hebben op oorspronkelijke inhoud die door gebruikers is geüpload, bijvoorbeeld wanneer het recht op dataportabiliteit krachtens de AVG door een gebruiker wordt aangevoerd om zijn of haar profiel over te zetten naar een andere sociaalnetwerkprovider. Waar het profiel van een gebruiker deels wordt beschermd door het auteursrecht van de sociaalnetwerkprovider, is er sprake van parallelle wettelijke rechten op de data, namelijk de rechten van de gebruiker op zijn of haar persoonsdata en het auteursrecht op oorspronkelijke inhoud dat in licentie is gegeven aan de sociaalnetwerkprovider. Bij sectie 5.2 hieronder wordt geanalyseerd hoe onder de richtlijnen voor het recht op dataportabiliteit, zoals gepubliceerd door de Groep Databescherming Artikel 29 (Groep Databescherming Artikel 29, 2017), wordt gepoogd om de relatie van het recht op dataportabiliteit krachtens de AVG in overeenstemming te brengen met intellectuele eigendomsrechten en bedrijfsgeheimen.

Bescherming van databanken onder het auteursrecht en onder het sui generis databankenrecht

Naast de bescherming van data die onder het auteursrecht kunnen worden aangemerkt als oorspronkelijke inhoud, kunnen databanken worden beschermd onder bepaalde intellectuele eigendomsrechten. De databankrichtlijn voorziet in een tweeledig regime van bescherming van databanken door auteursrechtelijke bescherming te verlenen aan de structuur van oorspronkelijke databanken en door een sui generis recht in te stellen ter bescherming van de inhoud van databanken. Als een dataset volgens de richtlijn wordt aangemerkt als een 'databank', en voldoet aan de overige voorwaarden, kan deze ook profiteren van de twee parallelle vormen van bescherming. We zullen deze rechten nu achtereenvolgens toelichten.

De databankrichtlijn definieert een databank als *'een verzameling van werken, data of andere zelfstandige elementen, systematisch of methodisch geordend, en afzonderlijk met elektronische middelen of anderszins toegankelijk'*.⁵⁹ De interpretatie hiervan voor de doeleinden van dit onderzoek houdt in dat een dataset kan worden aangemerkt als een databank, mits de inhoud ervan kan worden opgedeeld in afzonderlijke items met onafhankelijke waarde⁶⁰, en kan worden geclassificeerd met behulp van een bepaalde methode waarmee ook individuele stukjes informatie kunnen worden opgehaald. Gezien de gestructureerde manier waarop technologiebedrijven data opslaan, is het waarschijnlijk dat hun datasets zoals bestudeerd in het kader van dit onderzoek kunnen worden beschouwd als databanken voor de doeleinden van de databankrichtlijn. Als een dataset is aangemerkt als een databank, kan deze profiteren van bescherming onder het auteursrecht en/of sui generis databankenrecht onder bepaalde voorwaarden, waaraan hier aandacht zal worden besteed.

⁵⁸ Zie G. Moody, 'Facebook Says Some of Your Personal Data Is Its 'Trade Secrets or Intellectual Property'', Techdirt, 12 oktober 2011, beschikbaar op <https://www.techdirt.com/articles/20111011/12190216306/facebook-says-some-your-personal-data-is-its-trade-secrets-intellectual-property.shtml>

⁵⁹ Artikel 1, lid 2 van Richtlijn 96/9/EG van het Europees Parlement en de Raad van 11 maart 1996 betreffende de rechtsbescherming van databanken (databankrichtlijn) [1996] PB L 77/20.

⁶⁰ Ter illustratie van het begrip 'onafhankelijkheid': volgens overweging 17 van de databankrichtlijn valt 'de vastlegging van een audiovisueel, cinematografisch, literair of muzikaal werk als zodanig' 'niet binnen de werkingsfeer van de richtlijn, omdat deze alleen waarde hebben als compleet oeuvre.

Om voor auteursrechtelijke bescherming in aanmerking te komen, moeten databanken 'een eigen intellectuele schepping van de maker' zijn door de manier waarop hun inhoud is geselecteerd of gerangschikt, zoals bepaald in Artikel 3, lid 1 van de databankrichtlijn. De auteursrechtelijke bescherming van een databank heeft alleen betrekking op de structuur van de data en strekt zich niet uit tot de inhoud.⁶¹ Ter verduidelijking hiervan stelde het Hof van Justitie dat een databank de eigen intellectuele schepping van de maker vormt 'wanneer de maker ervan, bij de keuze of de rangschikking van de data van deze databank, zijn creatieve vaardigheden op een originele manier tot uiting heeft kunnen brengen door het maken van vrije en creatieve keuzes en zijn werk de "persoonlijke noot" van de maker bevat'.⁶² Aan het oorspronkelijkheidsvereiste wordt niet voldaan 'wanneer voor de samenstelling van de databank technische overwegingen, regels of beperkingen gelden die geen ruimte laten voor creatieve vrijheid'.⁶³ Door deze inzichten toe te passen in het kader van dit onderzoek, kunnen datasets in aanmerking komen voor auteursrechtelijke bescherming zolang er ruimte is voor de auteur om vrije en creatieve keuzes te maken over de selectie of rangschikking van de data. De selectie of rangschikking van data komt niet in aanmerking voor auteursrechtelijke bescherming als daaraan puur technische overwegingen ten grondslag liggen. Met het oog op de verscheidenheid aan manieren waarop technologiebedrijven hun datasets kunnen rangschikken en het ontbreken van uniforme standaarden, is het waarschijnlijk dat hun dataets voldoen aan de voorwaarden voor bescherming onder het auteursrecht. Naast auteursrechtelijke bescherming kan een databank ook profiteren van bescherming op grond van het sui generis databankenrecht dat is ingesteld krachtens Artikel 7, lid 1 van de databankrichtlijn, indien 'de verkrijging, de controle of de presentatie van de inhoud in kwalitatief of kwantitatief opzicht getuigt van een substantiële investering'. Waar voor auteursrechtelijke bescherming van databanken een oorspronkelijke uiting van de creatieve vrijheid van de maker vereist is, is "investering" het relevante criterium voor bescherming onder het sui generis (wat "uniek" betekent) databankenrecht. Een databank die niet voldoet aan de oorspronkelijkheidsvoorwaarde van het auteursrecht kan daarom toch profiteren van bescherming onder het sui generis recht, zolang er aanzienlijke menselijke, technische of financiële middelen zijn geïnvesteerd in de fabricage ervan.⁶⁴ Voor het verwerken van deze inzichten in dit onderzoek is het een relevante vraag in hoeverre investeringen in datagenererende activiteiten buiten beschouwing moeten worden gelaten bij het bepalen of een dataset in aanmerking komt voor sui generis bescherming. Op basis van arresten van het Hof van Justitie zijn alleen investeringen in de verkrijging, controle of presentatie van reeds bestaande inhoud relevant voor sui generis databankenbescherming. Investerings in het creëren van elementen die de inhoud van een databank vormen komen niet in aanmerking.⁶⁵

In het licht van het bovenstaande is het voor datadeling in de technologiesector een cruciale vraag in hoeverre het verzamelen van door computers gegenereerde data of sensordata moet worden beschouwd als creatief investeren in plaats van simpelweg als het verzamelen van data.⁶⁶ Blijkens de evaluatie van de databankrichtlijn door de Europese Commissie uit 2018, zou het efficiënt zijn voor de dataeconomie om het verzamelen van door computers gegenereerde data en sensordata aan te merken als investeringen in het creëren van data, en zo dergelijke data uit te sluiten van sui

⁶¹ Artikel 3, lid 2 van de databankrichtlijn.

⁶² Zaak C-604/10 *Football Dataco Ltd en Anderen v. Yahoo! UK Ltd en Anderen*, ECLI:EU:C:2012:115, par. 38.

⁶³ Zaak C-604/10 *Football Dataco Ltd en Anderen v. Yahoo! UK Ltd en Anderen*, ECLI:EU:C:2012:115, par. 39.

⁶⁴ Overweging 7 van Richtlijn 96/9/EG van de databankrichtlijn.

⁶⁵ Uitspraak in *Fixtures Marketing Ltd v. Oy Veikkaus Ab*, C-46/02, ECLI:EU:C:2004:694, par. 34 en 37; uitspraak in *The British Horseracing Board Ltd en Anderen v. William Hill Organization Ltd*, C-203/02, ECLI:EU:C:2004:695, par. 31 en 34; uitspraak in *Fixtures Marketing Ltd v. Svenska Spel AB*, C-338/02, ECLI:EU:C:2004:696, par. 24 en 27; uitspraak in *Fixtures Marketing Ltd v. Organismos prognostikon agonon podostairou AE (OPAP)*, C-444/02, ECLI:EU:C:2004:697, par. 40 en 43.

⁶⁶ Werkdocument van de diensten van de Commissie – Evaluatie van Richtlijn 96/9/EG inzake de wettelijke bescherming van databanken, SWD (2018) 146 definitief, 25 april 2018, blz.35-40.

generis bescherming.⁶⁷ Toegang tot dergelijke data met het oog op hergebruik zou dan niet afhankelijk zijn van toestemming van de eigenaar van de databank. Mits aan de relevante voorwaarden wordt voldaan, kunnen datasets in aanmerking komen voor auteursrecht- en/of sui generis bescherming onder de databankrichtlijn. Beide soorten bescherming laten eventuele rechten op basis van onderliggende aspecten of data waaruit de databank bestaat, zoals auteursrechten (bijvoorbeeld met betrekking tot oorspronkelijke inhoud geüpload door gebruikers), en ook databescherming onverlet.⁶⁸

Bescherming van data als bedrijfsgeheim

Ten slotte kunnen data ook profiteren van bescherming op grond van de richtlijn bedrijfsgeheimen. Daartoe gelden de volgende voorwaarden: (1) de informatie is geheim in die zin dat zij, in haar geheel dan wel in de juiste samenstelling en ordening van haar bestanddelen, niet algemeen bekend is bij of gemakkelijk toegankelijk is voor personen binnen de kringen die zich gewoonlijk bezighouden met de desbetreffende soort informatie; (2) de informatie bezit handelswaarde omdat zij geheim is; en (3) de informatie is door de persoon die rechtmatig daarover beschikt onderworpen aan redelijke maatregelen, gezien de omstandigheden, om deze geheim te houden.⁶⁹

Wat betreft het soort informatie dat voor bescherming in aanmerking komt, wordt in de overwegingen van de richtlijn bedrijfsgeheimen vermeld dat dit kan gelden voor *'een breed scala aan informatie, van technologische kennis tot handelsdata zoals informatie over klanten en leveranciers, bedrijfsplannen, en marktonderzoek en marktstrategieën'*.⁷⁰ In het kader van dit onderzoek betekent dit dat diverse soorten data of datasets die commerciële waarde creëren voor hun houders wanneer ze zijn afgeschermd van derden, in aanmerking kunnen komen voor bescherming als bedrijfsgeheim voor zover aan de drie voorwaarden van bovengenoemde richtlijn bedrijfsgeheimen is voldaan.⁷¹ Het voorbeeld aan het begin van dit onderdeel over het geschil tussen Max Schrems en Facebook toont aan dat technologiebedrijven zich beroepen op bescherming van bedrijfsgeheimen voor hun datasets.

Rechten verleend aan houders van auteursrechten, databankenrechten en bedrijfsgeheimen

Elk van de drie soorten bescherming verleent verschillende rechten aan de houders van data(sets).

Het auteursrecht geeft auteurs het recht om te verhinderen dat derden een tweetal onrechtmatige handelingen verrichten, namelijk (1) reproductie en (2) mededeling aan het publiek.⁷²

Het sui generis databankenrecht geeft makers van een databank het recht om te verhinderen dat derden een tweetal onrechtmatige handelingen verrichten: (1) de opvraging en/of het hergebruik van het geheel of een in kwalitatief of kwantitatief opzicht substantieel deel van die inhoud⁷³ en (2) het herhaald en systematisch opvragen en/of hergebruiken van niet-substantiële delen van de inhoud van de databank, in strijd met een normale exploitatie van die databank of waardoor

⁶⁷ Werkdocument van de diensten van de Commissie – Evaluatie van Richtlijn 96/9/EG inzake de wettelijke bescherming van databanken, SWD (2018) 146 definitief, 25 april 2018, blz. 35.

⁶⁸ Zie respectievelijk Artikelen 3, lid 2 en 7, lid 4 van de databankrichtlijn.

⁶⁹ Artikel 2, lid 1 van Richtlijn (EU) 2016/943 van het Europees Parlement en de Raad van 8 juni 2016 inzake de bescherming van niet openbaar gemaakte knowhow en bedrijfsinformatie (bedrijfsgeheimen) tegen het onrechtmatig verkrijgen, gebruiken en openbaar maken daarvan (richtlijn bedrijfsgeheimen) [2016] PB L 157/1.

⁷⁰ Overweging 2 van de richtlijn bedrijfsgeheimen.

⁷¹ Bovendien kunnen algoritmen in aanmerking komen voor bedrijfsgeheimbescherming.

⁷² Artikelen 2 en 3 van Richtlijn 2001/29/EG van het Europees Parlement en de Raad van 22 mei 2001 inzake de harmonisatie van bepaalde aspecten van het auteursrecht en naburige rechten in de informatiemaatschappij (Richtlijn Informatiemaatschappij) [2001] PB L 167/10 zoals gewijzigd door Richtlijn (EU) 2019/790 van het Europees Parlement en de Raad van 17 april 2019 inzake het auteursrecht en naburige rechten in de digitale interne markt [2019] PB L 130/92.

⁷³ Artikel 7, lid 1 van de databankrichtlijn.

ongerechtvaardigde schade wordt toegebracht aan de rechtmatige belangen van de fabrikant van de databank⁷⁴.

In tegenstelling tot auteursrechtelijke en sui generis databankenbescherming, geeft de bescherming van het bedrijfsgeheim de rechthebbenden geen exclusief recht om te verhinderen dat derden de data(sets) gebruiken, maar kan uitsluitend een beroep worden gedaan op dergelijke bescherming ingeval van onrechtmatige verkrijging, onrechtmatig gebruik en onrechtmatige openbaarmaking ervan.⁷⁵ Bescherming van het bedrijfsgeheim is dus niet van kracht bij het gebruik van legitiem verkregen informatie, die bijvoorbeeld onafhankelijk is ontdekt of gecreëerd.⁷⁶ Geheimhoudings- of vertrouwelijkheidsovereenkomsten worden vaak gebruikt om commercieel waardevolle informatie geheim te houden en bescherming van het bedrijfsgeheim te waarborgen.

Relevantie voor datadeling

De mate waarin data worden beschermd door bepaalde intellectuele eigendomsrechten of bedrijfsgeheimen is relevant voor datadeling, omdat het bestaan van rechten of plichten ten aanzien van toegang tot data en het creëren van aanvullende rechten of plichten in dat verband de draagwijdte van de bescherming onder het auteursrecht en databankenrechten met betrekking tot datasets kan beperken. Bestaande mechanismen voor datadeling, zoals het recht op dataportabiliteit krachtens de AVG, vallen al samen met deze rechten, zoals blijkt uit het geschil tussen Max Schrems en Facebook aan het begin van dit hoofdstuk. Door houders van auteursrechten of databankenrechten te verplichten om data(sets) uit te wisselen, of door rechten te creëren voor derden om toegang te krijgen tot data(sets) die in potentie beschermd zijn onder deze rechten (zoals het reeds bestaande recht op dataportabiliteit krachtens de AVG), wordt het recht van de houders van deze rechten om derden uit te sluiten van het gebruik van de data(sets) beperkt. Intellectueel eigendom geniet niet alleen bescherming in de vorm van specifieke rechten, maar heeft ook de status van een grondrecht onder Artikel 17 van het Handvest van de Grondrechten van de Europese Unie en Artikel 1 van Protocol nr. 1 bij het Europees Verdrag inzake de Rechten van de Mens. Tegelijkertijd is de bescherming van intellectueel eigendom niet absoluut, maar kan deze worden afgewogen tegen andere rechten of belangen. Weigeringen om licenties voor intellectuele eigendomsrechten te verlenen zijn bijvoorbeeld als onrechtmatig beschouwd op grond van het EU-mededingingsrecht, wat heeft geleid tot licentieverplichtingen en dus tot beperking van de exclusiviteit verleend aan houders van intellectuele eigendomsrechten.⁷⁷

In de Mededeling "Een Europese datastrategie" uit 2020, heeft de Europese Commissie haar voornemen uitgesproken om het kader voor intellectueel eigendom te evalueren "*met het oog op een verdere verbetering van de toegang tot en het gebruik van data*". Specifiek wordt melding gemaakt van een mogelijke herziening van de databankrichtlijn en een verduidelijking van de toepassing van de richtlijn bedrijfsgeheimen als ondersteunend kader.⁷⁸ In het kader van de Datawet, die de Commissie momenteel in voorbereiding heeft en die naar verwachting in 2021 zal worden gepubliceerd, wordt een leidraad verwacht voor de afweging van de voordelen van datadeling tegen de mogelijke inbreuk op de reikwijdte van de bescherming geboden door

⁷⁴ Artikel 7, lid 5 van de databankrichtlijn.

⁷⁵ In leden 2-5 van Artikel 4 van de richtlijn bedrijfsgeheimen worden de omstandigheden vermeld waarin de verkrijging, het gebruik en het openbaar maken van een bedrijfsgeheim als onrechtmatig worden beschouwd.

⁷⁶ In artikel 3 van de richtlijn bedrijfsgeheimen worden de situaties opgesomd waarin de verkrijging, het gebruik en de openbaarmaking van een bedrijfsgeheim als rechtmatig worden beschouwd.

⁷⁷ Zie gevoegde zaken C-241/91 en C-242/91, *Magill*, ECLI:EU:C:1995:98; zaak C-418/01, *IMS Health*, ECLI:EU:C:2004:257; en zaak T-201/04, *Microsoft*, ECLI:EU:T:2007:289.

⁷⁸ Mededeling van de Commissie aan het Europees Parlement, de Raad, het Europees Economisch en Sociaal Comité en het Comité van de Regio's 'Een Europese datastrategie', COM (2020) 66 definitief, 19 februari 2020, blz. 13. Het kabinet verwelkomt ook het voornemen van de Commissie om het intellectueel eigendomsrechtkader te evalueren met het oog op de verdere verbetering van datatoegang en –gebruik. Zie BNC-fiche 'Mededeling over een Europese datastrategie', beschikbaar op <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2020/04/14/mededeling-over-een-europese-datastrategie/fiche-1-mededeling-over-een-europese-datastrategie.pdf>.

eerdergenoemde wetgevingsinstrumenten. Een voorbeeld van de uitkomst van een dergelijke eerdere afweging is de goedkeuring van vrijstellingen van auteursrechten en sui generis databankenbescherming om tekst- en datamining uit te voeren onder de voorwaarden zoals vastgelegd in Richtlijn 2019/790.⁷⁹ Indien dit wenselijk wordt geacht, kunnen dergelijke vrijstellingen verder worden uitgebreid om de toegang tot data op EU- of nationaal niveau te bevorderen. Voorgesteld is om ook nieuwe rechten of plichten ten aanzien van datadeling te integreren in de intellectuele eigendomsregimes zelf, bijvoorbeeld door het recht op intellectuele eigendomsbescherming voor data(sets) afhankelijk te stellen van verplichte licentieverlening van diezelfde data door de rechthebbende (Drexl, 2018). Hoewel de Europese Commissie dit als mogelijkheid noemde in haar evaluatie van de databankrichtlijn uit 2018,⁸⁰ kan een dergelijke benadering in de praktijk moeilijk te implementeren zijn, vanwege onzekerheden over hoe het begrip 'data' moet worden gedefinieerd en de vorm van de relevante intellectuele eigendomsrechten die het juridisch mogelijk moeten maken voor rechthebbenden om aan derden bepaalde vormen van onrechtmatig gebruik te verbieden.

Zoals in de punten hieronder wordt aangetoond, kunnen de bestaande databeschermings- en mededingingsrechtregimes datadeling al vergemakkelijken en daarbij een wisselwerking aangaan met de bescherming die wordt geboden aan data(sets) onder het auteursrecht, het sui generis databankenrecht en bedrijfsgeheimen.

4.2 EU-wetgeving inzake bescherming van data

De AVG⁸¹ bevat wettelijke regels die van toepassing zijn op de verwerking van persoonsdata. Zoals blijkt uit Artikel 1, streeft de AVG een tweeledig doel na: het waarborgen van het grondrecht op databescherming, en het stimuleren van het vrije verkeer van persoonsdata binnen de Europese Unie. Daartoe geeft de AVG betrokkenen⁸² rechten die hun controle over persoonsdata versterken, en legt zij de verwerkingsverantwoordelijken⁸³ plichten op om het verzamelen en gebruiken van persoonsdata te beperken.

Reikwijdte en toepassing van de AVG

Om de AVG van toepassing te laten zijn, moet de informatie in kwestie kunnen worden aangemerkt als 'persoonsdata', zoals gedefinieerd in Artikel 4, lid 1 als "alle informatie met betrekking tot een geïdentificeerde of identificeerbare natuurlijke persoon". Het begrip persoonsdata wordt breed geïnterpreteerd en omvat informatie die "wegens haar inhoud, doel of gevolg gelieerd is aan een bepaalde persoon".⁸⁴ Zelfs onbedoelde effecten van dataverwerking en de redelijke waarschijnlijkheid van identificeerbaarheid kunnen informatie tot persoonsdata bestempelen. In verband met deze interpretatie is geopperd dat, gegeven de vooruitgang op het gebied van data-

⁷⁹ Artikelen 3 en 4 van Richtlijn (EU) 2019/790 van het Europees Parlement en de Raad van 17 april 2019 inzake auteursrechten en naburige rechten in de digitale eengemaakte markt en tot wijziging van Richtlijnen 96/9/EG en 2001/29/EG PB L 130/92

⁸⁰ Werkdocument van de diensten van de Commissie - Evaluatie van Richtlijn 96/9/EG inzake de rechtsbescherming van databanken, SWD (2018) 146 definitief, 25 april 2018.

⁸¹ Verordening (EU) 2016/679 van 27 april 2016 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsdata en betreffende het vrije verkeer van die data en tot intrekking van Richtlijn 95/46/EG (AVG) [2016] PB L 119/1.

⁸² In Artikel 4, lid 1 van de AVG wordt betrokkene gedefinieerd als 'een identificeerbare natuurlijke persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identifier zoals een naam, een identificatienummer, locatiedata, een online identifier of van een of meer elementen die kenmerkend zijn voor de fysieke, fysiologische, genetische, psychische, economische, culturele of sociale identiteit van die natuurlijke persoon'.

⁸³ In Artikel 4, lid 7 van de AVG wordt een verwerkingsverantwoordelijke gedefinieerd als 'een natuurlijke persoon of rechtspersoon, een overheidsinstantie, een dienst of een ander orgaan die/dat, alleen of samen met anderen, het doel van en de middelen voor de verwerking van persoonsdata vaststelt'.

⁸⁴ Zaak C-434/16 Nowak, ECLI:EU:C:2017:994, par. 35.

analyse, elke informatie als persoonsdata kan worden beschouwd, waarmee databescherming de 'wet van alles' wordt (Purtova, 2018).

Hoewel het niet duidelijk is hoe effectief anonimiseringstechnieken nog zijn met de toenemende mogelijkheden tot heridentificatie (Groep Databescherming Artikel 29, 2017), is de AVG niet van toepassing op anonieme informatie, namelijk "data die geen betrekking hebben op een geïdentificeerde of identificeerbare natuurlijke persoon of op persoonsdata die zodanig anoniem zijn gemaakt dat de betrokkene niet of niet meer identificeerbaar is".⁸⁵ In het kader van datadeling betekent dit dat de AVG niet van toepassing is op de uitwisseling van persoonsdata die anoniem zijn (gemaakt). Er moet echter rekening mee worden gehouden dat de anonimisering op zich de verwerking inhoudt van persoonsdata waarop de bepalingen van de AVG van toepassing zijn. Onder de bepalingen van de AVG wordt datadeling zowel beperkt als gefaciliteerd, zoals hieronder wordt besproken.

Rechtmatigheid, doelbinding en minimale dataverwerking als beperkingen voor datadeling

Onder de AVG is de verwerking (inclusief de uitwisseling) van persoonsdata als zodanig niet verboden, maar zijn de verwerkingsvoorwaarden vastgelegd. Een belangrijke voorwaarde is de rechtmatigheid (of rechtsgrond) inzake de verwerking van persoonsdata.⁸⁶ Rechtsgronden die relevant zijn voor onze doeleinden zijn onder meer toestemming van de betrokkene, uitvoering van een overeenkomst, voldoen aan een wettelijke verplichting, en gerechtvaardigde belangen van de verwerkingsverantwoordelijke.⁸⁷

Veel onlinediensten gaan uit van de toestemming van de betrokkene als rechtsgrond voor de verwerking van persoonsdata, die doorgaans worden verzameld nadat de gebruiker het privacybeleid of de algemene voorwaarden van een dienst heeft geaccepteerd. Een vraag die momenteel door wetenschappers wordt besproken is of het opleggen van een verplichting om data uit te wisselen onder het mededingingsrecht een rechtsgrond is voor dataverwerking in de zin van artikel 6, lid 1 van de AVG.⁸⁸ Is dit het geval, dan is toestemming van betrokkenen niet nodig om persoonsdata uit te wisselen.⁸⁹ Afgezien van de theorie en de technische details, lijkt het echter discutabel om een verwerkingsverantwoordelijke toe te staan persoonsdata uit te wisselen met derden zonder dat dit enige positieve actie van de kant van betrokkenen vereist. Zoals hieronder nader wordt besproken, vormt de spanning tussen enerzijds de behoefte aan datadeling als motor voor concurrentie en innovatie en anderzijds de bevoegdheid van individuele personen wier persoonsdata worden uitgewisseld een van de belangrijkste uitdagingen op dit beleidsgebied.

Naast de voorwaarde van rechtmatigheid, zijn ook de andere principes voor de verwerking van persoonsdata, zoals gesteld in Artikel 5, lid 1 van de AVG, relevant. Twee van deze principes kunnen leiden tot beperking van datadeling, namelijk doelbinding⁹⁰ en minimale dataverwerking⁹¹.

- **Het principe van doelbinding** houdt in dat persoonsdata uitsluitend voor specifieke doeleinden kunnen worden verzameld en niet verder kunnen worden verwerkt op een manier die in strijd is met die doeleinden. In het kader van datadeling houdt dit in dat persoonsdata niet kunnen worden uitgewisseld als een dergelijke uitwisseling niet strookt met het oorspronkelijke doel waarvoor de data zijn verzameld. Als het nieuwe doel echter verenigbaar is met het

⁸⁵ Overweging 26 van de AVG.

⁸⁶ Artikel 5, lid 1 van de AVG

⁸⁷ Artikel 6, lid 1 van de AVG

⁸⁸ (Graef, 2016; Kathuria & Globocnik, 2019)

⁸⁹ Aangezien het bedrijf in kwestie anders zou afzien van naleving van een mededingingsrechtelijke voorziening en dus aansprakelijk zou worden gesteld, lijkt het erop dat een mededingingsrechtelijke verplichting vanuit een puur juridisch-technisch perspectief kan worden aangemerkt als een wettelijke verplichting tot datadeling voor de doeleinden van de eis tot een wettige reden voor verwerking op grond van de AVG.

⁹⁰ Artikel 5, lid 1(b) van de AVG

⁹¹ Artikel 5, lid 1(c) van de AVG

oorspronkelijke doel, is er geen andere rechtsgrondslag vereist dan die op grond waarvan de verzameling van persoonsdata werd toegestaan.⁹² Een belangrijke overweging voor de beoordeling of datadeling is gebaseerd op een verenigbaar doel, is of betrokkenen redelijkerwijs hadden kunnen verwachten dat de verwerkingsverantwoordelijke de data in kwestie zou uitwisselen met derden (Graef, Tombal & De Streel, 2019).⁹³

- **Het principe van minimale dataverwerking** schrijft voor dat persoonsdata adequaat en relevant moeten zijn, en zich moeten beperken tot wat nodig is ten aanzien van de doeleinden waarvoor ze worden verwerkt. Voor datadeling betekent dit dat de hoeveelheid uit te wisselen data zo beperkt mogelijk moet zijn. Dientengevolge moet het doel waarvoor de data worden uitgewisseld duidelijk worden omschreven, en moet de verwerkingsverantwoordelijke die persoonsdata uitwisselt ervoor zorgen dat geen data worden uitgewisseld die niet direct noodzakelijk zijn om dat doel te bereiken.

Recht op dataportabiliteit als facilitator van datadeling

Hoewel de vereisten krachtens de AVG inzake rechtmatigheid, doelbinding en minimale dataverwerking beperkingen opleggen aan de mate waarin persoonsdata kunnen worden uitgewisseld, heeft de AVG ook een recht op dataportabiliteit ingevoerd dat uitwisseling en hergebruik van persoonsdata faciliteert op verzoek van de betrokkene. Volgens Artikel 20, lid 1 van de AVG heeft betrokkene het recht de hem betreffende persoonsdata, die hij aan een verwerkingsverantwoordelijke heeft verstrekt, in een gestructureerde, gangbare en machineleesbare vorm te verkrijgen (d.w.z. verzoek met het recht op verkrijgen), en hij heeft het recht die data aan een andere verwerkingsverantwoordelijke over te dragen, zonder daarbij te worden gehinderd. Volgens Artikel 20, lid 2 van de AVG heeft betrokkene het recht om de persoonsdata 'indien dit technisch mogelijk is' rechtstreeks van de ene verwerkingsverantwoordelijke naar de andere door te zenden.

Het recht op dataportabiliteit is van toepassing wanneer de verwerking is gebaseerd op toestemming of uitvoering van een contract, en wanneer de verwerking langs geautomatiseerde weg wordt uitgevoerd. In 2017 publiceerde de Groep Databescherming Artikel 29 richtlijnen met betrekking tot het recht op dataportabiliteit, ter verduidelijking van de voorwaarden waaronder het nieuwe recht van toepassing zou zijn. De richtlijnen zijn niet juridisch bindend, maar hebben een gezaghebbende status.

Een van de kwesties die de richtlijnen hebben verduidelijkt was welke persoonsdata geacht worden te vallen onder de term 'verstrek door' betrokkene. Volgens de Groep Databescherming Artikel 29 vallen onder 'verstrekte' persoonsdata zowel data die bewust en actief zijn verstrekt door de betrokkene, zoals een gebruikersnaam, e-mailadres of iemands leeftijd, als data die zijn waargenomen bij de activiteiten van gebruikers, waaronder door een slimme meter verwerkte ruwe data, activiteitenoverzichten of een zoekgeschiedenis. Afgeleide of verkregen data die door de verwerkingsverantwoordelijke zijn aangemaakt op basis van data die door de betrokkene zijn verstrekt, zoals gebruikersprofielen of algoritmische resultaten van een gezondheids- of financiële beoordeling, vallen echter buiten de reikwijdte van het recht op dataportabiliteit (Groep Databescherming Artikel 29, 2017).

⁹² Overweging 50 van de AVG.

⁹³ Om de compatibiliteit te beoordelen op basis van Artikel 6, lid 4 van de AVG, moet de verwerkingsverantwoordelijke bovendien rekening houden met: (a) een eventuele koppeling tussen die doeleinden en de doeleinden van de voorgenomen verdere verwerking; (b) het kader waarin de data zijn verzameld; met name de redelijke verwachtingen van de betrokkenen op basis van hun verhouding met de verwerkingsverantwoordelijke betreffende het verdere gebruik ervan; (c) de gevolgen van de voorgenomen verdere verwerking voor de betrokkenen; en (e) passende waarborgen bij zowel de oorspronkelijke als de voorgenomen verdere verwerkingen.

De opname van waargenomen data als onderdeel van persoonsdata die zijn 'verstrekkt door' betrokkene, heeft geleid tot beweringen dat de interpretatie van de Groep Databescherming Artikel 29 verder gaat dan de oorspronkelijke bedoeling van de wetgever, en niet zorgt voor het juiste evenwicht met de belangen van verwerkingsverantwoordelijken die hebben geïnvesteerd in het verzamelen van waargenomen data.⁹⁴ Voor zover wij weten is er momenteel geen sprake van een handhavingsactie van een databeschermingsautoriteit of een rechterlijke uitspraak waardoor dit debat over de grenzen van de persoonsdata die binnen de reikwijdte van het recht op dataportabiliteit vallen kan worden beslecht.

Naast de notie van door betrokkene 'verstrekte' persoonsdata kan nog een andere relevante beperking worden aangetroffen in Artikel 20, lid 4 van de AVG, waarin wordt gesteld dat het recht op dataportabiliteit geen afbreuk doet aan de rechten en vrijheden van anderen. Hieronder vallen de databeschermingsrechten van andere betrokkenen, van wie de persoonsdata zijn vermengd met die van de betrokkene die een beroep doet op het recht op dataportabiliteit (bijvoorbeeld iemands interacties op een sociaal netwerk), en de intellectuele eigendomsrechten of bedrijfsgeheimen van verwerkingsverantwoordelijken ter bescherming van de investering in dataverzameling en data-analyse (bijvoorbeeld auteursrecht-, databanken- en bedrijfsgeheimenbescherming) (Graef, Husovec & Purtova, 2018).

Er vindt nog discussie plaats over hoe een dergelijke afweging van belangen op de juiste manier tot stand moet komen.⁹⁵ Inzake het intellectueel eigendom heeft de Groep Databescherming Artikel 29 toegelicht dat intellectueel eigendom en bedrijfsgeheimen in overweging moeten worden genomen alvorens een verzoek tot dataportabiliteit wordt beantwoord, maar dat "het resultaat van die overwegingen niet mag leiden tot een algehele weigering om informatie te verstrekken aan betrokkene" (Groep Databescherming Artikel 29, 2017). Daarom kan een verwerkingsverantwoordelijke een verzoek om dataportabiliteit niet afwijzen, omdat de persoonsdata van degene in kwestie ook worden beschermd door het intellectueel eigendom en de bedrijfsgeheimen van de verwerkingsverantwoordelijke. Volgens de Groep Databescherming Artikel 29 zouden verwerkingsverantwoordelijken mogelijkheden moeten onderzoeken om persoonsdata verstrekt door betrokkenen door te geven in een zodanige vorm dat geen informatie die onder bedrijfsgeheimen of intellectuele eigendomsrechten valt wordt prijsgegeven (Groep Databescherming Artikel 29, 2017). Er wordt echter niet aangegeven wat er moet gebeuren als dit niet mogelijk blijkt te zijn.

Ten aanzien van de effectiviteit van het recht op dataportabiliteit lopen de meningen uiteen over de vraag of Artikel 20 van de AVG alleen een eenmalige vorm van portabiliteit inhoudt, waar data eenmalig op verzoek van betrokkene aan een nieuwe verwerkingsverantwoordelijke worden doorgegeven, of dat het zo kan worden geïnterpreteerd dat het ook betrekking heeft op continue en real-time dataportabiliteit, totdat betrokkene het portabiliteitsverzoek intrekt.⁹⁶ Een dergelijke interpretatie zou het recht op dataportabiliteit geschikter maken voor dynamische settings (waarvan er veel bijvoorbeeld zijn verbonden met IoT-apparaten). Real-time portabiliteit vereist echter een technische infrastructuur die continue datastromen kan hanteren. En aangezien Artikel 20, lid 2 van de AVG verwijst naar 'technische haalbaarheid', zullen de beschikbare technische mogelijkheden waarschijnlijk van invloed zijn op de wijze waarop het recht op dataportabiliteit wordt gehandhaafd. Hoewel het recht op dataportabiliteit onder de AVG van toepassing is op elke vorm

⁹⁴ Zie bijvoorbeeld *Legal Memo with Respect to the Article 29 Guidelines on the Right to Data Portability*, European Telecommunications Network Operators' Association, 16 februari 2017, 8-9, beschikbaar op https://etno.eu/data/positions-papers/2017/170131%20ETNO_Data%20Portability_Memo/170131%20ETNO_Data%20Portability_Memo.pdf.

⁹⁵ Zie bijvoorbeeld het Witboek "Charting a Way Forward. Data Portability and Privacy", gepubliceerd door Facebook in september 2019. Beschikbaar op <https://about.fb.com/wp-content/uploads/2020/02/data-portability-privacy-white-paper.pdf>.

⁹⁶ (Cr mer, De Montjoye & Schweitzer, 2019; Graef, Husovec & Van den Boom, 2020)

van verwerking van persoonsdata, ongeacht de bedrijfstak, zal de realisatie ervan dus waarschijnlijk per sector verschillen, afhankelijk van de aanwezigheid van technische randvoorwaarden.

Met betrekking tot het recht op dataportabiliteit krachtens de AVG is het vermeldenswaard dat in het kader van het onderscheid tussen horizontale, verticale en sector-overschrijdende datadeling, Artikel 20 van de AVG zich agnostisch opstelt ten aanzien van het doel om data over te dragen. Het recht op dataportabiliteit als zodanig kan elk van deze drie soorten datadeling faciliteren. Horizontale datadeling vindt plaats wanneer de betrokkene data doorgeeft aan een concurrerende verwerkingsverantwoordelijke die een dienst biedt die een equivalent (of vervanging) is van de dienst geboden door de oorspronkelijke verwerkingsverantwoordelijke (bijvoorbeeld de overdracht van een gebruikersprofiel tussen vergelijkbare sociale netwerken). Verticale datadeling vindt plaats wanneer data worden overgedragen aan een verwerkingsverantwoordelijke die een aanvullende dienst biedt (bijvoorbeeld het gebruik van energieverbruiksdata om aanbiedingen te kunnen vergelijken in een vergelijkingsdienst). Sector-overschrijdende datadeling is mogelijk wanneer data worden overgedragen tussen bedrijfstakken (bijvoorbeeld de overdracht van bankrekeninginformatie aan de Big Techs). Met betrekking tot laatstgenoemde vorm van uitwisseling zijn sommigen van mening dat het recht op dataportabiliteit gevestigde technologiebedrijven helpt om nog meer data te bemachtigen en hun marktpositie te versterken door gebruikers over te halen om data van andere diensten over te dragen.⁹⁷

4.3 EU-mededingingsrecht

De structuur van het EU-mededingingsrecht bestaat uit een aantal takken. Voor het doel van dit onderzoek beperken we ons tot de takken van mededingingsregelingen (Artikel 101 van het VWEU) en misbruik van machtspositie (Artikel 102 van het VWEU), waarbij we een aantal inzichten opnemen uit het concentratieonderzoek (EU-Concentratieverordening). Als een regeling die de interne markt stimuleert en het concurrentieproces beschermt ten gunste van de consument, kan het EU-mededingingsrecht datadeling zowel beperken als mogelijk maken.

Informatieverkeer als beperking van datadeling

Datadeling is in strijd met Artikel 101 van het VWEU wanneer deze collusie tussen concurrenten bevordert. Hoewel een datadeelovereenkomst (ook wel 'datapooling' genoemd) een concurrentiebevorderend effect kan creëren door toetredingsdrempels te verlagen en de keuze van de consument te vergroten, kan deze concurrentiebeperkend werken wanneer de datadeling marktspelers in staat stelt zich bewust te worden van elkaars marktstrategieën, of wanneer de toegang tot data wordt beperkt tot bepaalde marktspelers.

In hoeverre toegangsvoorwaarden van een datapoolingsysteem concurrentiebeperkend werken is in het geding bij het concurrentieonderzoek dat in mei 2019 door de Europese Commissie in gang is gezet naar de organisatie Insurance Ireland (Europese Commissie, 2019a).⁹⁸ Met betrekking tot

⁹⁷ Zie de opmerkingen van de voorzitter van de Franse mededingingsautoriteit Isabelle de Silva in L. de Klein & L. Zhou, 'French, Australian agency chiefs share ideas on privacy, competition interplay', PARR, ABA Antitrust Voorjaarsbijeenkomst (2019), beschikbaar op https://www.acuris.com/assets/PaRRABAreport2019_1.pdf: 'Krachtens de AVG moeten ook bedrijven gebruikers toestaan hun data naar een ander platform te verplaatsen. De Autorité hoorde vervolgens van muziekstreamingdienst Deezer dat als gebruikers mogen vertrekken met een hele lijst met muziek die ze hebben gemaakt, het voor Apple of andere grote spelers gemakkelijker kan worden om de gebruiker over te nemen, zei De Silva'.

⁹⁸ Insurance Ireland beheert een databank waaraan daarbij aangesloten bedrijven continu data over verzekeringsclaims bijdragen, met als doel de detectie door verzekeraars van mogelijk frauduleus gedrag van eisers te vergemakkelijken en de juistheid van de door potentiële klanten verstrekte informatie te waarborgen. De Commissie maakt zich zorgen over het feit dat de voorwaarden die Insurance Ireland aan bedrijven heeft opgelegd om toegang te krijgen tot de databank die

de manier waarop datadeelovereenkomsten collusie kunnen bevorderen door informatieverkeer, wordt in de horizontale richtlijnen van de Commissie aandacht besteed aan hoe bepaalde factoren het informatieverkeer tussen concurrenten gevoelig kunnen maken voor collusie. Deze factoren zijn onder meer het strategische karakter van de informatie, het marktgebied van de bedrijven in kwestie, de geïndividualiseerde of geaggregeerde aard van de betreffende bedrijfsinformatie, hoe oud de informatie is, de frequentie van het informatieverkeer, de openbare of niet-openbare aard van de data, en of het informatieverkeer openbaar of niet-openbaar is.⁹⁹ Hoewel technisch informatieverkeer inzake de ontwikkeling van nieuwe producten of de interoperabiliteit tussen bestaande producten voor het overgrote deel niet tot problemen lijkt te leiden, zijn datapoolingovereenkomsten waarbij partijen strategische en concurrentie-informatie uitwisselen over prijzen of nieuwe innovaties, waarschijnlijk in strijd met Artikel 101 van het VWEU. Op de schaal tussen die twee uitersten bevinden zich datapools waarin niet rechtstreeks commercieel gevoelig informatieverkeer plaatsvindt, maar waarin informatie wordt uitgewisseld over een groot aantal klanten op een manier die uiteindelijk een lid van de pool in staat kan stellen om concurrerende inzichten te verkrijgen op basis van data-analyse (Graef, Tombal & De Streel, 2019). Dit zijn de datadeelovereenkomsten waarvan de wettigheid het moeilijkst kan worden beoordeeld (Lundqvist, 2018).

Een relevant precedent om de mate van aansprakelijkheid voor datadeling te bepalen is het *Asnef-Equifax* arrest, waarin het Hof van Justitie de compatibiliteit op grond van Artikel 101 van het VWEU van een door een aantal banken in Spanje opgezet kredietwaardigheidsregister heeft beoordeeld. Het doel van het register was om solvabiliteits- en kredietinformatie over klanten uit te wisselen om de risico's op leningen en kredietactiviteiten te kunnen bepalen. Het Hof voerde aan dat om concurrentiebeperking te voorkomen op grond van Artikel 101 van het VWEU: (1) de relevante markt in kwestie niet sterk geconcentreerd mag zijn; (2) het register niet de identiteit van de kredietverstrekkers moet kunnen onthullen, aangezien dit zou kunnen bijdragen tot de identificatie van de marktpositie of de commerciële strategie van hun concurrenten; en (3) het register zonder onderscheid toegankelijk moet zijn voor alle ondernemers die actief zijn in de relevante sfeer, zodat sommige ondernemers niet worden benadeeld omdat zij geen toegang hebben tot informatie die nodig is om risico's te beoordelen.¹⁰⁰ Een soortgelijke analyse kan worden gehanteerd om de compatibiliteit van uitwisselingsovereenkomsten voor andere soorten data te testen. Naar verwachting zal de Commissie duidelijkheid verschaffen over de voorwaarden waaronder datapooling verenigbaar is met Artikel 101 van het VWEU bij de voortgaande herziening van de horizontale richtlijnen.¹⁰¹ Zoals voorgesteld door Crémer, De Montjoye & Schweitzer (2019), kan de Commissie voor haar benadering van datapooling inspiratie putten uit haar beoordeling van octrooipooling en onderzoeks- en ontwikkelingsovereenkomsten.

Remedies voor datatoegang om datadeling mogelijk te maken

Hoewel datadeling dus onder bepaalde omstandigheden tot collusie kan leiden, kan deze tezelfdertijd worden opgelegd als remedie om een concurrentieprobleem op een markt aan te pakken. Het meest relevant zijn de gevallen waarin een onderneming met een machtspositie weigert om data uit te wisselen die "essentieel" of "onmisbaar" zijn voor een derde om een eigen product of dienst te ontwikkelen. Afhankelijk van de definitie van "essentieel", kan er een verband zijn met de essentiële faciliteiten-doctrine. Afgezien van de toepassing van deze doctrine, kan datadeling worden opgelegd als remedie onder het mededingingsrecht om andere vormen van misbruik van machtspositie aan te pakken (bijvoorbeeld discriminerende datatoegang) of om te

bedrijven mogelijk een concurrentienadeel hebben opgeleverd op de Ierse motorrijtuigenverzekeringmarkt ten opzichte van de bedrijven die al toegang hebben tot de databank (Europese Commissie, 2019a).

⁹⁹ Richtsnoeren inzake de toepasselijkheid van artikel 101 van het Verdrag betreffende de werking van de Europese Unie op horizontale samenwerkingsovereenkomsten [2011] PB C 11/1, par. 86-94.

¹⁰⁰ Zaak C-238/05, *Asnef-Equifax*, EU:C:2006:734, par. 58-61.

¹⁰¹ Zie https://ec.europa.eu/info/law/better-regulation/initiatives/ares-2019-4715393/public-consultation_en.

voorkomen dat bij fusies mededingingsbezwaren ontstaan. Hoewel de drempel voor het opleggen van datadeling in dergelijke situaties lager is dan onder de essentiële faciliteiten-doctrine, hangt de mogelijkheid van concurrentie-interventie in dit soort zaken af van de vraag of al dan niet sprake is van ander probleemgedrag. Hieronder beperken we ons dan ook tot onderzoek naar weigeringen tot datadeling die onder de essentiële faciliteiten-doctrine zouden vallen.

Weigering om diensten te verlenen is een vorm van misbruik van machtspositie onder Artikel 102 van het VWEU. Als remedie kan een mededingingsautoriteit of rechtbank de onderneming met een machtspositie verplichten om concurrenten toegang te verschaffen tot de gevraagde input onder de essentiële faciliteiten-doctrine. Dienovereenkomstig verplicht deze doctrine ondernemingen met een machtspositie om (onder bepaalde omstandigheden) toegang te verlenen tot input of activa die essentieel zijn voor concurrenten om een verwante markt te betreden.

De essentiële faciliteiten-doctrine is ontstaan in de VS, in een zaak uit 1980 onder de Sherman Act (*Verenigde Staten v. Terminal Railroad Association of St. Louis*). Een essentiële faciliteit is in het mededingingsbeleid vanouds "een faciliteit of infrastructuur waartoe concurrenten noodzakelijkerwijs toegang moeten hebben om diensten te kunnen verlenen aan hun klanten", zodanig dat deze een aanvulling moet zijn op een economische activiteit in een verwante maar afzonderlijke markt, concurrenten deze niet kunnen dupliceren, en de toegang ertoe noodzakelijk is om te kunnen concurreren in de verwante markt.¹⁰²

In het kader van het EU-mededingingsrecht is de doctrine ontwikkeld in de loop van een lange reeks zaken die te maken hadden met toegang tot fysieke infrastructuren en het verlenen van licenties ten aanzien van intellectuele eigendomsrechten.¹⁰³ Afgezien van interventies op nationaal niveau om datasets uit te wisselen met als doel om meer traditionele markten toegankelijk te maken, waaronder de markt voor gasleveranties¹⁰⁴ en loterijdiensten¹⁰⁵, hebben zich voor zover ons bekend tot dusver op EU-niveau geen mededingingszaken voorgedaan die datatoegang hebben afgedwongen op de data-gedreven markten waarop we ons richten in dit onderzoek. Een aantal gevallen op EU-niveau kan echter in ruimere zin worden geïnterpreteerd als verband houdend met informatiemiddelen. In de zaak *Magill* concludeerde het Hof van Justitie dat de weigering van drie Ierse omroepen om de uitgeverij Magill een auteursrechtlicentie te verlenen voor de wekelijkse overzichten van hun televisieprogramma's onrechtmatig was.¹⁰⁶ In de zaak *IMS Health* constateerde het Hof van Justitie misbruik in het geval van een weigering van IMS, een bedrijf dat zich bezighoudt met informatieverstrekking over de regionale verkoop van farmaceutische producten in Duitsland, om een licentie te verlenen aan zijn concurrent NDC voor het gebruik van de auteursrechtelijk beschermde bouwsteenstructuur die IMS had ontwikkeld en die een *de facto* standaard was geworden.¹⁰⁷ In de zaak *Microsoft* oordeelde het Gerecht dat Microsoft onrechtmatig handelde door te weigeren concurrenten te voorzien van interoperabiliteitsinformatie die nodig was voor werkgroepservers om via niet-Microsoft besturingssystemen te communiceren met Microsofts dominante Windows client-PC besturingssysteem.¹⁰⁸

¹⁰² Bergman, M.A. (2001), "The role of the essential facilities doctrine", *The Antitrust Bulletin* 2 blz. 403-434.

¹⁰³ Zie o.a. gevoegde zaken C-241/91 en C-242/91, *Magill*, ECLI:EU:C:1995:98; zaak C-7/97, *Oscar Bronner*, ECLI:EU:C:1998:569; zaak C-418/01, *IMS Health*, ECLI:EU:C:2004:257; zaak T-201/04, *Microsoft*, ECLI:EU:T:2007:289.

¹⁰⁴ Franse mededingingsautoriteit, Besluit n°17-D-06 (*GDF Suez*), 21 maart 2017, beschikbaar op <http://www.autoritedelaconurrence.fr/pdf/avis/17d06.pdf>.

¹⁰⁵ Besluit 2015-P/K-27 van 22 september 2015 van de Belgische mededingingsautoriteit, *Stanleybet Belgium/Stanley International Betting and Sagevas/World Football Association/Samenwerkende Nevenmaatschappij Belgische PMU v. Nationale Loterij*.

¹⁰⁶ Gevoegde zaken C-241/91 en C-242/91, *Magill*, ECLI:EU:C:1995:98.

¹⁰⁷ Zaak C-418/01, *IMS Health*, ECLI:EU:C:2004:257.

¹⁰⁸ Zaak T-201/04, *Microsoft*, ECLI:EU:T:2007:289.

De toepassing van de essentiële faciliteiten-doctrine op data-gedreven markten, naast de meer traditionele markten die aan de orde kwamen in eerdere zaken, is onderwerp van discussie in de literatuur (en ook in rapporten met betrekking tot het mededingingsbeleid). Hoewel een aantal wetenschappers heeft betoogd dat data zich vanwege de niet-homogene aard en de brede beschikbaarheid ervan niet lenen voor onderwerping aan de essentiële faciliteiten-doctrine (Colangelo & Maggiolino, 2018), zijn anderen van mening dat, afhankelijk van de feiten van een zaak, weigeringen tot datadeling de toepassing van de essentiële faciliteiten-doctrine kunnen rechtvaardigen (Drexl, 2017; Graef, 2016). De behoefte aan toegang tot data als voorwaarde voor concurrerende markten in de dataeconomie wint ook terrein in beleidskringen (Crémer, De Montjoye & Schweitzer, 2019; Europese Commissie, 2020). Ondanks de aandacht voor deze problematiek in literatuur en beleidsdebatten, bestaat er geen duidelijkheid over hoe de essentiële faciliteiten-doctrine in de praktijk in concrete scenario's moet worden toegepast op data, vanwege het ontbreken van zaken in data-gedreven markten.

Aangezien het opleggen van een verplichting om diensten te verlenen verstrekkende gevolgen heeft voor de contractvrijheid en het eigendomsrecht van een onderneming met een machtspositie, is er een hoge juridische drempel ten aanzien van de essentiële faciliteiten-doctrine. Alleen in 'uitzonderlijke omstandigheden', zoals omschreven in de uitspraken in de zaken *Magill*, *Bronner*, *IMS Health* en *Microsoft*, wordt een weigering om diensten te verlenen beoordeeld als machtsmisbruik, namelijk:

- als de input of activa onmisbaar zijn voor het verlenen van de downstreamdienst,
- als weigering van toegang leidt tot uitsluiting van effectieve concurrentie op de downstreammarkt, en
- als de onderneming met een machtspositie geen objectieve rechtvaardiging heeft om toegang te weigeren.¹⁰⁹

Hoewel deze voorwaarden over het algemeen adequaat op data kunnen worden toegepast (zoals hieronder in het kader van de casussen nader zal worden toegelicht), zijn enkele algemene beperkingen vermeldenswaard.

Ten eerste is, volgens de huidige interpretatie in de jurisprudentie, een belangrijke beperking van de essentiële faciliteiten-doctrine de voorwaarde dat de onderneming met een machtspositie de toegang tot een downstreammarkt waarop deze zelf al actief is uitsluit. In eerdere zaken hebben de rechtbanken van de EU dan ook de voorwaarde tot uitsluiting van effectieve concurrentie zodanig geïnterpreteerd dat de onderneming met een machtspositie zelf al actief moest zijn op de downstreammarkt die de toegangsoeker wil betreden.¹¹⁰ Met andere woorden, de weigering om diensten te verlenen komt voort uit het feit dat de onderneming met een machtspositie op de markt 'gebruik maakt' van die positie voor de input op de markt waarop een derde wil gaan concurreren met de onderneming met een machtspositie. Dit betekent dat scenario's waarbij toegangzoekers

¹⁰⁹ Gevoegde zaken C-241/91 en C-242/91, *Magill*, ECLI:EU:C:1995:98; zaak C-7/97, *Oscar Bronner*, ECLI:EU:C:1998:569; zaak C-418/01, *IMS Health*, ECLI:EU:C:2004:257; zaak T-201/04, *Microsoft*, ECLI:EU:T:2007:289. In zaken die betrekking hebben op activa die krachtens het intellectueel eigendomsrecht worden beschermd, is de aanvullende nieuweproductvoorwaarde toegepast om te beoordelen of een weigering om een licentie te verlenen neerkomt op misbruik van machtspositie. Aan de nieuweproductvoorwaarde is voldaan wanneer een weigering om een intellectueel eigendomsrecht in licentie te geven de introductie verhindert van een nieuw product waarnaar in potentie vraag van de consument bestaat. In de zaak *Microsoft* heeft het Gerecht de normen om aan de nieuweproductvoorwaarde te voldoen echter aanzienlijk verlaagd door te stellen dat een beperking van 'technische ontwikkeling' voldoet. Zie zaak T-201/04, *Microsoft*, ECLI:EU:T:2007:289, par. 647.

¹¹⁰ Zie zaak T-504/93, *Tiercé Ladbroke v. Commissie*, ECLI:EU:T:1997:84, par. 133. De zaak betrof een weigering door organisatoren van Franse paardenrennen om aan Ladbroke, die diensten verleende bij het afsluiten van weddenschappen in België, een uitzendvergunning te verstrekken voor geluid en beeld van de Franse paardenrennen. Afgezien van het ontbreken van de onmisbaarheidsvoorwaarde, oordeelde het Gerecht dat niet was voldaan aan de voorwaarde van uitsluiting van effectieve concurrentie, omdat de organisatoren van de Franse paardenrennen geen concurrenten van Ladbroke waren op de relevante markt voor het verlenen van diensten bij het afsluiten van weddenschappen in België en er dus geen sprake was van een poging van voorbehoud van die gerelateerde markt aan zichzelf.

de input willen gebruiken als basis voor de ontwikkeling van een nog niet door de onderneming met een machtspositie aangeboden product of dienst, buiten het bestek vallen van de essentiële faciliteiten-doctrine zoals die momenteel wordt geïnterpreteerd (Drexl, 2017; Graef, 2016). Een belangrijk kenmerk van data, en Big Data in het bijzonder, is dat het potentieel ervan vaak niet kan worden voorspeld ten tijde van de verzameling. Het weigeren van toegang tot data kan daardoor de opkomst van nieuwe markten verhinderen (Rousseva, 2010).

Of het waarschijnlijk is dat voornoemd probleem zich in de praktijk voordoet staat te bezien. Er zijn redenen om aan te nemen dat een platform met een machtspositie gaarne bereid zal zijn om toegang te verlenen tot zijn data wanneer dit leidt tot nieuwe innovatieve downstreamdiensten, tenzij de nieuwe downstreamdienst het platform of een van zijn downstreamdiensten dreigt te (des)intermediëren (Van Gorp & De Bijl, 2019).

Dit brengt ons bij een ander (meer conceptueel) probleem: de voorwaarde dat de onderneming met een machtspositie zelf actief moet zijn in de downstreammarkt die de toegangszoeker wil betreden. Op traditionele op infrastructuur gebaseerde markten is deze eis passend. Op dergelijke markten is het niet logisch dat de verwerkingsverantwoordelijke van de essentiële faciliteit een downstreamdienst blokkeert die niet concurreert met de eigen downstreamdiensten. De opkomst van een nieuwe downstreamdienst heeft immers geen gevolgen voor de huidige downstreaminkomsten en genereert uitsluitend extra upstream inkomsten.

Deze redenering kan ook – maar niet noodzakelijkerwijs – gelden voor een platform. Een nieuwe downstreamdienst kan het platform dreigen te intermediëren of disintermediëren, ook als het platform zelf niet actief is op de downstreammarkt (Van Gorp & De Bijl, 2019). In dergelijke gevallen is het weigeren van toegang tot de essentiële data concurrentiebepkend (in de zin van defensief) van aard, zelfs als het platform op geen enkele downstreammarkt actief is. Een verruiming van de interpretatie van de essentiële faciliteiten-doctrine tot gevallen waar geen sprake is van ‘gebruik maken’ lijkt haalbaar, aangezien het mededingingsrecht flexibel is en de toepassing ervan kan worden aangepast aan de relevante marktomstandigheden. Van Gorp en De Bijl (2019) stellen voor om allereerst de aandacht te richten op een analyse van objectieve rechtvaardigingen. Als die tekortschieten, kan volgens Van Gorp en De Bijl de conclusie luiden dat het gedrag concurrentiebepkend van aard is, zonder dat een relevante markt moet worden gedefinieerd of een machtspositie moet worden beoordeeld.

Een tweede uitdaging is de drempel voor onmisbaarheid van de input. De niet-concurrerende aard van data, de brede beschikbaarheid van data, het aanvullende karakter van data en het vermogen om data af te leiden kunnen een dataset niet-onmisbaar maken. Het vereist echter diepgaande technologische kennis om tot een dergelijke conclusie te komen. In een aantal fusiebeschikkingen op onlinemarkten (bijvoorbeeld Google/DoubleClick en Facebook/ WhatsApp) heeft de Commissie geconcludeerd dat ‘voldoende data’ beschikbaar blijven voor derden buiten de controle van de fuserende partijen.

In haar fusiebeschikking *Google/DoubleClick* uit 2007 voerde de Commissie bijvoorbeeld aan dat de combinatie van data van Google over het zoekgedrag van gebruikers met data van DoubleClick over webbrowsergedrag van gebruikers, de gefuseerde entiteit geen concurrentievoordeel opleverde dat niet kon worden geëvenaard door concurrenten.¹¹¹ Volgens de Commissie was een dergelijke combinatie van informatie al beschikbaar voor een aantal concurrenten van Google, waaronder Microsoft en Yahoo, die allebei zoekmachines gebruikten en ook advertentiediensten aanboden. Bovendien voerde de Commissie aan dat concurrenten data of targetingdiensten konden kopen van derden waaronder portalen, andere grote webuitgevers en aanbieders van

¹¹¹ Zaak COMP/M.4731 – *Google/DoubleClick*, 11 maart 2008, par. 366.

internetdiensten.¹¹² Hoewel Google ten tijde van de goedkeuring van de fusie nog geen webbrowser aanbood, in tegenstelling tot bijvoorbeeld Microsoft, kwam daarin snel verandering. Met de huidige kennis hadden sommige van deze beschikkingen wellicht kritischer moeten zijn ten aanzien van het belang en de substitueerbaarheid van data.

De Commissie hield ook rekening met de rol van schaalgrootte bij het verzamelen van data in de zaak betreffende de *Microsoft/Yahoo*-fusie in 2010. Een van haar argumenten om de fusie goed te keuren was de mogelijkheid dat *“als de transactie, door de schaaleardeffecten, leidt tot een sterkere concurrent met een groter vermogen tot innovatie, dit Google ook zal stimuleren om zijn innovatie-inspanningen op de markt op peil te houden of zelfs op te voeren”*.¹¹³ Dit kan worden geïnterpreteerd als de EC die stelt dat Bing en Yahoo afzonderlijk onvoldoende concurrentiedruk op de markt zouden uitoefenen.¹¹⁴ Enkele jaren later, in 2017 en 2018, werd door de Commissie geconcludeerd dat Google een absolute machtspositie had verworven op de markt voor algemene zoekdiensten.¹¹⁵ Zoals we in de casus hieronder uitleggen, is een van de redenen voor de machtspositie van Google mogelijk het ongeëvenaarde volume en de verscheidenheid van zijn data met betrekking tot zoekopdrachten en klikgedrag.

Vanwege de kenmerkende eigenschappen van data en de concurrentiedynamiek van de data-economie, is gepleit voor een lagere drempel voor het verplicht stellen van toegang tot data in vergelijking met toegang tot andere input. Het argument is dat, afhankelijk van de exacte setting, een toegangsverplichting mogelijk niet dezelfde negatieve invloed op investeringen zal hebben, met name wanneer data een nevenproduct zijn van de dienst en voor het verzamelen en de opslag ervan geen speciale investeringen nodig zijn (Schweitzer, Haucap, Kerber, & Welker, 2018).¹¹⁶ Het toepassen van een lagere drempel voor de onmisbaarheidsvoorwaarde lijkt haalbaar te zijn (ongeacht de wenselijkheid ervan), wederom omdat het mededingingsrecht flexibel is en de toepassing ervan kan worden aangepast aan de relevante marktomstandigheden.

Om de grenzen te verkennen van bestaande benaderingen en interpretaties zou proactieve handhaving door mededingingsautoriteiten nodig zijn. Als eenmaal een precedent is geschapen waarmee een plicht tot datadeling op basis van de essentiële faciliteiten-doctrine kan worden opgelegd, verschaft dit ook duidelijkheid aan andere marktspelers over de reikwijdte van het mededingingsrecht bij het eisen van datatoegang. Ondanks het feit dat mededingingshandhaving (met uitzondering van fusiecontrole) achteraf en per geval plaatsvindt, kan het voorafgaand daaraan al van invloed zijn op de toekomstige strategie van marktspelers ten aanzien van de door hen verstrekte datatoegang. Een dergelijk proces kost echter tijd en zal naar alle waarschijnlijkheid onderwerp zijn van langdurige rechtszaken, zodat in specifieke gevallen verdere regelgevende maatregelen vereist kunnen zijn die buiten het bestaande kader van het mededingingsrecht vallen,

¹¹² Zaak COMP/M.4731 – *Google/DoubleClick*, 11 maart 2008, par. 269-272 en 365.

¹¹³ Zaak COMP/M.5727 – *Microsoft/Yahoo*, 18 februari 2010, par. 219.

¹¹⁴ In 2010 volgde het Amerikaanse ministerie van Justitie (DOJ) een soortgelijke redenering om akkoord te gaan met de MICROSOFT/ YAHOO!-fusie *“De transactie zal de concurrentieprestaties van Microsoft verbeteren [omdat] de gecombineerde activiteiten Microsoft zullen voorzien van een nog veel grotere bron van data dan het momenteel heeft, of zonder deze transactie waarschijnlijk zal verwerven. Deze grotere datapool kan het testen effectiever maken en dus leiden tot snellere innovatie van potentiële nieuwe zoekgerelateerde producten, veranderingen in de presentatie van zoekresultaten en betaalde zoekvermeldingen, andere veranderingen in de gebruikersinterface, en veranderingen in de zoek- of betaalde zoekalgoritmen. Deze verbeterde prestatie zou, indien gerealiseerd, een overeenkomstig grotere concurrentiedruk op de markt moeten uitoefenen”*. Zie persbericht ministerie van Justitie, Verklaring van de Antitrustdivisie van het ministerie van Justitie met betrekking tot haar besluit om haar onderzoek naar de advertentieovereenkomst tussen Microsoft Corporation en Yahoo! inzake zoeken en betaald zoeken op internet af te sluiten.

¹¹⁵ Zie zaak AT. 40099 (*Google Android*) en zaak AT.39740 — *Google Search (Shopping)*

¹¹⁶ Zie bijvoorbeeld Schweitzer et al (2018) die het volgende stellen: ‘Er zijn goede redenen om aan te nemen dat, afhankelijk van de exacte setting, de drempel om te constateren dat een weigering om data te verstrekken als misbruik moet worden beoordeeld lager kan liggen dan de drempel om misbruik te constateren in het geval van weigering om toegang te verlenen tot infrastructuur of intellectuele eigendomsrechten. Dit geldt met name indien en voor zover de weigering om toegang te verlenen betrekking heeft op data die nagenoeg incidenteel en zonder bijzondere investering worden gegenereerd.’

afhankelijk van de marktomstandigheden en het nagestreefde beleidsdoel (Graef, Tombal & De Streef, 2019). Aanvullende maatregelen kunnen bijvoorbeeld gerechtvaardigd zijn om datatoegang als innovatie-instrument in bepaalde sectoren te bevorderen, ongeacht of er sprake is van een onderliggend concurrentieprobleem of marktfalen. Wij komen hierop terug in het volgende hoofdstuk.

5 Beoordeling van mogelijke verplichtingen tot datadeling in vijf casussen

5.1 Inleiding

Focus en onderzoeksvragen met betrekking tot de casussen

Zoals vermeld in de inleiding tot dit rapport richt dit onderzoek zich op de impact van datadeling op productiviteit, innovatie, concurrentie, keuzevrijheid en beveiliging/privacy. Er kunnen ook andere publieke belangen in het geding zijn, met mogelijk onderlinge relaties¹¹⁷. Wij richten ons echter op het publieke belang dat rechtstreeks baat kan hebben bij, of kan worden geschaad door, datadelingspraktijken en -beleid, of de afwezigheid daarvan.

De Nederlandse overheid heeft in haar visie op data de volgende criteria gedefinieerd om te bepalen wanneer een verplichting tot datadeling moet worden opgelegd:

1. Datadeling draagt bij aan het publieke belang (met name in verband met concurrentie, innovatie en keuzevrijheid).
2. De data zijn niet eenvoudig te (re)produceren of te verzamelen door andere partijen dan de partij (partijen) waarvoor een verplichting tot datadeling wordt overwogen.
3. Het lukt belanghebbenden niet of niet voldoende om op vrijwillige basis passende afspraken te maken over toegang tot de data.
4. Stimulansen voor investeringen in (het produceren en/of verzamelen van) data worden niet substantieel verminderd door de invoering van een verplichting tot datadeling.
5. De voordelen voor het publieke belang wegen voldoende op tegen mogelijke negatieve neveneffecten (met name in verband met concurrentie, innovatie, en privacy- en beveiligingsrisico's).

In dit hoofdstuk onderzoeken we in vijf casussen of aan deze criteria wordt voldaan, t.w.

1) zoekmachines, 2) sociale-medianetwerken, 3) betaaldiensten, 4) administratieve bedrijfssoftware en 5) een generieke situatie van verticale uitsluiting via het blokkeren van datatoegang. In aanvulling hierop worden in het kader van de casussen de huidige wettelijke bepalingen voor uitwisseling van data die relevant zijn voor de desbetreffende casussen besproken, alsmede juridische en technische kwesties die relevant zijn bij het opleggen van verplichtingen tot datadeling.

In de casus inzake zoekmachines onderzoeken we of horizontale uitwisseling van data over zoektermen en klikgedrag zinvol is. In de casus inzake sociale-medianetwerken onderzoeken we of het wenselijk is om sociale-mediaplatformen te verplichten om sociale data uit te wisselen. In de casus inzake betaaldiensten onderzoeken we het effect van de verplichting tot datadeling onder PSD2 op de concurrentiepositie van banken en of het nodig is om ook een verplichting tot datadeling op te leggen aan toegangzoekers. In de casus inzake administratieve bedrijfssoftware onderzoeken we in hoeverre een verplichting tot dataportabiliteit vendor lock-in kan verminderen.

¹¹⁷ Het bevorderen van concurrentie en verminderen van lock-in gaan bijvoorbeeld met elkaar gepaard en dragen tegelijkertijd bij aan bredere consumentenbelangen, zoals keuzevrijheid en de betaalbaarheid en kwaliteit van de dienstverlening. Hoe groter platforms worden, hoe meer ze een poortwachtersfunctie vervullen en hoe belangrijker de integriteit en de continuïteit van het platform worden. Evenzo geldt dat hoe afhankelijker consumenten worden van platforms voor de verspreiding van nieuws en media, hoe verantwoordelijker de platforms zijn (of zouden moeten zijn) voor pluralisme in het aanbod van nieuws en media, en hoe meer behoefte er is aan bescherming van consumenten tegen ongepaste beïnvloeding door overheden, door het platform zelf en door andere gebruikers van de platforms.

Ten slotte bespreken we een generieke situatie waarin een platform een deelnemer uitsluit van datatoegang.

De besprekingen van de casussen vatten de bevindingen en conclusies samen van interviews en deskresearch.

5.2 Casus: zoekmachines

Zoekmachines (zoals Bing en Google) koppelen informatiezoekers aan beschikbare informatie op internet. Zoektechnologie is gebaseerd op de volgende bouwstenen:

- **Webcrawlers** zijn robots die de inhoud van webpagina's registreren en vervolgens verwerken in een **index** (zoals een index aan het eind van een boek), met een vermelding van elk woord op elke geïndexeerde webpagina.
- Een **zoekalgoritme** dat de index verkent om websites aan zoekopdrachten te koppelen.
- Zoekalgoritmen leren door analyse de correlatie tussen door gebruikers ingevoerde **zoekopdrachten** en hun **klikgedrag**.
- Het algoritme kan (indien relevant) bij de interpretatie van zoekopdrachten ook **contextuele data** in aanmerking nemen, zoals zoekgeschiedenis, gebruikt apparaat, locatie enz.
- **Profieldata** (waaronder voorkeuren, leeftijd, nationaliteit enzovoort) worden vaak niet ingevoerd in het zoekalgoritme, vanwege het belang van objectiviteit van organische zoekresultaten. Profieldata (tezamen met de andere data) worden wel ingevoerd in het algoritme dat de geadverteerde zoekresultaten matcht om te worden weergegeven met de organische zoekresultaten (vaak met een kennisgeving).

Zoekopdrachten en klikgedrag zijn gedragsdata die worden toegevoegd aan zoekalgoritmen. Bovendien geven zoekopdrachten en klikgedrag crawlers een signaal om inhoud te gaan indexeren waarin gebruikers geïnteresseerd lijken te raken. Deze feedbackloops leiden tot leereffecten die aanleiding geven tot een voordeel van de first-mover en een risico op kanteling van de markt, waardoor concurrentie en innovatie kunnen worden geschaad. Een mogelijke remedie is horizontale uitwisseling van (gedrags)zoekdata tussen zoekmachines.

5.2.1 *Factoren voor verplichtingen in de Nederlandse visie op datadeling*

De zoekmarkt is sterk geconcentreerd, waarbij Google rond 90% van de markt bedient. Een hypothese is dat concurrenten hun achterstand niet kunnen wegwerken in de afwezigheid van datadeling. Als deze hypothese juist blijkt te zijn, rijst de cruciale vraag of verplichte datadeling toetreding kan vergemakkelijken en concurrentie en innovatie kan herstellen. Een aanvullende vraag betreft de implicaties voor privacy en innovatiestimulansen.

Draagt datadeling bij aan een publiek belang zoals concurrentie, keuzevrijheid of innovatie?

Op basis van de checklist in sectie 2.2 zijn er diverse redenen om aan te nemen dat beperkte toegang tot gedragsdata een belemmering vormt voor concurrentie op de zoekmarkt, en dat horizontale uitwisseling van die data het proces en het resultaat van concurrentie en innovatie kunnen ondersteunen.

- Zoek- en klikgedrag zijn vrij **uniek** en staan in hoge mate **exclusief** ter beschikking van Google. De exclusiviteit is het gevolg van het feit dat Google een marktaandeel van 90% heeft in zoekdiensten, en omdat de Google Chrome-browser de leidende positie van Microsoft op de browsermarkt heeft overgenomen.

- Er zijn **directe en indirecte datagerelateerde externe effecten**. Indirect: zoek- en klikgedrag van persoon A helpt bij het interpreteren van zoekopdrachten van persoon B. Direct: eerder zoekgedrag van persoon A verbetert de zoekresultaten voor persoon A.
- Deze externe effecten zorgen voor een voordeel van de first-mover door **toenemende schaalopbrengsten**. Hoewel de leercurve van een zoekalgoritme reeds afvlakt bij relatief kleine aantallen zoekopdrachten¹¹⁸, wordt het belang van schaalgrootte van gebruikers bepaald door het feit dat een aanzienlijk deel van de zoekopdrachten (ongeveer 25%) 'uitzonderlijk' is; concurrenten kunnen wellicht niet de voor dergelijke opdrachten benodigde hoeveelheid data realiseren.
- Zoekdata **devalueren** in een vrij hoog tempo. Webindexen worden regelmatig geüpdatet op basis van nieuwe zoekopdrachten die crawlers instrueren om zich op specifieke inhoud te richten, wat impliceert dat de houdbaarheid van data beperkt is. Dit draagt bij aan het belang van de schaal van aantallen gebruikers.
- Het is essentieel voor de kwaliteit van de zoekdienst dat **verkregen inzichten onmiddellijk worden toegepast** bij het updaten van de webindex, met name voor wat betreft actuele (en onverwachte) gebeurtenissen die in korte tijd veel aandacht trekken¹¹⁹. Ook in dit verband is de schaal van aantallen gebruikers van belang.

Daartegenover staat dat leereffecten maar één reden zijn voor een concurrentievoordeel bij online zoeken, wat erop zou wijzen dat **(enkel) het delen van data wellicht niet voldoende is** om meer concurrentie te bevorderen. Er zijn andere grote belemmeringen voor opschaling: i) hoge verzonken investeringen, ii) indirecte netwerkeffecten tussen zoekers en adverteerders, en iii) starheid van gebruikers, in combinatie met koppel- en bundelingsstrategieën.

- Tot de verzonken investeringen behoren investeringen in innovatieve activiteiten, zoals het efficiënter en frequenter updaten van de index¹²⁰, het ontwikkelen van een aparte index specifiek voor mobiel zoeken¹²¹, innovaties en testen (door mensen) van zoekalgoritmen¹²², en investeringen om laadtijden te versnellen¹²³. Daarbij komt dat een zoekdienst taalafhankelijk is, waardoor veel van de investeringen in innovatie en testen voor iedere extra taal gedupliceerd moeten worden.
- Met betrekking tot andere bepalende factoren voor de leidende positie van Google (zoals indirecte netwerkeffecten met adverteerders en het koppelen en bundelen van verschillende diensten) beweren sommige experts dat, zelfs als datadeling het mogelijk zou maken voor concurrenten om de kwaliteit van de marktleider te evenaren, dit mogelijk niet zou leiden tot een groei van hun marktaandeel zolang deze andere factoren blijven bestaan.

Al met al volgt uit de eerste vijf bovengenoemde punten dat data de positie van Google versterken. Het is echter niet evident dat data de belangrijkste bepalende factor zijn voor die positie. Een natuurlijk experiment op de Russische markt bevestigt dit. Omdat de Russische taal sterk verschilt van het Engels, had Google een groot datanadeel ten opzichte van Yandex toen het in 2006 de Russische markt betrad¹²⁴. In de afgelopen tien jaar zijn Google en Yandex om beurten marktleider

¹¹⁸ He et al. (2017) constateren dat de leercurve van een zoekalgoritme al afvlakt bij kleine hoeveelheden zoekopdrachten. De schaal die nodig is voor algoritmen om te leren is dus gemakkelijk haalbaar voor de meeste zoekopdrachten.

¹¹⁹ Dit geldt in het bijzonder voor zoekacties die verband houden met actuele gebeurtenissen die in korte tijd veel aandacht trekken (bijv. een nucleair ongeluk, een terroristische aanval, een nieuw Olympisch record, een koninklijke geboorte enz.).

¹²⁰ <https://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>
<https://www.brafton.co.uk/news/seo-case-study-sites-see-more-pages-indexed-by-google-than-bing-even-post-panda-800527170/>

¹²¹ <https://webmasters.googleblog.com/2016/11/mobile-first-indexing.html>
<http://www.thesempost.com/bing-search-will-not-mobile-first-search-index/>

¹²² <https://searchengineland.com/bing-search-quality-rating-guidelines-130592>

¹²³ Tijdens de interviews gaven respondenten (behalve Google) aan dat 'laadtijd' een belangrijk kwaliteitskenmerk is in de ogen van eindgebruikers. Het geval wil dat Google anderen hierin overtreft. <https://www.shoutmeloud.com/google-started-ranking-websites-based-on-load-time-and-speed.html>

¹²⁴ <https://www.cnbc.com/2019/01/18/yandex-is-beating-google-in-russia.html>
<https://www.nytimes.com/2006/12/18/technology/18google.html>

geweest op de Russische zoekmarkt. Het feit dat er geen sprake is geweest van een kanteling in de Russische markt impliceert dat er andere bepalende factoren voor concurrentieposities (naast leereffecten) zijn op de zoekmarkt. In 2009 introduceerde Yandex bijvoorbeeld een nieuw algoritme voor machine learning (MatrixNet), waarmee Yandex begin jaren '10 van deze eeuw erin slaagde om terrein te heroveren op Google¹²⁵. Ook het besluit uit 2017 van de Russische mededingingsautoriteit om Google te dwingen om de Chrome-browser en -zoekmachine los te koppelen van het Android-besturingssysteem leidde tot een snelle daling van het marktaandeel van Google in mobiel zoeken.

Met betrekking tot verplichte datadeling stellen Argenton & Prüfer (2012) een algemene, wederkerige verplichting voor zoekmachines voor. Zij stellen dat asymmetrische verplichtingen het speelveld voor de first-mover verstoren en een kanteling van de markt veroorzaken ten gunste van de partij die het tweede grootst is. Wederkerigheid zou dan in eerste instantie alleen gelden tussen de grootste speler en toegangszoekers, waarbij de grootste speler nog steeds een datavoordeel zou behouden¹²⁶. Andere spelers zouden deze voorsprong kunnen verkleinen door hun data ook op wederkerige basis uit te wisselen.¹²⁷ Hiervoor is geen verplichting nodig, omdat alle partijen er baat bij hebben, maar een expliciete uitzondering voor dergelijke samenwerking zou moeten worden opgenomen in Artikel 101 van het VWEU, uit hoofde waarvan horizontale samenwerking tussen concurrenten verboden wordt.

Zijn de data eenvoudig te (re)produceren of te verzamelen door anderen?

Zoals hierboven vermeld, is het zoek- en klikgedrag vrij uniek en grotendeels exclusief voor Google vanwege zijn machtspositie op de browsermarkt. Dit was tien jaar geleden anders, toen Microsoft zoekdata kon waarnemen via Internet Explorer. In deze situatie kwam verandering toen de Google Chrome-browser de leidende positie van Microsoft in de browsermarkt overnam. De recente uitspraak van de Europese Commissie in de zaak Google Android, waarmee Google wordt verplicht om zijn mobiele browser los te koppelen van Android, kan de concurrentie op de Europese browsermarkt vergroten en daarmee zoekdata minder uniek en exclusief maken.

Lukt het niet of niet voldoende om op vrijwillige basis afspraken te maken over datatoegang?

Data met betrekking tot zoekopdrachten en klikgedrag worden niet uitgewisseld met derden. Tijdens interviews toonden nichespelers geen interesse in toegang tot zoekdata en gaven grotere spelers aan niet snel bereid te zijn tot grootschalige uitwisseling van zoekdata, uit privacyoverwegingen en vanwege risico's op indirecte openbaarmaking van bedrijfsgeheimen door het risico op reverse engineering.

Zoekresultaten worden wel uitgewisseld via commerciële overeenkomsten met nichespelers, in aanvulling op de dienst van de toegangsverlener (bijvoorbeeld het bieden van privacy door DuckDuckGo bij Bing en Startpage bij Google). Het risico op reverse engineering lijkt klein te zijn, omdat toegangszoekers niet over de middelen beschikken om dergelijk onderzoek te financieren. Bovendien wordt reverse engineering verboden in de overeenkomsten.

Verlaagt verplichte datadeling de prikkels voor investeringen in data substantieel?

Gedragsdata worden grotendeels verstrekt en waargenomen, wat niet direct gepaard gaat met investeringen in data-analyse. Uitwisseling ervan is niet rechtstreeks van invloed op het verzamelen

¹²⁵ <https://www.raamoprusland.nl/dossiers/economie/786-how-even-yandex-is-being-brought-under-control>

¹²⁶ Als A verplicht wordt om data uit te wisselen met B en C op wederkerige basis, dan zou A nog steeds een grotere dataset hebben (van A+B+C) dan B en C (respectievelijk van B+A en van C+A).

¹²⁷ Als B en C op vrijwillige basis besluiten om hun data ook uit te wisselen, dan hebben alle partijen toegang tot de grotere dataset van A+B+C.

van deze data. Datadeling kan echter wel een impact hebben op investeringsprikkels in het algemeen (in webindexen, algoritmen enzovoort).

Op basis van een theoretisch model betogen Argenton & Prüfer (2012) dat wederkerige datadeling tussen zoekproviders innovatie, kwaliteit, consumentensurplus en het algemeen welzijn stimuleert, omdat als gevolg ervan *“geen enkele zoekmachine op zijn lauweren kan rusten, aangezien de enige manier om zijn marktaandeel en winsten te behouden is om alle efficiëntiewinst die voortkomt uit de exploitatie van externe netwerkfactoren in betere kwaliteit te investeren.”* Deze verklaring gaat ervan uit dat datadeling leidt tot evenwichtiger concurrentie, wat in de echte wereld misschien niet het geval is (zoals hierboven betoogd). Zelfs als dit wel van toepassing is, dan is de conclusie met betrekking tot de investeringsprikkels mogelijk niet correct wanneer toegang tot zoekdata reverse engineering van het algoritme of de zoekresultaten mogelijk maakt (zie hieronder).

Wat zijn de negatieve gevolgen op het vlak van publieke belangen?

Een risico op reverse engineering kwam aan het licht in 2011, toen Google een experiment uitvoerde om vermoedens te bevestigen dat Microsoft de zoekresultaten van Google reconstrueerde met data die werden opgehaald via Internet Explorer¹²⁸. In het kader van het experiment bewerkte Google zijn index door niet-bestaande woorden te koppelen aan willekeurige websites. Vervolgens gaf Google 20 medewerkers opdracht om de Googlewebsite regelmatig te bezoeken via Internet Explorer, de niet-bestaande woorden op te zoeken en op de resultaten te klikken. Google beweerde dat de ‘nepresultaten’ na een paar weken begonnen te verschijnen in de zoekresultaten van Bing. Microsoft bevestigde destijds dat het zoek- en klikdata verzamelde via zijn browser en werkbalk, maar ontkende dat het de resultaten van Google kopieerde¹²⁹. Het is echter normaal dat zoekalgoritmen combinaties van zoek- en klikdata gebruiken om te leren. Voor niet vaak voorkomende termen zijn dit wellicht de enige signalen die het algoritme en de webcrawlers opvangen. Als crawlers de inhoud van de website niet verifiëren, dan koppelen de index en het algoritme een zoekterm automatisch aan een website. Dit komt neer op reverse engineering van een index en zoekresultaten.

Daarnaast gaven grotere spelers aan beducht te zijn voor grootschalige uitwisseling van zoekdata vanwege privacyrisico's. Toch zijn die risico's relatief klein wanneer gedragsdata worden geanonimiseerd, mits ze niet kunnen worden gecombineerd met contextuele data die deanonimisering mogelijk maken. Geanonimiseerde data met betrekking tot zoekopdrachten en bijbehorend klikgedrag behouden vrij veel waarde voor het trainen van zoekalgoritmen.

5.2.2 Huidige wettelijke bepalingen voor verplichte datadeling

De AVG is van toepassing, maar is mogelijk niet effectief

Indien data met betrekking tot zoekopdrachten en klikgedrag kunnen worden herleid naar een individuele persoon, bijvoorbeeld door ze te koppelen aan een IP-adres of cookie, of door ze op te slaan in een profiel wanneer gebruikers inloggen bij de zoekdienst, gelden ze als persoonsdata. Dit betekent dat de AVG-bepalingen van toepassing zijn, inclusief het recht op dataportabiliteit¹³⁰. Dit recht kan echter alleen worden gebruikt om de zoekgeschiedenis van gebruikers één voor één en

¹²⁸ <https://googleblog.blogspot.com/2011/02/microsofts-bing-uses-google-search.html>

¹²⁹ <https://www.businessinsider.com/bing-is-cribbing-from-google-search-results-2011-2>

¹³⁰ Bovendien zou het klikgedrag van een gebruiker worden aangemerkt als ‘waargenomen data’, waarvan de Groep Databescherming Artikel 29 heeft aangevoerd dat deze vallen onder het recht op dataportabiliteit op grond van de AVG. De Groep Databescherming Artikel 29 verwijst in dit verband expliciet naar een zoekgeschiedenis als een vorm van waargenomen data die binnen het toepassingsgebied van Artikel 20 van de AVG vallen: ‘Waargenomen data verstrekt door betrokkene voortkomend uit het gebruik van de dienst of het apparaat [...] kunnen bijvoorbeeld de zoekgeschiedenis van een persoon bevatten’ (Groep Databescherming Artikel 29, 2017). Wij merken hierbij op dat de Groep Databescherming Artikel 29 geen rekening heeft gehouden met het risico op reverse engineering gerelateerd aan de uitwisseling van gedragsdata.

op hun verzoek over te dragen. Dit betekent dat de effectiviteit van dataportabiliteit als instrument om datadeling verplicht te stellen afhankelijk is van hoe actief er door personen een beroep op wordt gedaan. Een tweede beperking is dat het nog steeds onduidelijk is in hoeverre het recht op dataportabiliteit onder de AVG betrekking heeft op real-time continue datadeling tussen verwerkingsverantwoordelijken. Real-time portabiliteit lijkt effectiever te zijn om knelpunten op te lossen dan eenmalige portabiliteit. Een derde beperking is dat de AVG geen wederkerige datadeling kan afdwingen; dit kan alleen op verzoek van het individu.

Essentiële faciliteiten-doctrine kan van toepassing zijn, met creatieve interpretatie

Een mogelijkheid om datadeling met betrekking tot zoekopdrachten en klikgedrag verplicht te stellen is om een beroep te doen op de essentiële faciliteiten-doctrine onder het mededingingsrecht. Een weigering om data inzake zoekopdrachten en klikgedrag uit te wisselen kan als misbruik van machtspositie gelden in de zin van Artikel 102 van het VWEU wanneer het bedrijf dat de toegang tot data controleert een machtspositie heeft en er sprake is van vier 'uitzonderlijke omstandigheden': onmisbaarheid, uitsluiting van effectieve concurrentie, verhindering van het op de markt brengen van een nieuw product, en het ontbreken van een objectieve rechtvaardiging¹³¹. Echter, alleen zoekmachines met een machtspositie zouden dan op grond van het mededingingsrecht kunnen worden verplicht om informatie uit te wisselen, wat impliceert dat een uitwisselingsverplichting niet wederkerig kan zijn. Het loont echter de moeite om te onderzoeken of wederkerigheid kan dienen als een vorm van 'eerlijke en redelijke' compensatie¹³².

5.2.3 Belemmeringen/voorwaarden voor verplichte datadeling

Juridisch

Krachtens het mededingingsrecht moeten de data die voor een verplichting in aanmerking komen een essentiële voorziening zijn. Dit zou kunnen werken, en de criteria die we hebben toegepast bij het beantwoorden van de eerste vraag in paragraaf 5.2.1 kunnen hierbij helpen. Tot op heden is er echter geen precedent. Een nadeel van het mededingingsrecht is bovendien dat beslissingen vaak lang op zich laten wachten. Mogelijk is er tegen die tijd onherstelbare schade aangericht. Aanvullende maatregelen vooraf kunnen helpen om sneller te reageren. Deze optie moet hoe dan ook worden onderzocht door de Europese Commissie, die momenteel werkt aan nieuwe kaders voor ex ante regelgeving voor digitale poortwachters.

Opgemerkt wordt dat gedragsdata informatie kunnen bevatten die wordt beschermd onder de databankrichtlijn of de richtlijn bedrijfsgeheimen. Dergelijke data beschrijven interacties met een gebruikersinterface die resultaten toont van de (investeringen in de) structuur van databanken (bijvoorbeeld een webindex) en algoritmen. Analyse van deze data kan dus informatie vrijgeven die onder deze richtlijnen wordt beschermd. Voor zover er geen sprake kan zijn van reverse engineering zou dit geen probleem moeten zijn, maar er is nog geen precedent geschapen.

¹³¹ Hoewel er duidelijke juridische precedents zijn die duiden op een machtspositie van Google op de markt voor online zoeken (zie de beoordeling van de Commissie in de besluiten inzake Google Shopping en Google Android), zijn er geen duidelijke juridische precedents met betrekking tot de kwestie van onmisbaarheid van zoekdata die in het bezit zijn van één partij. In eerdere fusiezaken (bijvoorbeeld Google/DoubleClick) voerde de Europese Commissie aan dat de data van Google inzake het zoekgedrag van gebruikers in combinatie met de data van DoubleClick inzake het browsergedrag van gebruikers de gefuseerde entiteit niet een zodanig concurrentievoordeel verschafte dat dit niet door concurrenten kon worden geëvenaard. Hoewel dit misschien destijds het geval was, merken wij op dat dit vandaag de dag mogelijk niet meer zo is. Verder merken we op dat het risico op reverse engineering kan worden beschouwd als een objectieve rechtvaardiging om niet uit te wisselen.

¹³² In dat geval zou het mededingingsrecht alsnog een juridische basis kunnen bieden om toegang af te dwingen. Wederkerigheid zou dan in eerste instantie alleen gelden tussen de grootste speler en toegangszoekers, waarbij de grootste speler nog steeds een datavoordeel zou behouden. De andere spelers zouden deze voorsprong echter kunnen verkleinen door hun data ook op wederkerige basis uit te wisselen. Hiervoor zou geen verplichting nodig zijn, omdat alle spelers hier baat bij hebben, maar een expliciete uitzondering voor deze samenwerking moet worden opgenomen in Artikel 101 van het VWEU, dat horizontale samenwerking tussen concurrenten verbiedt.

Technisch

Horizontale datadeling vereist een technische opzet die bestaat uit API's en standaardisatie van databankstructuren. Dit betekent dat partijen gezamenlijk technische overeenkomsten moeten sluiten, eventueel onder toezicht van een overheidsinstantie. Indien alleen de grootste partij verantwoordelijk wordt gesteld (of de leiding neemt), kan dit proces worden belemmerd. Het is logischer om de andere partijen de standaard te laten ontwikkelen. Maar het is ook mogelijk dat de kleinere partijen zich minder bewust zijn van de impact van standaardisatie op innovatie door de grootste partij. Daarom werkt sectorale samenwerking onder toezicht van een bevoegde autoriteit waarschijnlijk het beste om een technische infrastructuur voor datadeling tot stand te brengen.

5.3 Casus: sociale netwerken

Een sociaal netwerk is een online platform waar gebruikers contact leggen en communiceren met vrienden, familie of andere gebruikers die bepaalde interesses of professionele achtergronden delen. Voorbeelden van sociale netwerken zijn Facebook en de sport-tracking-app Strava.

Sociale netwerken zijn bedoeld om de betrokkenheid van gebruikers bij andere gebruikers, apps van derden en websites te maximaliseren. Directe en indirecte netwerkeffecten stimuleren de groei van het platform. Zoals Facebook toont, worden inkomsten vaak gegenereerd via gerichte advertentiediensten, waarvoor **profieldata** (bijvoorbeeld een beschrijving van iemands voorkeuren) van essentieel belang zijn. Inkomsten kunnen ook worden gegenereerd via (premium)abbonementen, zoals bij Strava. Door gebruikers zelf gegenereerde inhoud, sociale data en gedragsdata vormen een continue input om de gebruikerservaring te verbeteren en om te leren en te innoveren. **Sociale data**, zoals contactlijsten, social graphs en communicatiedata, zijn van essentieel belang voor het faciliteren van gebruikersinteracties.

Sociale netwerken wisselen vaak data uit met apps en websites van derden om het ecosysteem te verrijken en betrokkenheid van gebruikers te stimuleren.

5.3.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling

Er zijn diverse socialemediadiensten. Veel mensen zijn geabonneerd op meerdere socialemediadiensten. Facebook (inclusief Instagram en WhatsApp) is echter uitzonderlijk groot en beheert communicatiekanalen die zo langzamerhand misschien onmisbaar zijn geworden. Een hypothese is dat bij gebrek aan datadeling het niet mogelijk is voor concurrenten om gelijke tred te houden met Facebook, wat schadelijk is voor concurrentie, innovatie en keuzevrijheid. Als deze hypothese juist blijkt te zijn, rijst de cruciale vraag of verplichte datadeling de toegang tot de markt kan vergemakkelijken en concurrentie en innovatie kan herstellen. Een bijkomende vraag is wat dit betekent voor privacy en innovatiestimulansen. Hieronder bespreken we deze problematiek in samenhang met de factoren die deel uitmaken van de overheidsvisie op datadeling tussen bedrijven.

Draagt datadeling bij aan een publiek belang zoals concurrentie, keuzevrijheid of innovatie?

Horizontale datadeling

Er zijn diverse redenen waarom horizontale datadeling het concurrentieproces en -resultaat kan ondersteunen:

- In de ogen van eindgebruikers kan door gebruikers gegenereerde inhoud die is geüpload door vrienden en relaties worden beschouwd als **uniek en exclusief** beschikbaar via het sociale netwerk. Beëindiging van het account zou verlies van toegang tot die data betekenen. Er is dus sprake van een zekere mate van **user lock-in** omdat concurrerende netwerken wellicht niet in

staat zijn tot exacte duplicatie van de door gebruikers gegenereerde inhoud van de contacten van een gebruiker.

- Datagebruik wordt in verband gebracht met **netwerkeffecten** (die toegang tot de markt kunnen belemmeren). Social graph data houden verband met directe netwerkeffecten. De essentie van sociale netwerken zit in de functionaliteit om contact te leggen met andere gebruikers (en deze te volgen). Daarom helpen data over of van een bepaalde gebruiker rechtstreeks om de sociaalnetwerkdienst voor anderen te verbeteren. Profieldata houden verband met indirecte netwerkeffecten, aangezien ze worden gebruikt om consumenten te koppelen aan adverteerders, providers en app-ontwikkelaars.
- In zekere (en mogelijk hoge) mate bieden data een voordeel van de first-mover door middel van **toenemende schaalopbrengsten**. Sociale data ondersteunen de waarde van het grotere aantal interacties wanneer meer gebruikers zich aansluiten bij een sociaal netwerk. Deze waarde neemt exponentieel toe, maar op een gegeven moment loopt de relevantie van meer (potentiële) aansluitingen terug, waardoor uiteindelijk de opbrengsten gaan afnemen. Het keerpunt kan echter ver in het verleden liggen, wat het voor nieuwkomers moeilijk maakt om de achterstand in te halen. Bovendien bestaat de mogelijkheid dat het gevestigde bedrijf ernaar streeft dit uit te stellen door continu te blijven investeren in infrastructuur en algoritmen, en daarmee de mogelijkheden voor relevante interacties te vergroten. Iets soortgelijks doet zich voor bij profieldata. Meer profieldata zijn meestal beter voor sociale netwerken, vooral wanneer netwerken voortdurend nieuwe functies toevoegen om de ecosysteemervaring te verbeteren. Ook hier schuift dit streven het punt voor zich uit waarop toenemende opbrengsten van data veranderen in afnemende opbrengsten.

Anderzijds zijn er diverse factoren die erop wijzen dat concurrentie niet afhankelijk is van datadeling:

- Bestaande social graph data zijn **niet exclusief en uniek**, aangezien min of meer vergelijkbare informatie elders kan worden verkregen (zie ook hieronder). Toegang tot iemands sociale data maakt zijn/haar vrienden echter nog geen lid van een platform.
- Sociale data zijn relatief statisch, maar dit geldt niet voor profieldata. Voor sociale netwerken die inkomsten genereren uit advertenties, is het van cruciaal belang om gebruikers betrokken te houden, waardoor gebruikersactiviteit en -gedragsdata worden gegenereerd, die in profielen worden verwerkt. Sommige profieldata, bijvoorbeeld data die worden gebruikt voor direct targeten van consumenten, kunnen **snell in waarde dalen**. Dit vermindert het voordeel van het voordeel van de first-mover.
- Data vormen één aspect van de concurrentiepositie van een sociaal netwerk, de **innovatieve merites** van het bedrijf is een ander. Bij bedrijfsmodellen die de aandacht van gebruikers willen trekken, is het niet alleen van belang te voorspellen in welke mate inhoud aantrekkelijk is voor bepaalde gebruikers. Het is net zo belangrijk om ervoor te zorgen dat gebruikers actief en aan hun scherm "gekluisterd" blijven. De functies en de gebruikersinterface zijn hiervoor grotendeels verantwoordelijk; de effectiviteit daarvan om gebruikersbetrokkenheid te stimuleren hangt in hoge mate af van menselijke creativiteit en innovatie.
- Aangezien sociale data (en in mindere mate sommige profieldata) vrij statisch zijn, en omdat bij gebruikers vaak sprake is van multi-homing, hebben concurrenten een bepaalde hoeveelheid tijd om hun databank op te bouwen door gebruikers aan te trekken met nieuwe en onderscheidende functies, en vervolgens netwerkeffecten in gang te zetten ten behoeve van groei.
- Een aandachtspunt is dat innovatieve functies en gebruikersinterfaces mogelijk heel goed te imiteren zijn. Zo wisten Snapchat en TikTok via onderscheidende functionaliteiten jonge Facebookgebruikers over te halen om zich ook bij hun platform aan te sluiten. Een gevestigd bedrijf kan dergelijke nieuwe functies vervolgens naar zijn eigen platform kopiëren. Dit weerhoudt consumenten er misschien niet van om zich bij een nieuw sociaal netwerk aan te

sluiten, maar het beperkt wel de tijd die op het nieuwe platform wordt doorgebracht, omdat dezelfde functies nu ook beschikbaar zijn op het platform van het gevestigde bedrijf. Datadeling heeft geen invloed op deze dynamiek.

Data zijn al met al van groot belang voor een sociaal netwerk: om een gebruikersbestand op te bouwen, om gebruikersbetrokkenheid te vergroten, om inkomsten te genereren door de aandacht van gebruikers te trekken en te verkopen, en om het voortgaande innovatieproces te voeden (bijvoorbeeld door middel van hoogfrequente A/B-testen). Uit de eerste drie bovengenoemde punten kan worden afgeleid dat data de leidende positie van een sociaal netwerk mede versterken. De tweede reeks punten geeft echter een ander beeld, waaruit niet direct blijkt dat data de belangrijkste motor zijn voor een machtspositie. Daarom kunnen nieuwkomers, met een strategie die is gebaseerd op nieuwe en onderscheidende functies, voet aan de grond krijgen en zelfs een existentiële bedreiging vormen voor gevestigde bedrijven.

We verwijzen daarnaast naar sectie 3.2.5, en merken op dat de uitwisseling van profieldata tussen platforms die advertentieruimte bieden, de onderlinge concurrentiemethode fundamenteel kan veranderen. Platforms moeten dan andere manieren vinden om zich te onderscheiden, wat van invloed kan zijn op de aard van innovatie. De algehele impact voor consumenten is onzeker.

Verticale datadeling

Interviewpartners bevestigden vanuit een zakelijk perspectief dat sociale netwerken vaak datadeelovereenkomsten hebben met downstreamdienstverleners, die het platform gebruiken om in contact te komen met - en diensten aan te bieden aan - consumenten. Innovaties door externe ontwikkelaars en downstreamdienstverleners ondersteunen de groei van nieuwe bedrijfsmodellen, doordat partijen de data van het platform gebruiken. Het platform zelf verhoogt de waarde van zijn ecosysteem, wat zijn (mogelijk dominante) concurrentiepositie versterkt. Op een gegeven moment kunnen downstreampartijen echter een gedeeltelijke of volledige vervanging van het sociale netwerk worden, waarmee de positie van het platform wordt ondermijnd. Verticale datadeling biedt de consument dus variatie en innovatie, alsook een potentieel voor horizontale concurrentie.

Zijn de data eenvoudig te (re)produceren of te verzamelen door anderen?

De informatie over de bestaande social graph van een persoon (opgeslagen in een gevestigd sociaal netwerk) kan grotendeels elders worden verkregen, zoals werd bevestigd door interviewpartners. Nieuwkomers kunnen hun eigen sociale dataset opbouwen: met een strategie op basis van onderscheidende functionaliteiten kunnen ze klanten aantrekken en hun eigen social graphs creëren. Bovendien zijn er andere manieren om sociale data sneller te verzamelen, zoals toegang tot digitale adresboeken van gebruikers. Verstrekte profieldata zijn relatief eenvoudig te verkrijgen, met name via het aanmeldingsproces van nieuwe klanten. Dit komt in de praktijk veel voor, zoals ook bleek uit interviews. Data over voorkeuren worden echter vaak afgeleid uit sociale en gedragsdata. De mate waarin deze profieldata repliceerbaar zijn hangt af van de beschikbaarheid van onderliggende, ruwe data.

Lukt het niet of niet voldoende om op vrijwillige basis afspraken te maken over datatoegang?

Horizontale datadeling

Sociale netwerken zijn wellicht niet bereid om deel te nemen aan horizontale uitwisseling vanwege de concurrentiedreiging die ontstaat als een vervangend platform wordt ondersteund. Tot voor kort verbood Facebook via zijn voorwaarden voor datatoegang uitdrukkelijk nauwe banden met concurrerende sociale netwerken (of apps die zijn kernfunctionaliteit reproduceren). In 2018 schafte het bedrijf dit beleid af. Toen Relive, een app die een animerende kaart maakt op basis van locatiedata, foto's en video's, de mogelijkheid introduceerde om vrienden te volgen, beëindigde

Strava zijn datadeelovereenkomst met Relive (dit voorbeeld en de verwijzingen kunnen worden aangetroffen in de casus inzake verticale uitsluiting).

Verticale datadeling

Sociale netwerken faciliteren doorgaans verticale datadeling met downstreampartijen. Dergelijke samenwerkingsverbanden vergroten de waarde van hun ecosystemen. Derden gebruiken datadeelovereenkomsten met sociale netwerken om de groei van hun eigen bedrijf te verwezenlijken.

Omdat de precieze aard van de uitgewisselde data een zwarte doos kan zijn, is het moeilijk om een volledig beeld te geven van huidige datadelingspraktijken. Verticale overeenkomsten hebben meestal betrekking op waargenomen en verstrekte sociale en profieldata. Uitwisseling van dergelijke data ondermijnt niet de inkomsten die door een sociaal netwerk worden gegenereerd. Dit kan anders zijn voor verkregen sociale en profieldata, die van grote waarde kunnen zijn om adverteerders aan te trekken.

Verlaagt verplichte datadeling de prikkels voor investeringen in data substantieel?

In het geval van waargenomen en verstrekte sociale en profieldata is het onwaarschijnlijk dat de verplichting tot datadeling de stimulansen om die data te verzamelen zal ondermijnen. Deze soorten data zijn eenvoudig te verzamelen en worden in de meeste gevallen vrijwillig uitgewisseld met bedrijven die een complementaire rol vervullen (zie verticale en sector-overschrijdende datadeling).

Het antwoord kan echter anders zijn in het geval van verkregen data. Verkregen profieldata, die diepere inzichten in individuele voorkeuren bevatten, zijn zeer waardevol voor een gerichte benadering van consumenten. Het uitwisselen van dergelijke data ondermijnt de exclusiviteit en kan nieuwe spelers lokken op de advertentiemarkt. Het vermindert dus het potentieel voor het genereren van inkomsten, zodat verplichte uitwisseling de investeringsbereidheid kan schaden.

Wat zijn de negatieve gevolgen op het vlak van publieke belangen?

Datadeling met derden brengt het risico met zich mee dat privacy- en beveiligingsmaatregelen verwateren. Hoe meer data worden verzameld en uitgewisseld, en hoe groter het aantal producenten van apparatuur, app-ontwikkelaars en advertentiebedrijven die toegang krijgen tot data en misschien data op hun eigen servers opslaan, hoe groter de privacy- en beveiligingsrisico's. Contractuele voorwaarden kunnen dit risico slechts in beperkte mate verkleinen: het kan moeilijk te controleren zijn wat er gebeurt met gebruikersdata die van een platform zijn verdwenen. Evenzo kan het moeilijk zijn om te voorkomen dat gebruikersdata van gebruikersapparatuur verdwijnen, bijvoorbeeld wanneer een apparaat data synchroniseert of een back-up maakt naar clouddiensten.

5.3.2 *Huidige wettelijke bepalingen voor verplichte datadeling*

De AVG bevat het recht op dataportabiliteit dat de uitwisseling en het hergebruik van persoonsdata op verzoek van betrokkene faciliteert. Volgens de regelgeving krijgen betrokkenen bijvoorbeeld het recht om persoonsdata van de ene verwerkingsverantwoordelijke over te dragen aan een andere (indien technisch haalbaar). De huidige wetgeving staat dus een beperkte vorm van datadeling toe. Dit schiet echter tekort ten aanzien van de interoperabiliteit van sociale netwerken. Uitgebreidere vormen van datadeling, zoals real-time portabiliteit, vallen mogelijk onder de verplichtingen in de zin van de AVG (afhankelijk van technische haalbaarheid), maar de daadwerkelijke reikwijdte en het effect van de verplichtingen is nog niet duidelijk.

Het verhinderen van datatoegang kan in sommige gevallen neerkomen op machtsmisbruik. In dergelijke gevallen kan het mededingingsrecht van toepassing zijn (zie voor verdere bespreking van dit onderwerp de casus inzake verticale uitsluiting).

De essentiële faciliteiten-doctrine kan ook relevant zijn, aangezien deze betrekking heeft op situaties waarin een bedrijf met een machtspositie niet bereid is data uit te wisselen. Het directe gevolg van toepassing van de essentiële faciliteiten-doctrine zou zijn dat verplichtingen worden opgelegd om essentiële data uit te wisselen. Een cruciale vraag is dan of de data in kwestie "essentieel" zijn. Op basis van de verkenning in het kader van de casus verwachten we dat het moeilijk, zo niet onmogelijk, zal zijn om deze vraag te beantwoorden.

5.3.3 *Belemmeringen/voorwaarden voor verplichte datadeling*

Juridisch

Voor data kan bescherming onder het intellectuele eigendomsrecht gelden, waaronder auteursrechtbescherming, sui generis databankenbankbescherming en bedrijfsgeheimenbescherming. Deze regels lijken de uitwisseling van verstrekte/waargenomen data niet in de weg te staan. Auteursrechten spelen geen rol, aangezien het niet mogelijk is om daar databankstructuren uit te halen of om reverse engineering toe te passen op bedrijfsgeheimen.

Technisch

In het licht van de gemeenschappelijkheid van datadeling door sociale netwerken, verwachten we geen technische problemen ingeval van verplichtingen. Een uitzondering zou een geavanceerde vorm van real-time interoperabiliteit kunnen zijn, die technisch uitdagender is.

5.4 Casus: Banken

Deze casus onderzoekt de impact op de banksector van de asymmetrische verplichting tot datadeling voor banken (PSD2). We bespreken de financiële dienstverlening door de Big Techs, de drijfveren, de eruit voortvloeiende voordelen en de implicaties voor concurrentie en innovatie. We onderzoeken of het ontbreken van wederkerigheid bij datadeling leidt tot een ongelijk speelveld tussen traditionele financiële dienstverleners en de Big Techs. De hypothese is dat, wanneer Big Techs toegang hebben tot meer verschillende datatypen dan banken (data die relevant zijn voor productinnovaties), een asymmetrische plicht tot datadeling het concurrentievermogen van banken kan belemmeren. De vraag is dan of wederkerigheid bij verplichtingen tot datadeling het gelijke speelveld kan herstellen, en hoe de reikwijdte en de handhaving van een dergelijke voorwaarde moeten worden bepaald.

5.4.1 *Factoren voor verplichtingen in de Nederlandse visie op datadeling*

Ten aanzien van mogelijke verplichte datadeling tussen bedrijven, neemt de overheid in haar visiedocument datadeling een vijftal factoren in overweging. Wij bespreken deze hieronder.

Draagt datadeling bij aan een publiek belang zoals concurrentie, keuzevrijheid of innovatie?

Horizontale en verticale datadeling

Technologiebedrijven, startende bedrijven en FinTech-bedrijven hebben toegang tot bankdata onder PSD2, waarmee toegang tot de markt wordt gefaciliteerd. Als ze PSD2-licenties aanvragen, hebben ze toegang tot essentiële en unieke data die voorheen uitsluitend beschikbaar waren voor banken. Toegang tot deze data stelt hen in staat financiële diensten aan te bieden, en hun toetreding tot de markt kan de concurrentie in de financiële sector intensiveren. Als toetreders

diensten aanbieden die vergelijkbaar zijn met wat banken aanbieden, leidt toegang tot data via PSD2 tot meer horizontale concurrentie. Als bedrijven toegang tot bankdata gebruiken als input voor complementaire diensten, leidt datatoegang via PSD2 tot verticale concurrentie. Opgemerkt wordt dat PSD2-datatoegang financiële inclusie kan verbeteren en de toegang tot voorheen onontgonnen markten kan faciliteren, wat een bijzonder belangrijk voordeel is voor burgers zonder bankrekening en voor veel kleine en middelgrote ondernemingen.¹³³

Beweerd wordt dat de toegang van de Big Techs tot bankdata uiteindelijk de concurrentie in financiële diensten kan verzwakken, gezien hun vermogen om een kanteling van de markt te bewerkstelligen. Daarbij moet echter in aanmerking worden genomen dat, zelfs wanneer de Big Techs de markt voor financiële diensten zouden veroveren, de aard van de concurrentie tussen de Big Techs (althans op dit moment) inherent dynamischer is dan de concurrentie tussen banken. Ook kunnen de Big Techs wellicht bankdata combineren met andere persoonsdata, waardoor ze nieuwe en verbeterde diensten kunnen aanbieden. Bovendien leidt de toetreding van technologiegerichte bedrijven tot de financiële dienstverlening door verbeterde efficiëntie mogelijk tot verlaging van de kosten van financiële diensten voor de consument, zowel op detailhandels- als op institutioneel niveau. Naar verwachting zullen klanten toegang krijgen tot financiële producten die goedkoper, gemakkelijker, op maat gemaakt en toegankelijk zijn. Wat het netto-effect is voor de consument moet nog worden gezien.

Banken verzamelen ook allerlei niet-financiële data. De belangrijkste reden waarom ze niet-financiële data willen gebruiken is om hun kernactiviteiten te versterken: productiviteit te verhogen, risicobeheersing te verbeteren en de klantervaring te optimaliseren (dus niet om diensten te ontwikkelen buiten de financiële dienstverleningsmarkt). Gesteld kan worden dat het verwerven van niet-financiële data deel uitmaakt van de digitale transformatie van traditionele diensten zodat deze aan de efficiëntie- en kwaliteitseisen van het digitale tijdperk voldoen. In dit verband kunnen niet-financiële data de efficiëntie helpen verbeteren en de toegenomen concurrentiedruk van technologiebedrijven het hoofd bieden, maar het is geen essentiële input voor bancaire diensten.¹³⁴

Sector-overschrijdende datadeling

De term sector-overschrijdend is soms verwarrend. Het feit dat PSD2 buitenstaanders toegang geeft tot bankdata betekent niet dat er sprake is van sector-overschrijdende datadeling, omdat de buitenstaanders in kwestie de data gebruiken om tot de banksector toe te treden (wat horizontale concurrentie bevordert). Sector-overschrijdende datadeling is datadeling met andere sectoren dan die waar de data worden verzameld. Als banken toegang zouden krijgen tot data van andere sectoren, bijvoorbeeld data van technologiebedrijven, dan zou er sprake zijn van sector-overschrijdende datadeling.

Sommige banken pleiten voor de noodzaak van sector-overschrijdende datadeling met de Big Techs. Zij stellen dat PSD2 asymmetrische regelgeving heeft ingevoerd, met als gevolg asymmetrische concurrentie tussen banken en de Big Techs. Feit is dat de verplichte en asymmetrische datadeling tussen financiële dienstverleners en de Big Techs laatstgenoemde een datavoordeel verschaft en een bedreiging vormt voor de inkomstenstromen, winsten en uiteindelijk de innovatiestimulansen van banken. Wederkerigheid in de toegankelijkheid van datadeling zou een gelijk speelveld kunnen bevorderen.

¹³³ De Big Techs bieden dergelijke diensten doorgaans aan op markten buiten de Europese Unie.

¹³⁴ Banken ontwikkelen geen diensten buiten de markt voor financiële dienstverlening. In bepaalde rechtsgebieden mogen banken geen niet-bancaire activiteiten ontplooiën. In andere rechtsgebieden zouden ze bepaalde niet-bancaire activiteiten kunnen ontplooiën, maar zou dat voor hen erg duur zijn vanwege de bankregelgeving (financiële stabiliteit, prudentieel bankieren) en nalevingsvereisten, zoals antiwitwaspraktijken, het ken-uw-cliëntbeginsel, en onboardingprocessen.

Zijn de data eenvoudig te (re)produceren of te verzamelen door anderen?

Sinds de PSD2-regelgeving van kracht is en na het verkrijgen van een licentie, hebben de Big Techs gemakkelijk toegang tot data die betrekking hebben op bankrekeningen van klanten. Zonder PSD2 zouden de Big Techs geen data met betrekking tot bankrekeningen van klanten kunnen reproduceren of verzamelen. PSD2 heeft stimulansen gecreëerd om innovatieve diensten te ontwikkelen, omdat hun diensten doorgaans worden geleverd via platforms die al door klanten worden gebruikt.

Banken en andere financiële instellingen zijn al online aanwezig, waardoor ze veel profieldata kunnen verzamelen of kopen bij derden; daartoe hebben ze geen datadeelovereenkomst met de Big Techs nodig. Om hun klantprofielen te versterken kunnen banken hun traditionele data aanvullen met deze andere data over klanten. Andere soorten data, bijvoorbeeld sociale relaties of zoekactiviteiten, zijn doorgaans echter niet beschikbaar voor banken en kunnen door hen niet eenvoudig worden verzameld of gereproduceerd.

Lukt het niet of niet voldoende om op vrijwillige basis afspraken te maken over datatoegang?

Horizontale en verticale datadeling

Essentiële data van banken naar de Big Techs worden (in potentie) uitgewisseld onder PSD2.¹³⁵ De bestaande bankregelgeving staat niet toe dat bankdata op andere manieren dan via PSD2 beschikbaar worden gesteld. Voor beperkte uitwisseling van bankdata op commerciële basis bestaan er samenwerkingsovereenkomsten tussen technologiebedrijven en banken, waarbij de primaire aard van de overeenkomst niet is gerelateerd aan de data, maar waarbij beperkte datadeling noodzakelijk is. Voorbeelden zijn de samenwerking van Amazon met Bank of America voor zijn kredietprogramma¹³⁶ en de samenwerking van Google met Citigroup en Stanford Federal Credit Union om betaalrekeningen aan te bieden¹³⁷.

Sector-overschrijdende datadeling

Sector-overschrijdende datadeling heeft hoofdzakelijk betrekking op banken die toegang hebben tot data in andere sectoren, zoals data van de Big Techs. Commerciële overeenkomsten zijn vaak niet openbaar. Voor zover wij weten zijn gepersonaliseerde data met betrekking tot zoeken (opdrachten en klikgedrag) of sociaalnetwerkdata niet rechtstreeks toegankelijk voor banken, zelfs ingeval van toestemming van de klant.¹³⁸ Desondanks zijn de Big Techs datadeelovereenkomsten aangegaan met derden (ontwikkelaars), ter ondersteuning van apps op hun platforms (zoals bijvoorbeeld beschreven in de casus inzake sociale netwerken). Apps met betrekking tot bancaire en financiële diensten hebben doorgaans toegang tot dezelfde data, onder dezelfde algemene voorwaarden als anderen¹³⁹. Er zijn echter enkele voorbeelden te noemen waar dergelijke data zijn onderschept en gebruikt om inzichten te verkrijgen. Zo biedt Kabbage kleine bedrijven geautomatiseerde kredietlijnen aan, waarbij kredietbeslissingen worden gebaseerd op diverse datafactoren, waaronder bedrijfspvolume, hoe lang het bedrijf bestaat, transactievolume, socialemedia-activiteit, en kredietsscore van de eigenaar.¹⁴⁰

¹³⁵ Of ze worden uitgewisseld onder vergelijkbare regelgeving, zoals Open Banking-regels in het Verenigd Koninkrijk.

¹³⁶ In het kader van deze overeenkomst verstrekt Bank of America kapitaal aan Amazon Lending. De belangrijkste reden voor de overeenkomst was om geen bankdata uit te wisselen. Zie <https://www.americanbanker.com/slideshow/how-amazon-is-shaking-up-financial-services>

¹³⁷ Zie <https://www.reuters.com/article/us-google-finance/google-pay-to-offer-checking-accounts-through-citi-stanford-federal-idUSKBN1XN1IQ>

¹³⁸ Artikel 6, lid 1 van de AVG zou de combinatie mogelijk maken van persoonsdata in het bezit van de Big Techs met de traditionele financiële informatie van de bank, maar dergelijke gevallen van commerciële datadeling zijn ons niet bekend.

¹³⁹ Algemene voorwaarden voor datatoegang zijn bijvoorbeeld te vinden op <https://developers.google.com/terms/api-services-user-data-policy> (Google), <https://developer.linkedin.com/docs/rest-api> (LinkedIn), of <https://developers.facebook.com/docs/graph-api/using-graph-api> (Facebook).

¹⁴⁰ <https://www.valuepenquin.com/small-business/kabbage-business-loan-review>

Verlaagt verplichte datadeling de prikkels voor investeringen in data substantieel?

Ten aanzien van datadeling door banken onder PSD2 zal er (naar verwachting) geen sprake zijn van enige impact, omdat de data in kwestie doorgaans worden waargenomen/verstrekkt en kunnen worden beschouwd als een nevenproduct van financiële dienstverlening. Een mogelijke verplichting voor technologiebedrijven om data uit te wisselen behoeft niet noodzakelijkerwijs van invloed te zijn op hun stimulansen om te investeren in de productie van dergelijke data, mits het uitsluitend data betreft die aan hen zijn verstrekt of door hen zijn waargenomen, en niet om data die zijn afgeleid of verkregen uit deze data.

Wat zijn de negatieve gevolgen op het vlak van publieke belangen?

De financiële dienstensector komt in aanraking met gevoelige informatie over personen en ondernemingen. Met de digitale transformatie in de financiële sector komen er meer door klanten verstrekte data beschikbaar in digitale formats en wordt een toenemende hoeveelheid informatie waargenomen tijdens transacties. Het combineren van grote datasets maakt het gemakkelijker om inzichten te genereren, maar maakt de data ook gevoeliger voor beveiligingslekken. Ook kan de combinatie van grote datasets, zelfs als de data zijn gedeanonimiseerd, de privacy in gevaar brengen.¹⁴¹ Daarom vormt kwetsbaarheid met betrekking tot privacy en beveiliging een bedreiging voor de opkomst van Big Data-oplossingen in de financiële sector.

Consumenten kunnen baat hebben bij datadeling, maar de alomtegenwoordigheid van data, en bijgevolg databeveiliging, vormen een serieuze uitdaging voor zowel banken als de Big Techs. Hoe meer data worden verzameld en uitgewisseld, en hoe groter het aantal deelnemers aan het datadelingsproces, hoe groter de beveiligings- en privacyrisico's.

5.4.2 *Huidige wettelijke bepalingen voor datadeling*

De PSD2-richtlijn staat datadeling van banken naar andere bedrijven al toe¹⁴², en persoonsdata die door de Big Techs worden verzameld kunnen met toestemming van de gebruikers worden overgedragen aan banken onder de AVG-bepalingen.

Het argument dat de verplichtingen tot datadeling onder PSD2 asymmetrisch zijn, en dat deze asymmetrie de concurrentiepositie van banken ten opzichte van de Big Techs schaadt, klopt dus niet helemaal. Dit betekent echter niet dat er in het geheel geen sprake is van asymmetrie, maar die heeft geen betrekking op datadeling. Momenteel gelden voor nieuwe zakelijke activiteiten van banken dezelfde regels als voor hun traditionele bancaire activiteiten. Zoals door een aantal interviewpartners werd aangegeven concurreren banken daarom niet met nieuwkomers onder

¹⁴¹ Datasets die zijn geanonimiseerd kunnen vaak worden gedeanonimiseerd. Deanonimisering betekent dat anonieme data worden gekoppeld aan (een) openbare dataset(s) om de persoon te ontdekken op wie de data betrekking hebben (Lubarsky, 2017).

¹⁴² Krachtens Artikel 36 hebben aanbieders van betalingsinitiatiediensten en rekeninginformatiediensten het recht op vrije toegang tot de accountdata van hun klanten, met toestemming van diezelfde klanten om betaaldiensten of rekeninginformatiediensten te verlenen. Krachtens Artikel 31 van de technische reguleringsnormen moeten banken aan de volgende eisen voldoen: (i) zij dienen aanbieders van rekeninginformatiediensten dezelfde informatie te verstrekken over aangewezen betaalrekeningen en bijbehorende betaaltransacties die beschikbaar zijn voor de betaaldienstgebruiker bij het rechtstreeks aanvragen van toegang tot de rekeninginformatie; (ii) zij dienen aanbieders van betalingsinitiatiediensten dezelfde informatie te verstrekken over de initiatie en uitvoering van de betaaltransactie die is verstrekt of beschikbaar gemaakt aan de betaaldienstgebruiker wanneer de transactie rechtstreeks door deze laatste is geïnitieerd; en (iii) zij dienen aan aanbieders van betaaldiensten te bevestigen of het voor de uitvoering van een betaaltransactie benodigde bedrag beschikbaar is op de betaalrekening van de betaler.

dezelfde regels en voorwaarden.^{143,144} Een mogelijke oplossing zit in het idee van op activiteiten gebaseerde regelgeving die minder beperkingen oplegt (ten opzichte van de huidige regelgeving in de financiële sector) aan banken die willen innoveren.

5.4.3 *Belemmeringen/voorwaarden voor verplichte datadeling*

Juridisch

Als zowel de PSD2-bepaling als de AVG worden geïnterpreteerd als uitsluitend van toepassing zijnde op verstrekte of waargenomen data, is er geen wettelijke verplichting tot uitwisseling van afgeleide of verkregen data, zoals gebruikersvoorkeuren. Een dergelijke verplichting is echter ook niet nodig wanneer de onderliggende ruwe data waarop ze zijn gebaseerd worden uitgewisseld onder de respectieve wetten.

De AVG legt echter geen bepaalde vorm op voor datadeling van technologiebedrijven naar banken, noch zijn er specifieke eisen en voorwaarden voor real-time dataoverdracht. De huidige AVG-regelgeving zou kunnen worden aangepast (wellicht in de vorm van een domein-specifieke technische reguleringsnorm) ten behoeve van dergelijke (bijna real-time) datadelingen. Er zou ook een niet-domeinspecifieke aanpassing van de AVG kunnen worden ingevoerd, die zou gelden voor alle (bancaire en niet-bancaire) sectoren, met real-time toegang via API's en in een gestandaardiseerde vorm.

Technisch

De technische reguleringsnormen stellen hoge eisen aan datadeling. Op basis daarvan hebben banken hun eigen API's ontwikkeld, die bepalen hoe andere bedrijven toegang kunnen krijgen tot bankdata.

Het wordt onwaarschijnlijk geacht dat zich grote technische problemen zullen voordoen bij verplichte datadeling voor technologiebedrijven (die banken datatoegang verschaffen). Technische problemen zouden kunnen ontstaan in het geval van real-time toegang tot een mogelijk grote hoeveelheid data en (in het geval van sociale netwerken) moeilijk van persoonsdata gescheiden te houden sociale data. Een ander technisch probleem betreft de noodzaak van een vertrouwde (digitale) omgeving die transparante en veilige datadeling mogelijk maakt. Een verplichting is waarschijnlijk voldoende om dit in gang te zetten. Het Data Transfer Project illustreert deze ontwikkeling, aangezien dit mogelijk ook kan worden gebruikt als een vertrouwde uitwisseling die toegang tot data verschaft aan banken.

5.5 Casus: Administratieve bedrijfssoftware

De markt voor administratieve bedrijfssoftware is veelzijdig en omvangrijk. Deze markt wordt aangevoerd door een relatief kleine groep leveranciers die complexe grootschalige softwaresystemen kunnen leveren voor grote ondernemingen en overheidsinstanties. In deze systemen moeten diverse applicaties en clouddiensten worden geïntegreerd, eventueel aangevuld met complementaire oplossingen van kleinere ontwikkelaars. Belangrijke parameters zijn de mogelijkheid van interoperabiliteit en datadeling tussen systemen (dataportabiliteit).

¹⁴³ ING bepleit het principe van “dezelfde diensten, dezelfde risico's, dezelfde regels”. Zie <https://www.ing.com/About-us/Regulatory-international-affairs/Viewpoints/Financial-innovation-and-licensing.htm>

¹⁴⁴ Het is ook vermeldenswaard dat de activiteiten van de Big Techs op het gebied van financiële dienstverlening in bepaalde rechtsgebieden onderhevig kunnen zijn aan regelgeving. Evenzo zijn er andere rechtsgebieden waar het wettelijk kader de uitoefening van commerciële activiteiten door depositobanken of hun dochterondernemingen verbiedt.

Deze casus richt zich op de infrastructuur voor dataportabiliteit. Relevante soorten data bevatten operationele data van interne en externe processen en door softwaregebruikers beheerde bewerkingen, de sociale data van de organisatie die zijn opgeslagen in CRM en SRM systemen, en transactiedata. De activiteiten van gebruikers van bedrijfssoftware kunnen afhankelijk worden van hun software-infrastructuur om toegang te krijgen tot deze data. Dit leidt tot een bepaalde afhankelijkheid van hun softwareleverancier, wat vendor lock-in tot gevolg heeft en horizontale concurrentie beperkt. Afhankelijkheid kan ook de keuzevrijheid frustreren, vooral wanneer systemen beperkte flexibiliteit bieden om downstream-applicaties van andere ontwikkelaars te integreren.

5.5.1 *Factoren voor verplichtingen in de Nederlandse visie op datadeling*

Ten aanzien van mogelijke verplichte datadeling tussen bedrijven, neemt de overheid in haar visiedocument datadeling een vijftal factoren in overweging. Wij bespreken deze hieronder.

Draagt datadeling bij aan een publiek belang zoals concurrentie, keuzevrijheid of innovatie?

Verticale datadeling

Door de opkomst van clouddiensten in bedrijfssoftware stappen ontwikkelaars af van offline IT-systemen met closed source of propriëtaire software. Cloudoplossingen zijn doorgaans gebaseerd op open standaarden en open source software, en zorgen voor meer dataportabiliteit, interoperabiliteit en transparantie in de basisregels en codes van hun software. Cloudoplossingen bevatten een infrastructuur waarop derden oplossingen ontwikkelen. De cloud-omgeving functioneert daarbij als platform en verbetert het bedrijfsmodel van zowel leiders als nieuwkomers: kleine ontwikkelaars kunnen oplossingen bieden aan gebruikers van de infrastructuur van een leider, terwijl marktleiders een meer gedifferentieerde oplossingenportefeuille kunnen bieden die op hun infrastructuur wordt gehost.

Voor cloudoplossingen is verticale datadeling tussen de cloud-infrastructuur en de applicaties erboven nodig. Dit resulteert in rijke ecosystemen met een verscheidenheid aan nieuwe applicaties, die leiden tot verhoging van de efficiëntie van het uitgebreide netwerk van applicaties en databanken, wat economische voordelen oplevert, zoals synergiën, efficiëntere interne processen, beter geïnformeerde besluitvorming en meer innovatie.

Horizontale datadeling

Zoals hierboven uiteengezet, dragen beperkingen van dataportabiliteit bij aan vendor lock-in en beperken ze gebruikers bij het overstappen naar concurrerende leveranciers.

Er zijn geen stimulansen voor leveranciers van 'traditionele' offline propriëtaire systemen, noch voor leveranciers van de 'nieuwe' online cloud-infrastructuren, om hun klanten de mogelijkheid te bieden om over te stappen naar concurrenten. Concurrentie tussen providers leidt tot concurrerende biedingen in aanbestedingsfasen, maar zodra een klant is 'veroverd', ontstaan winstkansen of worden deze gecreëerd. Overstappen van de ene leverancier naar een andere kan technisch en operationeel ingewikkeld zijn, en bedrijven willen geen risico lopen voor de operationele continuïteit of dataverliezen lijden. De resulterende overstapkosten vergroten de mogelijkheid om hogere servicekosten te vragen voor voortzetting van de dienstverlening.

Dataportabiliteit (vrijwillig of verplicht) kan de drempels verlagen voor concurrentie tussen aanbieders van vergelijkbare software. Dataportabiliteit vermindert lock-in en verlaagt toetredingsdrempels, wat de effectiviteit van concurrentie op het gebied van prijs, kwaliteit en productaanbod (innovatie) vergroot.

Opgemerkt wordt dat beperkte dataportabiliteit niet de enige belemmerende factor is voor concurrentie. Ten eerste zijn er in de cloud-omgeving maar weinig leveranciers, vanwege de grote schaal die nodig is om een globaal netwerk van datacenters en aanverwante faciliteiten te kunnen onderhouden. En zelfs als data gemakkelijk kunnen worden overgedragen, kan het aanzienlijke financiële investeringen vergen om softwaresystemen uit te wisselen. De vraag naar maatwerk drijft deze overstapkosten nog verder op.

Zijn de data eenvoudig te (re)produceren of te verzamelen door anderen?

Deze vraag is in deze casus niet relevant. Het kernpunt hier is dat softwareaanbieders controle hebben over de toegang tot data doordat zij de infrastructuur beheren waarin organisaties hun (operationele en sociale) data opslaan. Deze data vormen de belangrijkste input voor de prestaties van een bedrijf en zijn een primaire bron van operationele, tactische en strategische besluitvorming. Als deze data niet kunnen worden overgedragen leidt dit tot lock-in.

Lukt het niet of niet voldoende om op vrijwillige basis afspraken te maken over datatoegang?

Verticale datadeling

Zoals hierboven uiteengezet, zijn 'traditionele' propriëtaire systemen meestal gesloten in de zin van verticale interoperabiliteit met en datatoegang voor externe applicatieontwikkelaars, terwijl op de cloud gebaseerde infrastructuren vaak opener zijn. De migratie van traditionele systemen naar cloud-systemen gaat dus gepaard met meer verticale datadeling.

Horizontale datadeling

Noch leveranciers van propriëtaire systemen noch aanbieders van cloud-infrastructuren hebben sterke drijfveren om het mogelijk te maken voor klanten om over te stappen via horizontale datadeling. Desalniettemin zijn recente initiatieven van bedrijven, zoals het Open Data Initiative (van Adobe, Microsoft en SAP) gericht op het bevorderen van datadeling binnen een ecosysteem¹⁴⁵. Hoewel dit een stap kan zijn in de richting van sectorbrede datadeling, kan het ook uitmonden in een nieuw ecosysteem dat leidt tot een lock-in voor zakelijke gebruikers. Openbare zelfreguleringsinitiatieven zoals de SWIPO-werkgroep voor belanghebbenden in de cloud (opgezet door de Europese Commissie) hebben tot doel mechanismen te ontwikkelen voor het vrije verkeer van niet-persoonsgebonden data door middel van gedragscodes en normen. Dit initiatief heeft fricties ervaren als gevolg van de heterogene belangen van de diverse belanghebbenden. Een van die belanghebbenden, Cigref (dat 150 grote Franse bedrijven en openbare besturen vertegenwoordigt), beschuldigde enkele van 's werelds toonaangevende clouddienstproviders ervan dat zij de missie van SWIPO frustreerden om *“de kern van hun bedrijven en hun vermogen om een lock-in te bewerkstelligen bij hun klanten te verdedigen”*¹⁴⁶. Tot dusverre heeft geen van deze initiatieven geleid tot een doorbraak in de zin van sectorbrede horizontale dataportabiliteit.

Verlaagt verplichte datadeling de prikkels voor investeringen in data substantieel?

Toonaangevende softwareontwikkelaars faciliteren nu al verticale dataportabiliteit met applicaties bovenop hun eigen systemen. Implementatie van horizontale dataportabiliteit kan de winstgevendheid verminderen. Sommige beargumenteren dat hiermee prikkels om te investeren in technologie en datainfrastructuur zouden afnemen. Echter, de toename van concurrentie zal hen waarschijnlijker aansporen om door middel van investeringen concurrerend te blijven.

¹⁴⁵ <https://www.techzine.nl/nieuws/cloud/420753/microsoft-sap-en-adobe-willen-open-data-initiative-versnellen/>

¹⁴⁶ Zie <https://www.cigref.fr/swipo-failure-regulate-european-cloud-market>. Opgemerkt wordt dat de beweringen van Cigref worden ondersteund door verklaringen van belanghebbenden die zijn betrokken bij dit onderzoek.

Wat zijn de negatieve gevolgen op het vlak van publieke belangen?

Standaarden voor datadeling zouden het moeilijker kunnen maken om aan de grote vraag naar maatoplossingen te voldoen. Het invoeren van standaardisatie zou het aanbod van op maat gemaakte en gedifferentieerde diensten kunnen schaden en leiden tot "one-size-fits-all" oplossingen.

5.5.2 Huidige wettelijke bepalingen voor verplichte datadeling

Er bestaat geen alomvattend, algemeen toegepast juridisch kader op nationaal of Europees niveau dat datadeling tussen leveranciers van administratieve software faciliteert. Omdat operationele data niet-persoonsgebonden zijn, vallen ze niet onder de AVG. Bijgevolg is het recht op dataportabiliteit zoals vastgelegd in de AVG niet van toepassing.

De zelfregulerende gedragscodes van SWIPO voor dataportabiliteit tussen cloud-providers, zoals ingevoerd met de verordening betreffende niet-persoonsgebonden data, kunnen relevant zijn.¹⁴⁷ Artikel 6, lid 1 van de Verordening (EU) 2018/1807 van het Europees Parlement en de Raad inzake een kader voor het vrije verkeer van niet-persoonsgebonden data in de Europese Unie, verplicht de Europese Commissie om de ontwikkeling van zelfregulerende gedragscodes op EU-niveau aan te moedigen en te faciliteren, waaronder "beste praktijken ter vergemakkelijking van het veranderen van dienstverlener en dataportabiliteit in een gestructureerde, algemeen gangbare en voor machines leesbare vorm, met inbegrip van open standaardformaten". Tot dusverre zijn zelfregulerende gedragscodes onvoldoende effectief gebleken, zoals werd bevestigd door interviewpartners.

Dataportabiliteit kan worden opgelegd als remedie in het kader van het mededingingsrecht onder de essentiële faciliteiten-doctrine, maar soortgelijke beperkingen als elders besproken zijn van toepassing: een leverancier moet een machtspositie hebben en de data moeten essentieel zijn (overeenkomstig de juridische connotatie).

5.5.3 Belemmeringen/voorwaarden voor verplichte datadeling

Juridisch

Interviewpartners bevestigden dat er geen zwaarwegende juridische obstakels zijn voor het creëren van uitwisselbaar en interoperabel gebruik van softwaresystemen. Voor bedrijven in de rol van softwaregebruikers is er geen equivalent van de AVG waaronder zij portabiliteit van hun operationele (verstrekke) data kunnen afdwingen. Opgemerkt wordt dat het vaststellen van AVG-achtige regelgeving voor bedrijven, inclusief een verplichting om data over te dragen (met name operationele data), de effectiviteit van initiatieven zoals SWIPO en het Open Data Initiative zou kunnen vergroten.

Technisch

Er lijken geen onoverkomelijke technische obstakels te zijn als het erom gaat portabiliteit en interoperabiliteit tussen administratieve systemen mogelijk te maken. Grote, geavanceerde zakelijke gebruikers eisen steeds vaker dat zij toegang krijgen tot hun data en deze kunnen ophalen, waarmee druk wordt uitgeoefend om technische problemen op te lossen.

¹⁴⁷ Zie <https://swipo.eu/> voor een overzicht van de zelfregulerende codes overeengekomen door de SWIPO-Werkgroep.

5.6 Casus: Verticale uitsluiting

Deze casus richt zich niet op een specifieke markt of sector, maar onderzoekt verticale datadeling tussen verticaal geïntegreerde platforms en derden, zoals downstreamdienstverleners, onafhankelijke websites en app-ontwikkelaars. Het doel is om in wat algemenere zin de impact van verticale uitsluiting van datatoegang op concurrentie en innovatie te onderzoeken.

We illustreren deze casus met het voorbeeld van datadeling tussen Strava en Relive, waartoe we openbare bronnen in de media gebruiken. Via Strava kunnen atleten hardloop- en fietsactiviteiten volgen en analyseren, met data die zijn verzameld via telefoons, horloges en volgapparatuur van derden. Dit ontwikkelde zich tot een sociaal netwerk waar atleten contact konden leggen en communiceren. Gebruikers profiteren van apps van derden die de functionaliteiten van Strava aanvullen. Ze baseren zich op gebruikersdata die zijn geüpload naar Strava en uitgewisseld met derden. Een voorbeeld is Relive, dat op basis van locatiedata animerende kaarten genereert verrijkt met zelf-gegenereerde foto- en videomateriaal. Het geheel wordt gevisualiseerd in een 3D kaart. Kort nadat Relive een sociaalnetwerkfunctionaliteit had geïntroduceerd, weigerde Strava Relive toegang tot zijn API. De overeenkomst stelt dat partners *"... geen applicatie mogen creëren die Strava slechts repliceert of ermee concurreert"*¹⁴⁸. Daarnaast vermeldt de overeenkomst *"... Strava kan nu of in de toekomst producten en diensten ontwikkelen die vergelijkbaar zijn of concurreren met uw ontwikkelaarsapplicaties"*¹⁴⁹ Strava kan dus de functionaliteit van een partner kopiëren en vervolgens stoppen met het verstrekken van data¹⁵⁰.

5.6.1 Factoren voor verplichtingen in de Nederlandse visie op datadeling

Ten aanzien van mogelijke verplichte datadeling tussen bedrijven, neemt de overheid in haar visiedocument datadeling een vijftal factoren in overweging. Wij bespreken deze hieronder.

Draagt datadeling bij aan een publiek belang zoals concurrentie, keuzevrijheid of innovatie?

Horizontale datadeling

Hoewel het in deze casus om verticale relaties gaat, is het nuttig om stil te staan bij horizontale kwesties. Een digitaal platform en een downstreamleverancier beginnen doorgaans met een complementaire of verticale relatie. Door de toevoeging van nieuwe functies kan een downstreamdienstverlener dichter bij het bedrijfsmodel van het platform komen. Dit is vermoedelijk gebeurd in het geval van Strava en Relive. Als gevolg van een dergelijke ontwikkeling ontstaat een horizontale concurrentiedreiging voor het platform. Door datatoegang te weigeren, kan een platform proberen zijn marktpositie te verdedigen tegen deze horizontale concurrent. Door middel van de API-overeenkomst van Strava worden partners die een horizontale concurrentiedreiging uitoefenen uitgesloten van datatoegang. Het verbieden van dit soort clausules zou horizontale concurrentie ondersteunen en zowel platforms als downstream-app-ontwikkelaars stimuleren om meer te innoveren.

Verticale datadeling

De relatie tussen een platform en een downstreamleverancier is per definitie verticaal. De leverancier heeft baat bij toegang tot de data van het platform, bijvoorbeeld om een innovatieve dienst te ontwikkelen of de groei van zijn klantenbestand een impuls te geven. Het platform breidt het aantal en de verscheidenheid van gebruikers uit, wat zijn ecosysteem verrijkt. Derhalve draagt datadeling bij aan innovatie en verscheidenheid (van nieuwe applicaties en diensten) voor zowel platformgebruikers als klanten van de downstream-app.

¹⁴⁸ <https://www.strava.com/legal/api>, tweede punt onder de inleidende paragraaf (geraadpleegd op 10 juni 2020).

¹⁴⁹ <https://www.strava.com/legal/api>, §11.B (geraadpleegd op 10 juni 2020).

¹⁵⁰ Dit gebeurde in het geval van "Segment Challenge"; zie <https://webbindustries.com/hackernews/story/20421625>.

Het belang van verticale datatoegang is met name relevant voor ontoegankelijke data die onmisbaar zijn om een bedrijf te runnen. Toegang tot dergelijke essentiële data — in de zin van een essentiële faciliteit (zie ook hieronder) — faciliteert dan de toegang tot de markt. In dergelijke gevallen zijn de data in kwestie uniek en exclusief, bijvoorbeeld omdat ze niet zonder rechtsbevoegdheid kunnen worden verkregen, of omdat het verzamelen en verwerken ervan grote (verzonken) investeringen vergt. Een aandachtspunt daarbij is echter dat situaties met essentiële data een uitzondering lijken te zijn (zie hieronder).

Zijn de data eenvoudig te (re)produceren of te verzamelen door anderen?

Deze vraag kan over het algemeen niet worden beantwoord. Daarvoor zouden de kenmerken van de voor een bepaald bedrijf onmisbare data moeten worden beoordeeld. Het blokkeren van API-toegang door Strava had bijvoorbeeld geen fatale gevolgen voor Relive. Voor een deel baseerde Relive zich op Strava voor toegang tot locatiedata, maar klanten van Relive kunnen de app ook rechtstreeks verbinden met een grote verscheidenheid aan volgapparatuur, bijvoorbeeld van Suunto, Polar, Apple, Garmin en Adidas¹⁵¹. Het blokkeren van datatoegang heeft mogelijk ook geen fatale gevolgen voor de sociaalnetwerkfunctionaliteiten van Relive, aangezien Relive (waarschijnlijk) gebruik kan maken van Facebook voor de toegang tot sociale data van gebruikers. Een andere app-ontwikkelaar die door Strava werd afgesloten van datatoegang, Segment Challenge, was echter volledig afhankelijk van de API van Strava¹⁵² en ging kort nadat Strava de datatoevoer had afgesloten ten onder.¹⁵³ Ondanks deze dramatische afloop moet echter worden vastgesteld dat de data in kwestie niet per se essentieel of niet-repliceerbaar hoeven te zijn.

Meer in het algemeen kan het problematisch zijn om aan te tonen dat de data van een gevestigd bedrijf onmisbaar zijn om te kunnen concurreren. Mededingingsautoriteiten kunnen echter data meenemen in de beoordeling van concurrentieproblemen. Op dit moment is er weinig empirisch bewijs van gevallen waarin data als essentieel zijn aangemerkt. Dit is wellicht een gevolg van het feit dat data en de kenmerken ervan moeilijk objectief te meten zijn, iets wat veel gemakkelijker is voor prijzen en kwaliteitsniveaus.

Lukt het niet of niet voldoende om op vrijwillige basis afspraken te maken over datatoegang?

Horizontale datadeling

Het is mogelijk dat platforms niet bereid zijn om deel te nemen aan horizontale datadeling vanwege de concurrentiedreiging die ontstaat door ondersteuning van een (potentiële) vervangende dienst. Toen de sport-app Relive bijvoorbeeld de mogelijkheid introduceerde om vrienden te volgen, beëindigde volg-app Strava de datadeelovereenkomst met Relive.

Vrijwillige horizontale datadeling kan desalniettemin een feit worden. In het verleden verleende Facebook (net als Strava) geen datatoegang aan apps die zijn kernfunctionaliteit repliceerden. In 2018 werd dit beleid echter afgeschaft door het sociale netwerk.

Verticale datadeling

Doorgaans faciliteren platforms verticale datadeling met downstreampartijen, omdat de waarde van hun ecosysteem erdoor toeneemt. Derden hebben baat bij datadeelovereenkomsten, omdat die hun dienstverlening verbeteren of hun groei bevorderen. Men mag dus verwachten dat verticale samenwerkingsverbanden automatisch tot stand komen. Naarmate downstream-apps echter dichter bij de zakelijke kernfunctionaliteiten van een platform komen, kunnen ze een horizontale

¹⁵¹ Zie <https://www.relive.cc/> (geraadpleegd op 10 juni 2020).

¹⁵² <https://news.ycombinator.com/item?id=20421625>

¹⁵³ <https://twitter.com/segmntchallenge>

concurrentiedreiging gaan vormen en (afhankelijk van de voorwaarden van de overeenkomst) worden afgesloten van datatoegang.

Verlaagt verplichte datadeling de prikkels voor investeringen in data substantieel?

Het antwoord op deze vraag hangt af van het soort platform en de aard van de data in kwestie.

In het geval van waargenomen en verstrekte sociale, contextuele of profieldata is het onwaarschijnlijk dat verplichte datadeling de stimulansen om dergelijke data te verzamelen zullen ondermijnen. Deze soorten data zijn gemakkelijk te verzamelen en worden vaak vrijwillig uitgewisseld met bedrijven die complementaire diensten aanbieden, dat wil zeggen in verticale relaties. Maar zelfs de uitwisseling van deze data met (downstreampartners die zich ontwikkelen tot) horizontale concurrenten zullen naar alle waarschijnlijkheid niet van invloed zijn op de stimulansen van het platform om de data te verzamelen. Deze data zijn een nevenproduct van de productie van de kernfunctionaliteit. Zoals door ons aangetoond, zou een verplichting om locatiedata en sociale data uit te wisselen met een potentiële concurrent als Relive, de stimulansen van Strava om deze data te verzamelen niet schaden.

In andere gevallen van verticale uitsluiting dan door ons geïllustreerd, kunnen verkregen sociale en profieldata diepere inzichten in individuele voorkeuren bevatten. Dit maakt ze zeer waardevol voor gerichte benadering van de consument, wat vooral relevant is voor platforms die de aandacht van de consument te gelde maken via de verkoop van advertentieruimte. De uitwisseling van verkregen data ondermijnt de exclusiviteit ervan en kan leiden tot meeliften door anderen. Het risico op meeliften is kleiner bij verticale partners, waarvan de waarde wordt gerealiseerd binnen het ecosysteem van een platform, zodat het platform mee kan delen in die waarde via zijn prijsstelling. Het risico op meeliften is groter wanneer afgeleide of verkregen data moeten worden uitgewisseld met horizontale concurrenten. Dat vermindert het potentieel om inkomsten te genereren, zodat verplichte uitwisseling de investeringsbereidheid kan schaden.

In het geval van gedragsdata hangt het antwoord af van de aard van het bedrijfsmodel en het risico op reverse engineering (zie bijvoorbeeld de casus inzake online zoeken).

Wat zijn de negatieve gevolgen op het vlak van publieke belangen?

Datadeling met derden brengt (per definitie) het risico met zich mee dat de bescherming van privacy en beveiliging verwateren. Hoe meer data worden verzameld en uitgewisseld, en hoe groter het aantal makers van apparaten, app-ontwikkelaars en advertentiebedrijven die toegang krijgen tot data en mogelijk data opslaan op hun eigen servers, hoe groter de privacy- en beveiligingsrisico's worden. Contractuele voorwaarden kunnen dit risico slechts in beperkte mate verkleinen: het kan moeilijk te controleren zijn wat er gebeurt met gebruikersdata die zijn verdwenen van een platform. Evenzo kan het moeilijk zijn om te voorkomen dat gebruikersdata verdwijnen van gebruikersapparaten, bijvoorbeeld wanneer een apparaat data synchroniseert of back-ups maakt naar clouddiensten.

5.6.2 Huidige wettelijke bepalingen voor verplichte datadeling

De aard van bestaande voorwaarden voor datadeling hangt af van het geval in kwestie. Met name als verticale uitsluiting persoonsdata betreffen, dan kunnen betrokkenen het upstream platform opdracht geven om de data uit te wisselen met - of over te dragen aan - de downstream app die is afgesloten van datatoegang. De AVG biedt de app-ontwikkelaar echter geen rechtsgrond om dergelijke toegang te eisen.

In situaties waar dit van toepassing is bevat de AVG een recht op dataportabiliteit dat op verzoek van betrokkene de uitwisseling en het hergebruik van persoonsdata faciliteert. In veel gevallen kan de AVG verticale uitsluiting echter niet aanpakken, omdat de verordening simpelweg niet van toepassing is op de situatie in kwestie.

Uitsluiting van datatoegang kan in bepaalde situaties gelden als misbruik van machtspositie, wat in strijd is met het mededingingsrecht. De essentiële faciliteiten-doctrine is met name relevant omdat deze betrekking heeft op situaties waarin een bedrijf met een machtspositie niet bereid is tot datadeling. Het directe gevolg van toepassing van de essentiële faciliteiten-doctrine zou zijn dat verplichtingen worden opgelegd om essentiële data uit te wisselen. De bewijslast die nodig is om de essentiële faciliteiten-doctrine te kunnen toepassen is echter zwaar. Een cruciale vraag is dan of de data in kwestie "essentieel" zijn. We verwachten dat het in veel gevallen erg moeilijk zal zijn om deze vraag te beantwoorden¹⁵⁴.

In het algemeen geldt dat in gevallen waarbij de datatoegang wordt geblokkeerd in een setting die overeenkomt met verticale uitsluiting, de huidige wetgeving mogelijk niet effectief is, zoals blijkt uit actuele beleidsinitiatieven met betrekking tot ex ante interventies op platforms met een machtspositie (of platforms met poortwachtersposities). Zie Van Gorp en De Bijl (2019) voor een uitgebreid onderzoek naar dit onderwerp. Zij stellen oplossingen voor waarbij het upstream platform moet bewijzen dat het uitsluitingsgedrag objectieve redenen dient. Als het platform dit niet kan bewijzen, kan het worden gedwongen om een eind te maken aan dit gedrag. In een van de voorgestelde oplossingen wordt vermeden dat een machtspositie of exclusiviteit moet worden vastgesteld op basis van een structurele analyse van markt- en productkenmerken, maar kan een machtspositie worden afgeleid uit waargenomen gedrag.

5.6.3 *Belemmeringen/voorwaarden voor verplichte datadeling*

Juridisch

Of er juridische problemen kunnen optreden inzake verplichte datadeling hangt af van de situatie. Data kunnen vallen onder intellectuele eigendomsrechten, waaronder auteursrechtelijke bescherming, sui generis databankenbescherming en bedrijfsgeheimbescherming. Voor een specifiek voorbeeld zie onder andere de casus inzake sociale netwerken.

Het definiëren en verkennen van de precieze aard van verplichtingen tot datadeling om verticale uitsluiting aan te pakken zal moeilijk worden. Zo moet bijvoorbeeld worden vastgesteld tot welke laag van de data-infrastructuur er toegang moet zijn. Het kan ook zijn dat mededingingsautoriteiten moeten aantonen dat een dergelijke interventie proportioneel is, dat wil zeggen dat er geen minder ingrijpende interventies voorhanden zijn om hetzelfde doel te bereiken.

Technisch

Of er technische problemen kunnen optreden inzake verplichte datadeling hangt af van de situatie. In het geval van sociale netwerken lijken er bijvoorbeeld geen technische problemen te zijn bij verplichtingen. Meer in het algemeen hebben marktpartijen in sectoren waar verticale datadeling gangbaar is manieren gevonden om daarmee om te gaan.

¹⁵⁴ Een uitzondering wordt gevormd door slimme meters voor energieverbruik: in sommige landen hebben distributienetbeheerders een wettelijk monopolie om huishoudens aan te sluiten op het energienetwerk en slimme meters te installeren bij die huishoudens. Hierdoor hebben ze rechtstreeks en exclusief toegang tot de energiedata van huishoudens. De EU-regelgeving voorziet echter in beperkte toegang tot energieverbruiksdata, waardoor buitenstaanders de mogelijkheid hebben om huishoudens een slimme thermostaat aan te bieden die data van de slimme meter ontvangt.

5.7 Samenvatting

In haar visiedocument datadeling heeft de Nederlandse overheid criteria geformuleerd voor het opleggen van verplichtingen tot datadeling. In dit onderzoek zijn die criteria onderzocht in vijf casussen: **1) zoekmachines**, onderzoek naar de redenen voor horizontale datadeling met betrekking tot zoek- en klikgedrag; **2) sociale netwerken**, onderzoek naar de noodzaak tot verplichte datadeling voor sociale-mediaplatformen (met name van sociale data); **3) betaaldiensten**, onderzoek naar het effect van verplichte datadeling onder PSD2 op de concurrentiepositie van banken en de wenselijkheid van verplichte datadeling voor toegangzoekers; **4) administratieve bedrijfssoftware**, onderzoek naar de mate waarin vendor lock-in kan worden aangepakt door middel van verplichte dataportabiliteit; en **5) verticale uitsluiting**, met betrekking tot situaties waarin een platform een verticale partner uitsluit van datatoegang. Onderstaande tabel geeft een overzicht van hoe per casus aan de overheidscriteria wordt voldaan.

Criteria	Zoeken	Sociale media	Bancair	B2B software	Verticale uitsluiting
Soort data	gedragsdata	sociaal	profieldata	operationele data	alle soorten
Uitwisselingsrichting	horizontaal	horizontaal	sector-overschrijdend	horizontaal	verticaal
Publiek belang	kantelen van de markt	kantelen van de markt	speelveld	Gebruikersmobiliteit	misbruik
Repliceerbaarheid	niet meer	In hoge mate	niet door banken	niet relevant	hangt ervan af
Vrijwillig uitgewisseld	H: nee V: niet relevant	H: soms V: ja	nee	H: nee V: ja	H: niet relevant V: nee
Effect op investeringen	in data: 0 andere: +/- ^a	in data: 0 andere: 0	in data: 0 andere: + ^b	in data: 0 andere: 0/- ^a	in data: 0 andere: + ^c
Negatieve neveneffecten	risico op reverse engineering.	privacyisico's	privacyrisico's	standaarden kunnen de mogelijkheid om te voldoen aan de vraag naar maatwerk beperken	mogelijke objectieve redenen voor discriminatie

^a Afhankelijk van hoe sterk de negatieve neveneffecten (zoals vermeld in de tabel) zijn.

^b Als banken toegang hebben tot profieldata, kunnen ook zij (en niet alleen technologiebedrijven) de correlaties tussen bankdata en profieldata onderzoeken. Dit betekent dat meerdere entiteiten kunnen innoveren met dezelfde data.

^c Niet-discriminatoire toegang tot data verhoogt de beschikbaarheid van de datapool voor een groter aantal innovators.

In hoofdstuk 6 gaan we dieper in op de algemene conclusies uit de casussen, in samenhang met de inzichten uit hoofdstukken 2, 3 en 4, gevolgd door een bespreking van beleidsmaatregelen.

6 Conclusie en beleidsimplicaties

Dit hoofdstuk brengt de inzichten uit de voorgaande hoofdstukken samen, op basis waarvan mogelijke varianten van verplichte datadeling worden besproken. Ten eerste geven we een overzicht van de belangrijkste kansen die datadeling biedt. Vervolgens bespreken we de beleidsimplicaties door twee beleidsdoeleinden verder uit te werken. Het eerste doeleinde is om te profiteren van de kansen die datadeling biedt op het gebied van innovatie. Dit kan door te zorgen voor de nodige randvoorwaarden en, waar nodig, verplichtingen op te leggen. Het tweede doeleinde is het adresseren van mogelijke concurrentieproblemen. We gaan daarbij in op vraag 6 van het ministerie, zoals verwoord in het inleidende hoofdstuk (Welke specifieke varianten van verplichte datadeling in de technologiesector zijn te overwegen?). We behandelen tevens vraag 7 (Welke juridische, economische en technische consequenties zijn te verwachten van de verschillende varianten van verplichtingen tot datadeling?) door de implicaties te schetsen van verplichte datadeling, bijvoorbeeld ten aanzien van standaardisatie en juridische implementatie.

6.1 Belangrijkste kansen geboden door datadeling

De kernvraag die in dit rapport wordt behandeld luidt: **Wat is de rol van data in de technologiesector en wat zijn de kansen en risico's op verplichte datadeling?** De antwoorden op deze complexe vraag geven inzicht in kosten en baten van varianten van verplichtingen tot datadeling met en tussen technologiebedrijven. Dergelijke bedrijven (in de technologiesector) houden zich bezig met de ontwikkeling en productie van op elektronica gebaseerde technologieproducten (zoals smartphones of zelfrijdende voertuigen) of bieden op elektronica gebaseerde technologie aan als dienst (zoals navigatie, communicatie, cloud-computing enzovoort). De Big Techs werken met bedrijfsmodellen waarin data een sleutelrol spelen en die blijf geven van sterke netwerkeffecten. Met behulp van (data)technologie organiseren ze interacties met en tussen waardeketens.

Hieronder vatten we op algemeen niveau onze bevindingen samen uit de analyse en de casussen. We maken onderscheid tussen de impact van datadeling op innovatie, concurrentie, gebruikersmobiliteit en privacy/beveiliging.

Impact van datadeling op innovatie

Verticale datadeling is gericht op het verminderen van operationele inefficiëntie en het faciliteren van de ontwikkeling van nieuwe diensten. Veel platforms ondersteunen verticale datadeling om de diversiteit van de beschikbare diensten binnen hun ecosystemen te vergroten. Deze innovatie-ecosystemen bieden een **vertrouwde omgeving voor datadeling**. Zonder een dergelijke omgeving kunnen scheppers van data en/of verwerkingsverantwoordelijken minder geneigd zijn om data uit te wisselen, waardoor het innovatiepotentieel wordt beperkt. Een beleidsvoorstel is om de ontwikkeling van dergelijke vertrouwde omgevingen voor datadeling te stimuleren.

Verticale datadeling kan ook partijen van buiten de bedrijfstak helpen om de waardeketen te betreden met data uit andere sectoren. Op die manier kan verticale datadeling **sector-overschrijdende data-analyses** faciliteren, wat de mogelijkheden om nog meer innovaties te ontdekken vergroot. Wanneer sector-overschrijdende datastromen echter slechts één richting opgaan, kan dit leiden tot een probleem ten aanzien van een gelijk speelveld ten behoeve van het vermogen om te innoveren. Een beleidsvoorstel is om te zorgen voor wederkerigheid van datastromen, in de zin dat **"u toegang krijgt tot (al) mijn data als ik toegang krijg tot (al) uw**

data". Een dergelijke wederkerigheid zou een gelijk speelveld tot gevolg hebben en het innovatiepotentieel van sector-overschrijdende datadeling maximaliseren. Bovendien genereert het netwerkeffecten omdat het innovatiepotentieel van alle deelnemers toeneemt met elke extra deelnemer. Dit betekent dat naarmate het aantal deelnemers groeit, het voor andere spelers (uit verschillende sectoren) steeds aantrekkelijker wordt om ook mee te doen aan de datadelingsovereenkomst. Merk op dat deze voordelen niet voortvloeien uit financiële compensatiemechanismen: het innovatiepotentieel van een bedrijf groeit niet als het wordt betaald voor gegevenstoegang. Bovendien leidt financiële compensatie niet tot netwerkeffecten.

Horizontale uitwisseling van data kan bijdragen tot een **gelijkwaardiger toegang tot trainingsdata**. Als gevolg hiervan zullen concurrerende bedrijven waarschijnlijk actiever gaan concurreren door te investeren in algoritmeontwikkeling en innovatie, in plaats van zoveel mogelijk data te verzamelen in een poging om een kanteling van de markt in hun voordeel te bewerkstelligen. Echter, horizontale datadeling kan innovatie ook negatief beïnvloeden. De casus inzake zoekmachines bevestigt dat de uitwisseling van gedragsdata een risico op reverse engineering met zich meebrengt. Dit kan innovatiestimulansen schaden. Een verplichting om gedragsdata uit te wisselen kan in strijd zijn met de wetgeving inzake bepaalde intellectuele eigendomsrechten, die ten doel heeft om die stimulansen te beschermen.

Impact van datadeling op concurrentie (en innovatie)

Het is mogelijk dat een bedrijf niet zonder meer bereid is tot datadeling wanneer die data **onmisbaar, uniek en exclusief** zijn, en wanneer die exclusiviteit bescherming biedt tegen concurrentiedruk. Deze druk moet hier worden gezien vanuit een hoek van 360 graden: binnen en tussen markten, en ook in de zin van onderhandelingsposities in de waardeketen.

Exclusiviteit kan voortkomen uit het verleden (bijvoorbeeld banken in het pre-PSD2-tijdperk) of het gevolg zijn van een strategie om een concurrentiepositie te versterken (zoals de beslissing van Google om de controle te krijgen over de browsermarkt door middel van bundeling). Wanneer dergelijke afgeschermd concurrentieposities de verticale datastroom in de waardeketen belemmeren, kan een **verticale verplichting tot datadeling** helpen om data-gedreven innovaties in de waardeketen te ontsluiten en **toegang tot de bedrijfstak bevorderen**.

Exclusiviteit kan ook het gevolg zijn van netwerk- en leereffecten die een kanteling van de markt veroorzaken, mogelijk in combinatie met (concurrentiebeperkende) strategieën die dataposities versterken. Het gaat daarbij met name om sociale data, profieldata, gedragsdata en (mogelijk) operationele data. Als er geen sprake is van concurrentiebeperkend gedrag, kan een **verplichting tot horizontale uitwisseling** van deze data **het risico op kanteling van de markt beperken**, aangezien dit concurrenten helpt om netwerk- en leereffecten in gang te zetten. De mogelijkheid bestaat echter dat een verplichting tot horizontale datadeling niet effectief is als gevolg van concurrentiebeperkende strategieën. Als Google bijvoorbeeld wordt verplicht om zoekdata uit te wisselen, zal dit de groei van concurrerende zoekmachines niet bevorderen wanneer Google zijn browser en zoekmachine koppelt aan zijn Android mobiele besturingssysteem. Tegelijkertijd is een verplichting tot datadeling misschien niet meer nodig als Google wordt gedwongen om die diensten los te koppelen van Android. Concurrerende browsers kunnen dan een groter marktaandeel verwerven en zoekdata inzien die in de Google-zoekmachine zijn ingevoerd, waardoor die data niet meer exclusief zijn. In zulke gevallen **moet concurrentiebeperkend gedrag eerst worden aangepakt**, voordat een verplichting tot horizontale datadeling wordt overwogen.

Bij het opleggen van verplichte datadeling moet rekening worden gehouden met **wederkerigheid**. Wederkerigheid is van belang om het risico dat de markt alsnog kantelt te beperken en te **voorkomen dat het op een na grootste bedrijf de nieuwe speler met een machtspositie**

wordt. In dergelijke gevallen impliceert wederkerigheid de onderlinge uitwisseling van specifieke en vergelijkbare soorten data tussen horizontale concurrenten. We wijzen erop dat dit een beperktere vorm van wederkerigheid is dan die waaraan hierboven wordt gerefereerd in relatie tot sector-overschrijdende datadeling gericht op het stimuleren van innovaties, die zou moeten gelden voor alle en verschillende soorten data tussen verschillende belanghebbenden.

Impact op gebruikersmobiliteit

Datadeling kan rechtstreeks bijdragen aan de mobilisatie van consumenten en hen in staat stellen om over te stappen en te multi-homen (zich tegelijkertijd abonneren op verschillende dienstverleners). Met name horizontale overdracht van (waargenomen en verstrekte) **sociale data** en door gebruikers **zelf-gegenereerde inhoud** (inclusief operationele data) maakt het mogelijk voor gebruikers om over te stappen naar vergelijkbare diensten of gelijktijdig gebruik te maken van gelijksoortige diensten (multi-homing). Als data niet kunnen worden overgedragen leidt dit tot lock-in, zoals geïllustreerd in de casus inzake administratieve bedrijfssoftware.

Er kan sprake zijn van **andere obstakels voor overstappen**. Bij administratieve bedrijfssoftware wordt vendor lock-in ook veroorzaakt door de operationele integratie van software, die moeilijk te ontwarren is. Voor socialemediadiensten leidt migratie van sociale data niet direct tot lagere overstapdrempels, omdat de sociale contacten zelf niet mee migreren. De uitwisseling van sociale data kan echter een ander sociaal netwerk helpen om de vrienden van een persoon uit te nodigen door middel van gerichte marketingcampagnes via andere platforms of communicatiekanalen. Facebook faciliteert dergelijke campagnes voor partners, en recente wijzigingen in de gebruiksvoorwaarden voor zakelijke gebruikers voorkomen dat ze worden afgesloten wanneer ze vergelijkbare functionaliteiten aanbieden. Om overstappen verder te vergemakkelijken, zou het datamobiliteitsbeleid kunnen worden aangevuld met beleid dat is gericht op een lichte vorm van interoperabiliteit op dienstniveau. De Britse mededingings- en marktautoriteit¹⁵⁵ heeft onlangs bijvoorbeeld geconcludeerd dat het nuttig zou zijn om een functionaliteit te hebben waarmee gebruikers bestaande contacten kunnen uitnodigen om zich aan te sluiten bij een nieuw netwerk. Ook stelt de autoriteit een functionaliteit voor waarmee berichten tegelijkertijd op meerdere platformen geplaatst kunnen worden. Deze lichte vormen van verplichte interoperabiliteit zou overstappen en multi-homing ondersteunen, zonder de strengere standaardisatie die nodig zou zijn om via elke willekeurige interface berichten op meerdere platformen in te kunnen zien en erop te reageren.

Impact van datadeling op privacy en beveiliging

Het verzamelen en opslaan van data brengt beveiligingsrisico's met zich mee. Voor datadeling is openheid, transport en overdracht van data vereist, wat dergelijke risico's vergroot. De toegenomen openheid en de benodigde logistiek vergen extra investeringen om deze risico's te verkleinen. De risico's op datadeling hangen onder andere af van de dataformats en de aard van de toegang. Datatoegang door middel van downloads, bijvoorbeeld, zoals in het geval van dataportabiliteit, draagt (kopieën van) de data over buiten het informatiesysteem van een platform en (mogelijk) buiten de controle van de originele datahouder. Met API-toegang kunnen apps en websites rechtstreeks gebruik maken van data, zonder kopieën te hoeven maken. Zo kunnen datahouders de identiteit van API-gebruikers en hun datagebruik controleren.

Naast deze technische kwesties hangen de risico's die gepaard gaan met datadeling af van de soorten data. Sociale data en gebruikersprofielen verliezen het grootste deel van hun waarde na anonimisering. De uitwisseling van dergelijke data brengt privacyrisico's met zich mee en vereist een juridische basis, zoals een wettelijke verplichting of toestemming van eindgebruikers. Een dergelijke legitimiteit sluit echter niet alle privacyrisico's uit. Voor datadeling op grotere schaal zijn

¹⁵⁵ (Competition and Market Authority, 2020)

dus passende maatregelen nodig om privacy en beveiliging te waarborgen. Dit geldt in mindere mate voor gedrags- en operationele data, die anoniem kunnen worden uitgewisseld, omdat ze hun waarde behouden na anonimisering.

6.2 Beleidsopties

We structureren de beleidsopties die voortkomen uit de analyse aan de hand van twee beleidsdoelstellingen, en bespreken de juridische implicaties in dit verband. De **eerste doelstelling** is om te profiteren van de kansen die datadeling biedt, m.n. voor innovatie. Dit kan door te zorgen voor de noodzakelijke randvoorwaarden en, waar nodig, verplichtingen op te leggen. De **tweede doelstelling** is de aanpak van concurrentieproblemen.

Beleidsdoelstelling I: data mobiliseren en innovatie stimuleren

Op basis van bovenstaande discussie kunnen drie kansen voor datadeling worden onderscheiden:

- I. **Waardeketens optimaliseren en ecosystemen verrijken:** Uitwisseling van operationele **data** binnen een waardeketen leidt tot grotere efficiëntie en vermindert marktfricties. De reden hiervoor is dat datadeling in een waardeketen helpt om manieren te ontdekken om productiviteit te verhogen en kosten te verlagen. Voorbeelden hiervan zijn te vinden in de logistiek, de zorg, de bouw en in de land- en tuinbouw. Vermindering van marktfricties is het directe gevolg van grotere transparantie door datadeling.
- II. **(Onvoorziene) innovaties in verschillende sectoren ontsluiten:** Deze situatie betreft de mogelijkheden die datadeling biedt door voorwaarden te scheppen voor onvoorziene innovaties als gevolg van uitwisseling van data met andere sectoren dan die waar de data worden verzameld. De voordelen van datadeling tussen sectoren komen voort uit verscheidenheid: met elke nieuwe soort data in een dataset groeit het aantal data-combinaties dat kan worden verkend exponentieel, en daarmee de mogelijkheid om nieuwe innovaties te ontdekken. Op voorhand is het onmogelijk te zeggen welke soorten data van belang zijn en welke niet. Over het algemeen geldt dat hoe meer data worden uitgewisseld hoe beter. Dergelijke voordelen zijn vaak niet te voorzien en vergen riskante inspanningen.
- III. **Klanten mobiliseren:** Datadeling kan helpen om gebruikers of kopers te mobiliseren of hen in staat stellen gemakkelijker over te stappen. Belemmeringen voor het overzetten van data van het ene platform naar het andere dragen ertoe bij dat gebruikers zich gevangen voelen, aangezien ze overstapkosten dragen als ze van provider wisselen. Omdat het hier gaat om overstappen tussen vergelijkbare providers, is de gerelateerde vorm van datadeling doorgaans horizontaal. Dit betreft hoofdzakelijk sociale data (zoals contactdata en adresboeken) en door gebruikers zelf gegenereerde inhoud. Belemmeringen voor het overdragen van deze data leiden tot lock-in.

Deze mogelijkheden kunnen elkaar aanvullen. Zo zijn PSD2-verplichtingen gericht op het realiseren van kans I. Tegelijkertijd kunnen de Big Techs hierdoor tot de banksector toetreden met data die in andere sectoren zijn verworven, wat bijdraagt aan kans II. Dit kan echter een ongelijk speelveld tot gevolg hebben vanwege asymmetrische dataposities: Big Techs hebben bijvoorbeeld een enorme hoeveelheid data over zoekactiviteiten, locaties en sociale contacten. Om een gelijk speelveld te behouden, zou toegang tot de data van banken ook afhankelijk kunnen worden gesteld van het verlenen van datatoegang aan banken. Met andere woorden, het helpt als datadeling wederkerig is in de zin van *"u krijgt toegang tot mijn data als ik toegang krijg tot die van u"*. Meer algemeen gesteld leidt het toepassen van een dergelijke brede wederkerigheidsvoorwaarde in alle sectoren

waar data worden uitgewisseld ertoe dat techbedrijven die actief willen worden in andere sectoren zelf ook meer data uitwisselen. In andere gevallen creëren de hierboven vermelde kansen op zichzelf al winst. Horizontale datadeling draagt bijvoorbeeld bij aan het overstappen van klanten, zoals in het geval van de markt voor administratieve bedrijfssoftware (kans III).

Alvorens uitwisselingsverplichtingen te overwegen is het zinvol om aandacht te besteden aan de randvoorwaarden die een vruchtbare voedingsbodem vormen voor vrijwillige datadeling. Een aandachtspunt hierbij is dat er voor datadeling een infrastructuur en een governancestelsel nodig zijn om de toegang tot en overdracht van data te beheren en te orkestreren. Deze taken kunnen op verschillende manieren worden uitgevoerd. In een verticale setting kan een bedrijf bijvoorbeeld een gecentraliseerde architectuur hebben voor datadeling, zoals een sociaal netwerk dat data uitwisselt met downstream app-ontwikkelaars. In andere situaties worden deze functionaliteiten gedecentraliseerd uitgevoerd via een peer-to-peer architectuur, bijvoorbeeld door horizontale concurrenten die data met elkaar uitwisselen, of door banken die data uitwisselen met nieuwkomers in hun bedrijfstak. Datadeling kan ook worden geregeld door een afzonderlijke entiteit, opgezet door deelnemers uit de sector.

In al deze situaties brengen het beheer en de organisatie van datadeling verantwoordelijkheden met zich mee, zoals het ondersteunen van innovatie, gebruikersmobiliteit en concurrentie, met inachtneming van de rechten en belangen van scheppers van data. Het dragen van dergelijke verantwoordelijkheden dient het publieke belang, maar sluit ook aan bij bedrijfsmodellen van zelfstandige platformen die datadelen met en tussen ecosysteemdeelnemers faciliteren om zo veel mogelijk waarde voor en door deelnemers te (laten) creëren. Concurrerende krachten kunnen deze verantwoordelijkheden ondersteunen, maar niet noodzakelijkerwijs, vanwege de mogelijkheid van marktfalen. Daarom is het een no-regret-optie om datadelingspraktijken te onderwerpen aan eenvoudige principes die uitgaan van vertrouwen en een gelijk speelveld. Door dergelijke principes 'over de hele linie' toe te passen worden mogelijke belemmeringen voor vrijwillige datadeling en (dus) de noodzaak van potentiële verplichte datadeling weggenomen.

Hiertoe stellen wij voor dat **principes van 'good governance' van datadeling** (*Good Data Sharing Governance* - GDSG) worden vastgelegd in algemene wetten:

1. **Bescherming van de rechten van scheppers van data.** Scheppers van data — zowel personen als bedrijven — moeten de mogelijkheid hebben om (i) te bepalen welke data anderen over hen verzamelen, en (ii) het gebruik van hun data te controleren. Dus waar de AVG individuele personen al in staat stelt om de toegang tot persoonsdata te controleren, moeten bedrijven en organisaties ook in staat worden gesteld om toegang tot data te controleren die naar hen kan worden herleid (in het bijzonder informatie met strategische waarde). Het doel is in de eerste plaats om bij scheppers van data vertrouwen te wekken om data uit te wisselen en/of anderen in staat te stellen data over hen te verzamelen, en ten tweede om klanten te ondersteunen bij overstappen en multi-homing.
2. **Ondersteuning van een gelijk speelveld.** Om randvoorwaarden te creëren voor de vele mogelijkheden met betrekking tot concurrentie en innovatie in bedrijfsmodellen en waardeproposities, moet er gelijke toegang tot data zijn. Daarvoor gelden de volgende voorwaarden:
 - a. *Non-discriminatie op basis van transparantie:* het bedrijf dat datadeling orkestreert moet gelijke voorwaarden toepassen op downstreambedrijven die de data ontvangen, ongeacht of het downstreambedrijf een eigen dochteronderneming is of een derde partij. Door deze voorwaarde neemt het risico op markt afscherming door verticaal geïntegreerde bedrijven af. De voorwaarde moet gelden voor alle relevante aspecten, zoals de kosten van

datatoegang, technische haalbaarheid en continue real-time toegang (aangezien een bedrijf dat data bezit of beheert continue real-time toegang heeft, mogen anderen niet met een achterstand beginnen).

- b. *Wederkerigheid van toegang tot datadeling*: een bedrijf dat om (toegang tot) data verzoekt moet transparant zijn over de eigen data en openstaan voor uitwisseling daarvan met de andere partij of, afhankelijk van de implementatie van datadeling, met een centrale architectuur voor datadeling. Hiertoe zijn mogelijk data-audits benodigd. Wederkerigheid voorkomt dat bedrijven die om datatoegang vragen bestaande datavoordelen te gelde kunnen maken in andere markten. Het vermindert ook het risico op het kantelen van een markt door leereffecten en genereert netwerkeffecten ten voordele van ieders innovatievermogen. (Zoals hierboven uitgelegd, leiden financiële compensatiemechanismen voor gegevenstoegang niet tot dergelijke voordelen).

Beide principes zijn in de eerste plaats van toepassing op verstrekte/waargenomen data, aangezien bij die soorten data het risico op negatieve neveneffecten, zoals reverse engineering of schendingen van bepaalde intellectuele eigendomsrechten, minimaal is. Maar mocht een dergelijk risico bestaan, dan zouden uitzonderingen op principe 2 mogelijk moeten zijn. In dat geval moet de oorspronkelijke verwerkingsverantwoordelijke bewijzen dat er objectieve redenen bestaan om niet tot uitwisseling over te gaan; deze kunnen bijvoorbeeld verband houden met efficiëntie, privacyrisico's, schendingen van het intellectueel eigendomsrecht of het risico op reverse engineering. We verwachten dat hiervan zelden of nooit sprake zal zijn bij verstrekte/waargenomen data die geen betrekking hebben op of niet herleidbaar zijn tot scheppers van data. Daar komt nog bij dat, als ze al kunnen worden herleid naar scheppers van data, principe 1 van toepassing is.

Principe 1 kan worden geïmplementeerd via algemene wetgeving die van toepassing is op elke entiteit die data verzamelt en verwerkt. Met betrekking tot persoonsgegevens (verstrekt / nageleefd) legt de AVG al de toepassing van principe 1 op. Aangezien de AVG geen betrekking heeft op bedrijfsgegevens, is het nuttig om een zakelijk equivalent van de AVG in te voeren. Merk op dat er onzekerheid bestaat over de vraag of dataportabiliteit onder de AVG continu en real-time kan zijn. Dit zou naar onze mening het geval moeten zijn als de betrokkene dat wenst. Het is dus nodig om dit te verduidelijken (of eventueel om de juridische tekst aan te passen). Evenzo moet de overdraagbaarheid van bedrijfsgegevens continu en real-time zijn als de dataschepper dat wenst.

Principe 2 kan ook worden geïmplementeerd door middel van algemene wetgeving die van kracht is wanneer partijen besluiten om data uit te wisselen. De EU-verordening inzake platform-to-business relaties (P2B-verordening), die op 12 juli 2020 in werking is getreden, voorziet in een reeks regels die een "eerlijke, transparante en voorspelbare zakelijke omgeving" creëren voor kleinere bedrijven en handelaren op onlineplatforms. Het lijkt wellicht praktisch om principe 2 expliciet op te nemen in een herziene P2B-verordening, maar een nadeel hiervan is dat deze verordening enkel geldt voor onlineplatforms. Een andere kandidaat is de EU-verordening met betrekking tot vrij dataverkeer, maar die verordening heeft uitsluitend betrekking op niet-persoonsgebonden data. Daarom is er behoefte aan een nieuwe en algemenere verordening, gericht op een gelijk speelveld voor datadeling en van toepassing op elk soort bedrijf dat zich bezighoudt met datadeling. Verder verdient het aanbeveling om de mogelijkheden te verkennen voor opname in de aanstaande wetgevingsinitiatieven van de EU, die zijn aangekondigd in de Europese datastrategie (gepubliceerd door de Europese Commissie in 2020). De verwachte initiatieven (waaronder een mogelijke Datawet) hebben tot doel om B2B-datadeling te faciliteren en te ondersteunen. Op basis van wat in het data-strategiedocument wordt vermeld, overlappen de beoogde doelen van de initiatieven met principe 2, aangezien ze waarschijnlijk eisen van eerlijkheid, transparantie en non-discriminatie zullen bevatten waar uitwisselingsverplichtingen wenselijk zijn. Eventuele aanvullende sectorale verplichtingen tot datadeling kunnen dan naast

horizontale EU-wetgeving bestaan in het kader van de Europese datastrategie, waarin de GDSG-principes zijn opgenomen.

In sectorale verplichtingen kunnen desgewenst maatwerkisen worden opgenomen in aanvulling op de horizontale wetgeving, zoals in de vergunningsvereiste in PSD2. De reden voor aanvullende sectorale verplichtingen is dat de GDSG-principes voornamelijk zijn gericht op het hold-up-probleem, dat een belemmering kan zijn voor belanghebbenden om data uit te wisselen binnen en tussen waardeketens. Deze principes geven vorm aan de ontwikkeling van vertrouwde omgevingen voor datadeling, maar verplichten belanghebbenden niet om data uit te wisselen. Brede acceptatie van de GDSG-principes vormt echter een stimulans voor alle belanghebbenden in de dataeconomie om deel te nemen aan datadelingsecosystemen, met name vanwege de wederkerigheidsclausule die leidt tot netwerkeffecten. Zijn bedrijven ondanks deze principes nog steeds niet bereid tot verticale datadeling, dan kan het nodig zijn om verplichtingen op te leggen, vooral als bedrijven hun concurrentiepositie te danken hebben aan exclusieve toegang tot unieke data en ze om defensieve redenen niet bereid zijn om data uit te wisselen met andere belanghebbenden binnen en tussen waardeketens. Een dergelijke terughoudendheid belemmert concurrentie en innovatie in en tussen waardeketens. In dat geval kunnen sectorspecifieke verplichtingen tot verticale datadeling nodig zijn.

Het verbinden van de principes met de aanstaande EU-initiatieven is zelfs nog relevanter gezien de geplande evaluatie van het kader voor bepaalde intellectuele eigendomsrechten uit het oogpunt van verbetering van datatoegang en -gebruik (waaronder een mogelijke herziening van de databankrichtlijn en een mogelijke verduidelijking van de manier waarop de richtlijn bescherming bedrijfsgeheimen moet worden toegepast) die de Commissie voornemens is om in dezelfde context uit te voeren. De lessen die kunnen worden geleerd in het kader van een dergelijke evaluatie zullen ook van nut zijn bij een mogelijke implementatie van de GDSG-principes. De evaluatie kan met name meer duidelijkheid bieden over de relatie van het kader voor dit intellectueel eigendomsrecht met mechanismen voor datadeling die de exclusiviteit en geheimhouding van door bepaalde intellectuele eigendomsrechten en bedrijfsgeheimen beschermde data kunnen beperken.

Een belangrijk punt ter nadere overweging is waar de verantwoordelijkheid en aansprakelijkheid kunnen worden ondergebracht voor schade als gevolg van datadeling. Dit geldt zowel voor persoonsdata die onder de AVG vallen als voor data die worden beschermd door bepaalde intellectuele eigendomsrechten, bedrijfsgeheimen of commerciële geheimhoudingsovereenkomsten. Waar houdt bijvoorbeeld de verantwoordelijkheid van de oorspronkelijke verwerkingsverantwoordelijke op wanneer persoonsdata worden overgedragen aan een nieuwe verwerkingsverantwoordelijke? En moeten alle partijen die toegang hebben tot bedrijfsdata hoofdelijk aansprakelijk worden gesteld als commerciële data worden gelekt uit een gecentraliseerde datadelingseenheid? Om te zorgen voor een effectieve implementatie van de GDSG-principes, wordt aanbevolen om proactief opheldering te geven over dergelijke kwesties in nieuwe wetgeving, of in richtsnoeren zoals gepubliceerd met betrekking tot persoonsdata door het Europees Comité voor databescherming.

Beleidsdoelstelling II: aanpak van concurrentieproblemen die voortkomen uit netwerk- en leereffecten

Naast het stimuleren van innovaties en het mobiliseren van gebruikers (kansen I, II en III, besproken bij beleidsdoelstelling I hierboven), kan verplichte datadeling helpen om winner-takes-all-effecten te voorkomen. Specifieker gesteld: een verplichting tot horizontale datadeling per geval kan bijdragen tot concurrerende marktresultaten wanneer:

1. data leiden tot leer- en/of netwerkeffecten (d.w.z. gedragsdata, sociale data, profieldata),
en

2. deze data uniek zijn en exclusief ter beschikking staan van een marktleider (d.w.z. niet replicerbaar zijn door een concurrent).

Bij een beslissing per geval om horizontale uitwisseling te verplichten zou moet worden meegenomen of:

3. Concurrenten met een klein (zeg 5% tot 10%) marktaandeel zijn niet in staat om de omvang van de gegevens te realiseren die nodig zijn om de leider in te halen, omdat:
 - a. er geen sterk afnemende marginale opbrengsten van data zijn. Met andere woorden, een marktaandeel van 5% tot 10% is niet voldoende om de minimale efficiënte schaal voor leren te realiseren.
 - b. de relevantie van de gegevens snel afneemt. Sommige gegevens (bijv. Sociale grafiekgegevens) blijven jarenlang waardevol (met regelmatige updates). Andere gegevens (bijv. Zoekopdrachten) beginnen al binnen enkele weken of zelfs dagen aan waarde te verliezen. Uiteindelijk moet dit van geval tot geval worden beoordeeld, aangezien de waarde van gegevens wordt bepaald door de aard van de dienst en het toegepaste bedrijfsmodel.
 - c. er korte continue of zich herhalende leercycli zijn en er snelle acceptatie van nieuwe technologie plaatsvindt. 'Kort' verwijst hier naar een paar dagen / weken om nieuwe inzichten te ontdekken en toe te passen in de technologie. Wanneer zich bijvoorbeeld een grote onverwachte gebeurtenis voordoet, moet een zoekmachine zijn index en algoritmen onmiddellijk bijwerken op basis van signalen die via zoekopdrachten worden ontvangen.
4. Er (overige) concurrentiebeperkende strategieën door de marktleider worden gevolgd om zijn huidige positie af te schermen. In dat geval moet dit concurrentieverstorend gedrag sowieso worden aangepakt om een verplichting tot gegevensuitwisseling effectief te laten zijn. Wanneer gegevens exclusiviteit het gevolg is van concurrentieverstorend gedrag, kan het al voldoende zijn om dat gedrag op grond van het algemene mededingingsrecht rechtstreeks te stoppen, waardoor de verplichting om gegevens te delen overbodig wordt.
5. Het delen van gegevens verwaarloosbare risico's van reverse engineering met zich meebrengt. Dit is met name relevant voor gedragsgegevens. In het geval van risico's van reverse engineering kan een verplichting tot gegevensuitwisseling de innovatieprikkels verminderen zoals beschermd door het IER-kader, en met name door de databank- en handelsgeheimenrichtlijn. In elk geval moet de bewijslast op de primaire gegevensbeheerder rusten op basis van objectieve redenen om niet te delen (bijv. schendingen van bepaalde intellectuele eigendomsrechten en risico's van reverse engineering).

Nadat bovengenoemde omstandigheden in overweging zijn genomen, moet in de eerste plaats een verplichting worden opgelegd aan de marktleider om specifieke soorten data horizontaal uit te wisselen (bijvoorbeeld geanonimiseerde zoekopdrachten en klikgedrag in het geval van zoekmachines). Om echter te voorkomen dat de op een na grootste concurrent het nieuwe bedrijf met een machtspositie wordt, moet de datadeling wederkerig zijn (d.w.z. dat toegangzoekers in ruil voor datatoegang de marktleider toegang moeten verlenen tot dezelfde soorten data). Dit betekent dat horizontale concurrenten op hun beurt onderling data moeten uitwisselen, anders behoudt de marktleider een datavoordeel.¹⁵⁶ Kortom, **horizontale concurrenten moeten onderling specifieke soorten data uitwisselen, zodat alle spelers over dezelfde pool van relevante data kunnen beschikken.**

Dergelijke verplichtingen zijn mogelijk op te leggen op grond van het mededingingsrecht, op voorwaarde dat de relevante data als een 'essentiële faciliteit' worden beschouwd (wat

¹⁵⁶ Als A verplicht is om op wederkerige basis data uit te wisselen met B en C, dan zou A nog steeds een grotere dataset hebben (van A+B+C) dan B en C (respectievelijk van B+A en van C+A). Als B en C besluiten om vrijwillig hun data ook uit te wisselen, dan hebben alle partijen toegang tot de grotere dataset van A+B+C.

waarschijnlijk het geval is wanneer aan voorwaarden 1, 2 en 3 is voldaan), wederkerigheid wordt beschouwd als een 'redelijke toegangsprijs', en dat een groepsvrijstelling onder Artikel 101 van het VWEU het mogelijk maakt voor concurrenten om data uit te wisselen. Bovendien moeten de verplichtingen in overeenstemming zijn met de AVG voor zover het data betreft die niet zonder waardeverlies kunnen worden geanonimiseerd (met name sociale en profieldata). In dit verband kan een verplichting op basis van het mededingingsrecht wellicht worden beschouwd als een wettelijke verplichting die een rechtsgrond kan zijn voor de uitwisseling van persoonsdata in overeenstemming met de AVG (of anders kunnen de toestemming van betrokkene of de wettelijke belangen van de verwerkingsverantwoordelijke worden beschouwd als rechtsgronden). Daarnaast zijn andere AVG-principes, zoals minimale dataverwerking en doelbinding, van toepassing, en hiermee wordt erop toegezien dat de uitgewisselde persoonsdata zo beperkt mogelijk zijn voor het doel, zoals uiteengezet in het mededingingsbesluit.

Het voordeel van toepassing van het mededingingsrecht is dat het uitgaat van beoordelingen per geval. Het dwingt besluitvormers om rekening te houden met de concurrentiesituatie, inclusief de dynamiek van sector-overschrijdende concurrentie. Dit is met name van belang in de context van moligopolistische concurrentie. Bij dit soort concurrentie is het voorkomen van een kanteling van de markt wellicht niet welvaartsverhogend, maar wel het voorkomen van misbruik van daaruit voortvloeiende machtsposities, zoals koppelen, bundelen en andere vormen van uitsluitingsgedrag.

Tegelijkertijd hebben recente mededingingszaken aangetoond dat mededingingsautoriteiten worstelen met de marktdynamiek bij het definiëren van markten en het beoordelen van machtsposities. Hierdoor kosten mededingingszaken meestal veel tijd. Als er sprake is van schade, dan kan deze onherstelbaar zijn tegen de tijd dat er een uitspraak wordt gedaan, vanwege de netwerk- en leereffecten. Er is dus behoefte aan snellere besluitvorming, die tot stand kan komen aan de hand van drie opties: 1) herziening van de richtsnoeren om relevante markten te definiëren en machtsposities te beoordelen, 2) het creëren van een nieuw instrument waarmee mededingingsautoriteiten structurele concurrentieproblemen kunnen aanpakken zonder de noodzaak om een machtspositie vast te stellen, en 3) ex ante regulering.

Van deze drie mogelijkheden lijkt optie 1 de gemakkelijkste en snelste manier, maar deze optie kan leiden tot frictie met eerdere precedentes. Optie 2 wordt momenteel door de Commissie overwogen (onder de naam "Nieuw Mededingingsinstrument"). Deze optie legt de lat voor een interventie lager, maar creëert daardoor automatisch meer ruimte voor willekeurige beslissingen. Optie 3 vereist nog steeds een uitgebreide analyse van bedrijfsmodellen in hun competitieve omgeving om een dominante positie of poortwachterspositie vast te stellen die wordt ingenomen door een digitaal platform. Het ontbreken van een sectordefinitie die enger is dan de 'technologiesector' maakt dit er niet eenvoudiger op. Bovendien is de concurrentie op digitale markten onderhevig aan voortdurende veranderingen, wat inhoudt dat regelgevende beslissingen en de onderliggende analyse van concurrentie regelmatig moeten worden herzien¹⁵⁷. Daarnaast zal elke (herziening van een) beslissing aan kritische en langdurige rechterlijke toetsingen worden onderworpen¹⁵⁸. Tenslotte leggen Van Gorp en De Bijl (2019) uit dat ex ante regulering het risico vergroot van het opleggen van onnodige interventies die innovaties schaden. Merk op dat de Commissie mogelijkheden onderzoekt om ex ante regulering in te voeren voor zogenaamde poortwachtersplatforms die de mededingingsregels zouden aanvullen. Deze nieuwe ex-ante-verordening zou mogelijk ook taken voor poortwachtersplatforms kunnen omvatten om gegevens met rivalen te delen.

¹⁵⁷ Zeg elke twee tot drie jaar; in ieder geval vaker dan bij telecomregulering het geval is, omdat die markt als veel minder dynamisch wordt beschouwd dan digitale markten

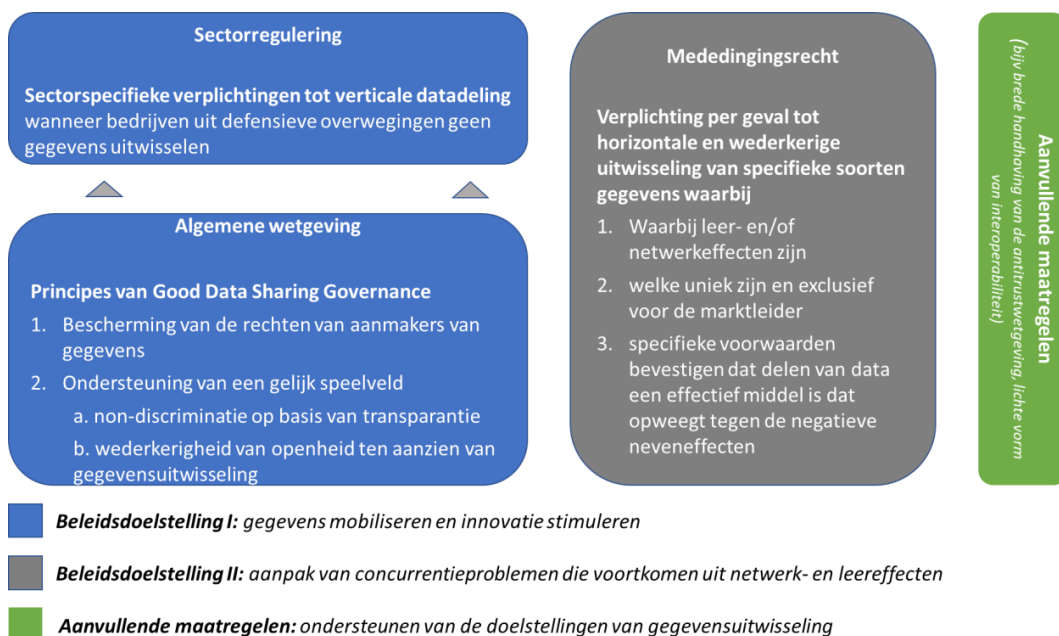
¹⁵⁸ Te meer vanwege het ontbreken van een duidelijk analytisch kader om digitale concurrentie te beoordelen.

Aanvullend beleid

Uit de voorgaande discussie blijkt dat enkel het stimuleren en/of verplicht stellen van datadeling niet altijd voldoende is om de beleidsdoelstellingen te realiseren. Aanvullende beleidsopties kunnen nodig zijn. Een algemene aanvullende beleidsmaatregel is het voorkomen van concurrentiebeperkende strategieën die volgen op een kanteling van de markt en gericht zijn op het monopoliseren van verwante of parallelle markten. Soms is sectorspecifiek beleid noodzakelijk als aanvulling op het datadelingsbeleid. Met betrekking tot socialemediadiensten heeft het overdragen van sociale data bijvoorbeeld nog niet tot gevolg dat iemands sociale netwerk naar een ander platform migreert. Om deze overstapdrempel te vermijden, zouden verschillende diensten een nog hogere mate van interoperabiliteit moeten hebben. In aanvulling op datadeling zou interoperabiliteit op dienstenniveau dus ook de communicatie tussen deelnemers van verschillende diensten moeten toestaan. Een dergelijke interoperabiliteit vereist standaardisatie, die de mogelijkheden kan beperken voor dienstverleners om te innoveren en zich met verschillende functionaliteiten te onderscheiden. Daarom stellen we voor om dergelijke verplichtingen te beperken tot lichte versies, zoals een functionaliteit waarmee gebruikers bestaande contacten kunnen uitnodigen om zich aan te sluiten bij een nieuw netwerk, of de mogelijkheid om (nieuwe) berichten op meerder platforms tegelijkertijd te kunnen posten. In elk geval moet bij interoperabiliteitsverplichtingen op dienstverleningsniveau worden voorkomen dat strenge standaardisatie-eisen de mogelijkheden van platforms om te innoveren en differentiëren beperken. In het geval van socialemediadiensten zou een dergelijke interoperabiliteitsverplichting kunnen worden opgelegd op basis van de Europese elektronische communicatiecode, aangezien het gaat om een interpersoonlijke communicatiefunctie die essentieel is voor de kerndienst van het platform.

Overzicht van beleidsopties

We sluiten dit hoofdstuk af met een figuur waarin de hierboven besproken beleidsopties worden samengevat. Daaronder wordt door middel van een aantal (deels hypothetische) voorbeelden de betekenis van de figuur in specifieke settings aangegeven.



Als voorbeeld bij het onderste blauwe vak, over algemene wetgeving, kan het productiebedrijf John Deere dienen, dat een breed scala aan machines produceert, waaronder tractoren. Door sensoren te installeren op tractoren kan het bedrijf data verzamelen over het gebruik ervan op landbouwgronden. Door middel van dergelijke data kan een reeks innovatieve applicaties voor boeren worden ontwikkeld. Zo stond het bedrijf aanvankelijk voor de keuze een eigen afdeling op te

zetten om dergelijke apps te ontwikkelen, dan wel een open platform te creëren voor boeren en onafhankelijke app-ontwikkelaars om verstrekte en waargenomen data uit te wisselen. In het eerste geval zouden er geen data zijn uitgewisseld, zodat de algemene wetgeving in het vak niet van toepassing is. Stel dat in het tweede geval dat de algemene wetgeving reeds van toepassing was, dan houdt principe 1 in dat boeren de mogelijkheid moeten krijgen om de toegang tot data die naar hen kunnen worden herleid te controleren. Principe 2(a) voorkomt discriminatie tussen ontwikkelaars, terwijl onder 2(b) iedere partij die toegang tot data vraagt moet vermelden welke soorten data (verstrekkt en waargenomen) deze partij zelf ter beschikking heeft, en dat die toegang moet verlenen tot die data wanneer daarom gevraagd wordt. Uiteraard dienen alle partijen de rechten van de scheppers van data in acht te nemen en te beschermen (principe 1).

De casus inzake banken illustreert een sectorspecifieke verplichting om verticaal data uit te wisselen (blauwe vak linksboven). Onder PSD2 werd een sectorspecifieke verplichting tot datadeling opgelegd aan banken, die niet welwillend stonden tegenover datadeling omdat dat de deur zou openzetten voor meer (dynamische) concurrentie. De eenzijdige verplichting voor banken om data uit te wisselen zou echter kunnen leiden tot een probleem ten aanzien van een gelijk speelveld: een dergelijke verplichting nodigt Big Tech-spelers uit om hun datavoordeel (waarbij sprake is van een geheel andere dimensie) in te zetten in de financiële sector. Ons voorstel lost dit probleem op door de sectorale verplichting aan te vullen met de GDSG-principes (blauwe vak linksonder). De noodzaak van wederkerigheid die is ingebed in deze principes betekent dat wanneer een Big Tech-speler verzoekt om toegang tot bankdata, hij nu andere spelers inzicht moet geven in zijn eigen databestand en hen de mogelijkheid moet geven om toegang te verzoeken.

Als voorbeeld van het gebruik van het mededingingsrecht om een verplichting tot horizontale datadeling op te leggen (grijze vak), verwijzen we naar de casus inzake zoekmachines bij sectie 5.2. In deze casus wordt uitgelegd dat zoekmachinetechnologie wordt gekenmerkt door sterke leereffecten die het risico op een kanteling van de markt verhogen. Dit risico is mogelijk te verkleinen door onderlinge uitwisseling van (geanonimiseerde) zoekopdrachten en klikgedrag tussen zoekmachinebeheerders verplicht te stellen. De casus toont ook negatieve neveneffecten in de vorm van reverse engineering.

Bovengenoemde voorbeelden van John Deere en zoekmachines tonen de relevantie aan van de aanvullende maatregelen in het groene vak. Stel dat John Deere had besloten om een eigen app-ontwikkelingsafdeling te creëren zonder een datadelingsplatform. Stel dat deze afdeling zo succesvol wordt dat aanzienlijke marktmacht wordt verworven en dat andere tractorproducenten in feite worden gemarginaliseerd. In zulke gevallen is het waarschijnlijk dat concurrerende tractorproducenten toegang tot de data van John Deere kunnen afdwingen op basis van het mededingingsrecht, aangezien de data kunnen worden aangemerkt als essentiële input (of 'essentiële faciliteit', als algemeen gebruikte term in de antitrustwetgeving). Ten aanzien van de zoekcasus moet het mededingingsrecht in bredere zin worden toegepast om een einde te maken aan concurrentiebeperkende praktijken zoals koppeling van browsers aan besturingssystemen. Dit is in de eerste plaats nodig om ervoor te zorgen dat verplichte datadeling daadwerkelijk het evenwicht herstelt in de concurrentie op de zoekmarkt. Ten tweede maakt het zoekopdrachten en klikgedrag minder exclusief, waardoor een verplichting tot datadeling mogelijk overbodig wordt.

6.3 Organisatorische en technische aspecten

Datadeling vereist standaardisatie

Veel data-architecturen zijn uitsluitend ontworpen voor de uitvoering van interne processen, zonder interoperabiliteit met andere systemen. Desalniettemin is het ontwerp van veel onlineplatforms

geschikt voor datadeling en interoperabiliteit, doorgaans onder centraal beheer van een ecosysteem. Deelnemers moeten zich daarbij aan de heersende regels en normen houden.

Om data uitwisselbaar te maken binnen en tussen onlinediensten en ecosystemen, is niet per se een algehele herziening van bestaande data-infrastructuren nodig. Het is vaak voldoende om data-adapters te ontwikkelen om datasystemen interoperabel te maken¹⁵⁹. Er zijn hierbij geen technologische beperkingen (zoals werd bevestigd tijdens interviews). Het is echter lastig om standaardisatie op verschillende niveaus te realiseren. Wellicht moet de plicht tot het vaststellen van een standaard worden opgelegd en moeten de technische specificaties worden overgelaten aan marktspelers (zie hieronder).

Voor datadeling is technisch materiaal nodig dat bestaat uit API's en afspraken over hoe data worden uitgewisseld: welke data worden uitgewisseld, de taal, de datastructuur, protocollen, frequentie van actualisering enzovoort. Om dit te realiseren moeten voor datadeling standaarden worden ontwikkeld op de volgende niveaus:

- **Architectuurniveau:** Om een betrouwbare overdracht van data tussen dataleveranciers en datagebruikers te waarborgen, moet een technische specificatie worden opgesteld. Deze specificatie heeft betrekking op overdracht, ontvangst en berichtgeving, beschrijft de data die worden overgedragen, biedt de middelen om de data veilig over te dragen, en specificeert regels voor data-governance.
- **Taalniveau:** Net als natuurlijke taal is computertaal gebaseerd op *syntactische* en *semantische* standaarden.
 - Ten aanzien van data heeft syntaxis betrekking op de structuur, de vorm en de volgorde van de elementen. Als computer A bijvoorbeeld om een nummer vraagt en een letter ontvangt, is er sprake van een syntaxisfout. Syntaxisstandaarden zijn dus nodig om data te kunnen begrijpen.
 - Informatie kan syntactisch correct zijn, maar semantisch onduidelijk of onjuist, wat leidt tot vertaalfouten of uitvoeringsfouten. Of, in gewone taal, zinnen kunnen correcte woorden en grammatica bevatten, maar toch nergens op slaan (bijvoorbeeld "het roze idee juicht gelukkig"), geen betekenis hebben (bijvoorbeeld "de vrijgezel is getrouwd"), of hun betekenis kan verkeerd worden geïnterpreteerd. Om data correct te interpreteren, vooral in een sector-overschrijdende of internationale context, zijn semantische modellen nodig.
- **Prestatieniveau:** Voor datadeling zijn prestatie-eisen nodig. Het is absoluut noodzakelijk om de technische beperkingen van de dataoverdracht te begrijpen en prestatie-eisen te stellen die redelijk, proportioneel en doeltreffend zijn. Het voorbereiden van data (bijvoorbeeld transformatie, filteren, opschonen of verifiëren) of het ophalen van data, is bijvoorbeeld eenvoudiger zonder real-time portabiliteit en veel ingewikkelder in het geval van (bijna) real-time toegang. De juiste prestatie-eisen zijn afhankelijk van het specifieke gebruik en de toepassing.

Problemen bij het ontwikkelen van standaarden

Voor standaardisatie zijn afspraken tussen betrokken partijen vereist. Als de grootste partij verantwoordelijk is (of een dominante rol op zich neemt), kan dit het proces belemmeren, aangezien standaarden van invloed zijn op concurrentie. Samenwerking tussen belanghebbenden werkt waarschijnlijk het beste om standaarden op te stellen, mogelijk onder toezicht van een autoriteit of onder druk van een wettelijke verplichting om data uit te wisselen en standaarden te

¹⁵⁹ Opgemerkt wordt dat we interoperabiliteit op het niveau van de data laag bedoelen, zodanig dat data kunnen worden uitgewisseld. Dit is wat anders dan interoperabiliteit op dienstniveau, wat betekent dat diensten kunnen communiceren, bijvoorbeeld dat mensen een bericht plaatsen op Facebook terwijl ze LinkedIn gebruiken. De mate van standaardisatie die nodig is om interoperabiliteit op het niveau van de data laag te realiseren is over het algemeen veel minder gecompliceerd dan de mate van standaardisatie die nodig is om interoperabiliteit op dienstniveau te realiseren.

ontwikkelen.¹⁶⁰ Als gevolg van snelle technologische veranderingen is het opstellen van gedetailleerde specificaties van architectonische standaarden een ingewikkelde opgave. Over-specificatie van de kant van beleidsmakers kan leiden tot een ontoereikende oplossing, terwijl onder-specificatie verschillen in implementatie- en interoperabiliteitskwesties tot gevolg kan hebben.^{161, 162}

Een praktisch probleem wordt gevormd door de beperkingen die overleg met de Big Techs die op internationaal niveau opereren met zich meebrengen. Een klein land of een slecht georganiseerde sector heeft weinig mogelijkheden om dergelijke spelers in toom te houden. Europese wetgeving (met betrekking tot data-governance en -uitwisseling) versterkt de onderhandelingspositie van Europese overheden en bedrijven ten opzichte van internationale technologiebedrijven bij de ontwikkeling van Europese standaarden. De AVG dient hierbij als voorbeeld. Hoewel de AVG geen verplichting oplegt aan marktspelers om standaarden te ontwikkelen, zijn internationale technologiespelers na de invoering van de AVG in 2018 een samenwerkingsverband begonnen om bepaalde standaarden te ontwikkelen voor de uitwisseling van persoonsdata¹⁶³. Deze inspanningen dateren van vóór de AVG¹⁶⁴, maar het initiatief kreeg meer momentum toen Facebook, Microsoft, Twitter en Apple vanaf 2018 besloten deel te nemen. Sindsdien zijn de inspanningen gebundeld in het zogenoemde Data Transfer Project. Voor bedrijfsdata worden via openbare en particuliere initiatieven de eerste stappen gezet om te komen tot horizontale portabiliteit-overeenkomsten¹⁶⁵. Daarnaast worden op EU-niveau door SWIPO, een EU-werkgroep voor belanghebbenden, de eerste pogingen gedaan om een gedragscode op te stellen voor (het overzetten en overdragen van) data. Deze initiatieven zijn begonnen als zelfregulerende codes; door ze een dwingend karakter te geven kon verdere samenwerking op het gebied van standaarden worden gestimuleerd.

Uit deze ontwikkelingen blijkt nog eens dat er geen technische obstakels zijn waar het datadeling betreft. De obstakels zijn meestal van organisatorische aard. Ze kunnen worden aangepakt via wettelijke verplichtingen en economische stimulansen (zie ons voorstel met betrekking tot GDSD-principes, met name de voorwaarden inzake non-discriminatie en wederkerigheid).

Een andere praktische uitdaging is dat legacy-systemen de ontwikkeling van standaarden kunnen belemmeren. Dit blijkt uit een manifest van Nederlandse werkgeversorganisaties met betrekking tot zorginstellingen.¹⁶⁶ Het manifest wijst op efficiëntie- en privacy-problemen veroorzaakt door IT-systemen die geen data kunnen uitwisselen. Het onderliggende probleem was slechte en versnipperde implementatie van open standaarden.¹⁶⁷

Problemen zoals hierboven beschreven maken het lastig om standaarden voor sectoroverschrijdende datadeling te ontwikkelen. Sectorale data-architecturen zijn gemakkelijker te

¹⁶⁰ Zo verplichtte de Britse mededingings- en marktautoriteit de negen grootste Britse banken bijvoorbeeld om een gestandaardiseerde API te ontwikkelen om data uit te wisselen als onderdeel van het 'open banking' besluit (dat voorafging aan de PSD2-uitwisselingsverplichting) – zie Bijlsma, Van Der Cruisen, & Jonker (2020)

¹⁶¹ De Regulatory Technical Standards (RTS) for payment services is een voorbeeld waarbij de toezichthouder de implementatie aan de markt heeft overgelaten, en er zijn veel verschillen in de wijzen van implementatie. Zie: <https://hollandfintech.com/2019/04/dutch-psd2-bank-apis-diversity-in-rules-and-standardization/>.

¹⁶² Kaders voor datadeling kunnen ook worden gecreëerd door betrokkenheid van meerdere belanghebbenden. Zie <https://knowledge-base.connectedautomateddriving.eu/data-sharing/methodology-data-sharing-framework/>.

¹⁶³ <https://www.blog.google/outreach-initiatives/public-policy/our-preparations-europes-new-data-protection-law/>.

¹⁶⁴ Google werkt sinds 2007 aan een datamobiliteitssysteem - zie <https://opensource.googleblog.com/2018/07/introducing-data-transfer-project.html>.

¹⁶⁵ Onlangs hebben Adobe, Microsoft en SAP hun krachten gebundeld in een project dat open data initiative wordt genoemd en dat in silo's verzamelde bedrijfsapplicatiedata van de drie bedrijven harmoniseert en bundelt, en gebruikers toestaat om hun applicaties te gebruiken en erop voort te bouwen.

¹⁶⁶ MKB Nederland & VNO-NCW. (2019). Samen vooruit: datadeling een ambitie voor in de zorg. Beschikbaar op: https://www.mkb.nl/sites/default/files/nln19085_manifest_samen_vooruit_05.pdf

¹⁶⁷ Recente technologie biedt echter een oplossing voor dit probleem. Zie bijvoorbeeld de technologie van Ledger Leopard, dat ook samenwerkt met de Nederlandse gezondheidsautoriteiten om dit probleem aan te pakken. <https://ledgerleopard.com/industry/healthcare/>

realiseren, maar evolueren meestal binnen informatiesilo's. Het is een enorme uitdaging om standaarden te ontwikkelen die data-integratie tussen heterogene data-architecturen realiseren.

De noodzaak om wettelijk verplichte standaarden voor te schrijven met betrekking tot het voorgestelde beleid

Ten aanzien van beleidsdoelstelling I (t.w. data mobiliseren en innovatie stimuleren), kan het nodig zijn een verplichting op te leggen om een open standaard te hanteren. Indien zo'n standaard nog niet bestaat, betekent dit dat marktpartijen er een moeten ontwikkelen. Tegelijkertijd hoeven noch algemene wetgeving noch sectorale verplichtingen het ontwerp van standaarden bepalen, behalve om te specificeren welke soorten data moeten worden uitgesloten (zoals verkregen, afgeleide en geproduceerde data), en mogelijk om noodzakelijke functionaliteiten te specificeren die scheppers controle over hun data geven. Het standaardisatieproces kan vervolgens worden overgelaten aan belanghebbenden, eventueel onder toezicht van of gefaciliteerd door een onafhankelijke externe partij, en binnen een eventueel wettelijk vast te leggen tijdsbestek. Meer specifieke eisen kunnen de doeltreffendheid van de standaarden ondermijnen, aangezien ze de innovatieve inspanningen die verband houden met het ontwerpen van standaarden zouden kunnen verstoren.

Ten aanzien van beleidsdoelstelling II (t.w. aanpak van concurrentieproblemen die voortkomen uit leer- en netwerkeffecten) is het noodzakelijk om specifiek te zijn over syntactische en semantische standaarden (t.w. soorten data die, wanneer ze worden uitgewisseld, het onderliggende probleem aanpakken) en over prestatie-eisen.

7 Bibliografie

- Adar, E. (2007). User 4XXXXX9: Anonymizing Query Logs. *presented at the World Wide Web Conference*. Retrieved from <http://www.cond.org/anonlogs.pdf>
- Agrawal, A., Gans, J., & Goldfarb, A. (2019). Economic policy for artificial intelligence. *Innovation Policy and the Economy*, 19(1), 139-159.
- Anderson, J. (2020, February 11). *The dynamics of data accumulation*. Retrieved from Bruegel Blog: <https://bruegel.org/2020/02/the-dynamics-of-data-accumulation/>
- Argenton, C., & Prüfer, J. (2012). Search Engine Competition with Network Externalities. *Journal of Competition Law & Economics*, 73-105.
- Bekkers, R., Birkman, L., Canoy, M., De Bas, P., Lemstra, W., Ménière, Y., . . . Verbeek, A. (2014). *Patents and standards: a modern framework for IPR-based standardisation*. Rotterdam/Brussels: Ecorys, commissioned by the European Commission.
- Bijlsma, M., Van Der Crujisen, C., & Jonker, N. (2020, July 23). Betaaldienstenrichtlijn heeft nog geen revolutie ontketend. *Economische Statistische Berichten*. Retrieved from <https://esb.nu/esb/20060101/betaaldienstenrichtlijn-heeft-nog-geen-revolutie-ontketend>
- Choi, J. P., & Jeon, D. S. (2016). *A Leverage Theory of Tying in Two-Sided Markets*. Working Paper.
- Choudary, S. P. (2015). *Platform Scale: How an emerging business model helps startups build large empires with minimum investment*. San Francisco: Platform Thinking Labs.
- Cloud Security Agency. (2014, September). *Big Data Taxonomy*. Retrieved from [cloudsecurityalliance.org: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Taxonomy.pdf](https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Taxonomy.pdf)
- Competition and Market Authority. (2020). *Online platforms and digital advertising*. London: CMA. Retrieved from https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final_report_1_July_2020_.pdf
- Crémer, J., de Montjoye, Y., & Schweitzer, H. (2019). *Competition policy for the digital era*. European Commission, a study commissioned by DG Competition.
- De Bas, P., Batura, O., Yagafarova, A., Van Gorp, N., ..., & Pichler, E. (2017). *Business-to-Business relations in the online platform environment*. Rotterdam: Ecorys/WIFO commissioned by the European Commission DG Connect.
- De Bijl, P., & Van Gorp, N. (2016-2019). Gastcolleges in Digital Business Models and Strategy. *Modular Executive MBA in Business & IT: Digital Strategy and Transformation*. Breukelen: Nyenrode Univercity.
- De Bijl, P., & Van Gorp, N. (2018, December). Publieke belangen en de regulering van digitale platformen. *Economische Statistische Berichten (ESB)*, 103(4768S), 63-38.
- De Bijl, P., & Werner, G. (2019). Platformbedrijven als marktmeesters. *TPEdigitaal*(jaargang 13(2)), 33-49.
- De Standaard. (2017, 05 11). *Zo kwam Snap aan 2 miljard dollar verlies*. Retrieved 12 31, 2019, from [standaard.be: https://www.standaard.be/cnt/dmf20170511_02876608](https://www.standaard.be/cnt/dmf20170511_02876608)
- Deahl, R. (2019, August 8). *How to Write a Compelling Profile of a Person*. Retrieved from [thebalancecareers.com: https://www.thebalancecareers.com/writing-a-compelling-profile-of-a-person-2316038](https://www.thebalancecareers.com/writing-a-compelling-profile-of-a-person-2316038)
- Deloitte. (2013, July 24). *Data as the new currency*. Retrieved from [Deloitte.com: https://www2.deloitte.com/us/en/insights/deloitte-review/issue-13/data-as-the-new-currency.html#endnote-sup-1](https://www2.deloitte.com/us/en/insights/deloitte-review/issue-13/data-as-the-new-currency.html#endnote-sup-1)
- Digital Transport & Logistics Forum. (2018). *Data sharing in supply and logistics as commodity*. Retrieved april 6, 2020, from [ec.europa.eu: https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&doid=14275](https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&doid=14275)
- Ecorys. (2008). *A study on EU spending*. Rotterdam: European Commission.
- Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform Envelopment. *Strategic Management Journal*, 32(12), 1270-1285.
- European Commission. (2020, februari 19). *A European strategy for data*. COM(2020) 66 final. Retrieved from [ec.europa.eu: https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy](https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy)

- Evans, P., & Gawer, A. (2016). The Rise of the Platform Enterprise: A Global Survey. *The Emerging Platform Economy Series, Vol. 1*. Retrieved from http://thecege.net/wp-content/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf
- Ezrachi, A., & Stucke, M. (2017). Artificial Intelligence & Collusion: When Computers Inhibit Competition. *University of Illinois Law Review*, 1775-1810.
- FCC. (2010). *2010 open internet order*. FCC 10-201.
- Federal Trade Commission. (2016). *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*. Washington: FTC Report. Retrieved from <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>
- Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources*. Oxford: Oxford University.
- Furman, J., Coyle, D., Fletcher, A., McAuley, D., & Marsden, P. (2019). *Unlocking digital competition*. Report of the Digital Competition Expert Panel.
- Gawer, A. (2016). *Competition Policy and Regulatory Reforms for Big Data: Propositions to Harness the Power of Big Data while Curbing Platforms' Abuse of Dominance*. Organisation for Economic Co-operation and Development.
- GSMA. (2018). *Connected Women, The Gender Analysis & Identification Toolkit Estimating subscriber gender using machine learning*. London: GSM Association. Retrieved from <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/09/GSMA-Gender-Analysis-and-Identification-Report-GAIT-August-2018.pdf>
- Hagiu, A., & Wright, J. (2020). When data creates competitive advantage. *Harvard Business Review*, 98(1), 94-101.
- He, D., Kannan, A., Liu, T. Y., McAfee, R. P., Qin, T., & Rao, J. M. (2017). Scale Effects in Web Search. *International Conference on Web and Internet Economics* (pp. 294-310). Springer, Cham.
- Hogendorn, C., & Frischmann, B. (2020). Infrastructure and general purpose technologies: a technology flow framework. *European Journal of Law and Economics*. Retrieved from <https://doi.org/10.1007/s10657-020-09642-w>
- IBM. (2020, March 12). *What is Derived Data*. Retrieved from IBM Knowledge Center: https://www.ibm.com/support/knowledgecenter/SSQVNT_3.8.0/ilog.odms.ide.odm.enterprise.help/ODME/Userman/topics/Creating/ADM_AdvDerivedData.html
- ISO/IEC. (2017, September). *Information technology — Cloud computing — Cloud services and devices: Data flow, data categories and data use*. Retrieved from iso.org: https://standards.iso.org/ittf/PubliclyAvailableStandards/c066674_ISO_IEC_19944_2017.zip
- Katz, M. L., & Sallet, J. (2018). Multisided Platforms and Antitrust Enforcement. *Yale Law Journal*, 127(7), 2142-2172. Retrieved from <https://ssrn.com/abstract=3131933>
- Kessel, J. M., & Mozur, P. (2016, augustus 9). *How China is changing the internet*. Retrieved from New York Times: <https://www.nytimes.com/video/technology/100000004574648/china-internet-wechat.html>
- Mano, M. d., & Padilla, J. (2018). Big Tech Banking. *Journal of Competition Law & Economics*, 14(4), 494-526.
- Mayer-Schönberg, V., & Ramge, T. (2018). A Big Choice for Big Tech, Share Data or Suffer the Consequences. *Foreign Affairs*, 97(5), 48-54.
- Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A Survey on Machine Learning for Data Fusion. *Information Fusion*, 2020(57), 115-129.
- Nerdwriter1. (2018, March 28). *How Dark Patterns Trick You Online*. Retrieved from YouTube.com: <https://youtu.be/kxkrdLI6e6M>
- Nooren, P., Van Gorp, N., Van Eijk, N., & Fathaigh, R. Ó. (2018). Should we regulate digital platforms? A new framework for Evaluating policy options. *Policy & Internet*, 10(3), 264-301.
- OECD. (2014). *Summary of the OECD privacy expert roundtable, Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking*. Working Party on Security and Privacy in the Digital Economy. DSTI/ICCP/REG(2014)3: OECD.
- Opara-Martins, J., Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. *Journal of Cloud Computing*, 5(1). Retrieved from <https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-016-0054-z>
- Pariser, E. (2011, May 2). *Beware online "filter bubbles"*. Retrieved from YouTube: <https://youtu.be/B8ofWFx525s>

- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: how networked markets are transforming the economy and how to make them work for you*. New York City: WW Norton & Company.
- Petit, N. (2018). American tech giants are fiercely competitive monopolies. *ESB*, 103(4768S).
- Prüfer, J. (2019). *Competition Policy and Data Sharing on Data-driven Markets; Steps Towards Legal Implementation*. Bonn: A Friedrich-Ebert-Stiftung.
- Prüfer, J., & Schottmüller, C. (2017). Competing with Big Data. *TILEC Discussion Paper No. 2017-006; CentER Discussion Paper 2017-007*.
- Radicand Economics. (2018). *A review of ACM's game theory analysis*. Woerden: VodafoneZiggo.
- Retail 90. (2017, September 14). *Winning the Amazon Buy Box: What You Need to Know...* Retrieved from YouTube.com: <https://www.youtube.com/watch?v=9jOSLkCsAik>
- Rocher, L., Hendrickx, J., & De Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(3069). Retrieved from <https://doi.org/10.1038/s41467-019-10933-3>
- Rogers, D. L. (2016). *The digital transformation playbook: Rethink your business for the digital age*. New York: Columbia University Press.
- Schweitzer, H., Haucap, J., Kerber, W., & Welker, R. (2018). *Modernising the law on abuse of market power*. German Federal Ministry for Economic Affairs and Energy. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3250742
- Scott, M. (2020, march 3). *Coronavirus crisis shows Big Tech for what it is — a 21st century public utility*. Retrieved from Politico: <https://www.politico.eu/pro/coronavirus-big-tech-utility-google-facebook/>
- Senellart, P. (2016, September 15). What do Internet companies know about you. *Data & Competition Policy*. Brussel: CERRE Private Seminar . Retrieved from <https://pierre.senellart.com/talks/cerre-20160915.pdf>
- Stiglitz, J. E., & Rosengard, J. K. (2015). *Economics of the public sector* (Fourth international student edition ed.). New York: WW Norton & Company.
- The Economist. (2017, May). The world's most valuable resource is no longer oil, but data. *The World's Most Valuable Resource – Data and the new rules of competition*.
- The Economist. (2020). Are data more like oil or sunlight? *Special report on the data economy: the data economy*.
- The Guardian. (2016, november 2). *Facebook forces Admiral to pull plan to price car insurance based on posts*. Retrieved april 6, 2020, from theguardian.com: <https://www.theguardian.com/money/2016/nov/02/facebook-admiral-car-insurance-privacy-data>
- The Verge. (2016, november 2). *Facebook blocks insurer exploiting user data to find 'conscientious' drivers*. Retrieved april 6, 2020, from theverge.com: <https://www.theverge.com/2016/11/2/13496316/facebook-blocks-car-insurer-from-using-user-data-to-set-insurance-rate>
- TNO, Ecorys, & IViR. (2015). *Digital platforms: an analytical framework for identifying and evaluating policy options*. The Hague: TNO, commissioned by the Dutch Ministry of Economic Affairs.
- TNO, Ecorys, & IViR. (2015). *Digital Platforms: an analytical framework for identifying and evaluating policy options*. Den Haag: TNO, in opdracht van het Ministerie van Economische Zaken.
- U.K. Parliament. (2019). *Disinformation and 'fake news': Final Report*. London: the Digital, Culture, Media and Sport Committee of the British Parliament.
- UK CMA. (2019). Appendix E: The role of data. In *Online platforms and digital advertising market study* (pp. E1-E44). London: UK CMA. Retrieved from https://assets.publishing.service.gov.uk/media/5df9ecc040f0b609402e2838/Appendix_E_The_role_of_data.pdf
- United Nations Statistical Commission and Economic Commission for Europe. (2000). Terminology on Statistical Metadata. *CONFERENCE OF EUROPEAN STATISTICIANS STATISTICAL STANDARDS AND STUDIES – No. 53*. Geneva: United Nations. Retrieved from https://ec.europa.eu/eurostat/ramon/coded_files/UNECE_TERMINOLOGY_STAT_METADATA_2000_EN.pdf
- Van Gorp, N. (2018). Data en marktmacht: privacy is de oplossing, niet het probleem. *Tijdschrift voor Toezicht*(nr. 2-3), 88-89.
- Van Gorp, N., & Batura, O. (2015). *Challenges for competition policy in the digitalised economy*. Rotterdam: Ecorys, i.o.v. het Europese Parlement.

- Van Gorp, N., & De Bijl, P. (2019). *Digital Gatekeepers, assessing exclusionary conduct*. Zevenhuizen: e-Conomics, in opdracht van het Ministerie van Economische Zaken en Klimaat. Retrieved from <https://www.government.nl/documents/reports/2019/10/07/digital-gatekeepers>
- Van Gorp, N., Graef, I., Lemstra, W., Stokking, H., Nooren, P., Billeskov, K., . . . Valcke, P. (2016). *Study on future trends and business models in communication services*. Rotterdam: Ecorys, commissioned by the European Commission, DG Connect.
- Van Gorp, N., Graef, I., Lemstra, W., Stokking, H., Nooren, P., Billeskov, K., . . . Valcke, P. (2016). *Study on future trends and business models in communication services*. Rotterdam: Ecorys, commissioned by the European Commission, DG Connect.
- Van Til, H., Van Gorp, N., & Price, K. (2017). *Big Data and Competition*. Rotterdam: Ecorys, in opdracht van het Ministerie van Economische Zaken. Retrieved from <https://zoek.officielebekendmakingen.nl/blg-813928.pdf>
- World Economic Forum. (2014, Mei). *Rethinking Personal Data: A New Lens for Strengthening Trust*. Retrieved from http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf
- Wright, J. (2004). One-sided Logic in Two-sided Markets. *Review of Network Economics*, (3(1)), 44-64.

8 Bijlagen: casestudies

Zie separaat document.

Over Ecorys

Ecorys is een toonaangevend internationaal onderzoeks- en adviesbureau dat de belangrijkste uitdagingen van de samenleving aanpakt. Met op onderzoek gebaseerd advies van wereldklasse helpen we publieke en private klanten om weloverwogen beslissingen te nemen en te implementeren die leiden tot een positieve impact op de samenleving. Wij ondersteunen onze opdrachtgevers met gedegen analyse en inspirerende ideeën, praktische oplossingen en oplevering van projecten voor complexe markt-, beleids- en managementvraagstukken.

In 1929 richtten zakenlieden van de huidige Erasmus Universiteit Rotterdam het Nederlands Economisch Instituut (NEI) op. Het doel was om de tegengestelde werelden van economisch onderzoek en zaken te overbruggen - in 2000 werd dit gerespecteerde instituut Ecorys.

Door de jaren heen breidde Ecorys zich uit over de hele wereld, met kantoren in Europa, Afrika, het Midden-Oosten en Azië. Onze medewerkers zijn afkomstig uit veel verschillende culturele achtergronden en vakgebieden omdat we geloven in de kracht die verschillende perspectieven bieden aan onze organisatie en onze klanten.

Ecorys blinkt uit in zeven expertises:

- Economische groei;
- Sociaal beleid;
- Natuurlijke bronnen;
- Regio's en steden;
- Vervoersinfrastructuur;
- hervorming van de publieke sector;
- Veiligheid en justitie.

Ecorys biedt een duidelijk pakket producten en diensten:

- voorbereiding en formulering van beleid;
- programmamanagement;
- communicatie;
- capaciteit opbouwen;
- monitoring en evaluatie.

We waarderen onze onafhankelijkheid, onze integriteit en onze partners. We geven om de omgeving waarin we werken en leven. We voeren een actief beleid op het gebied van maatschappelijk verantwoord ondernemen, gericht op het creëren van gedeelde waarde die de samenleving en het bedrijfsleven ten goede komt. We zijn ISO 14001 gecertificeerd, ondersteund door al onze medewerkers.



P.O. Box 4175
3006 AD Rotterdam
The Netherlands

Watermanweg 44
3067 GG Rotterdam
The Netherlands

T +31 (0)10 453 88 00
F +31 (0)10 453 07 68
E netherlands@ecorys.com
Registration no. 24316726

W www.ecorys.nl

Sound analysis, inspiring ideas