De invloed van (technische) ontwikkelingen op het begrip persoonsgegevens in relatie tot de AVG

Bart van der Sloot, Sascha van Schendel & César Augusto Fontanillo López

Tilburg University

Understanding Society

De invloed van (technische) ontwikkelingen op het begrip persoonsgegevens in relatie tot de AVG

Uitgevoerd door
Tilburg Insititute for Law, Technology and Society van Tilburg University

Auteurs
Bart van der Sloot, Sascha van Schendel & César Augusto Fontanillo López

Opdrachtgever
Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC)

Inhoudsopgave

Introductie

De Algemene Verordening Gegevensbescherming (AVG) is misschien wel het belangrijkste kader voor het digitale domein in Europa en daarbuiten. De AVG stelt regels aan- en bevat normen voor de verwerking van gegevens, legt verplichtingen vast voor personen en organisaties die gegevens verwerken (verwerkingsverantwoordelijken) en kent rechten toe aan personen van wie gegevens worden verwerkt (betrokkenen). Hoewel pas in 2016 aangenomen, stammen de regels in essentie uit de jaren 70 van de vorige eeuw. Doorslaggevend voor de toepassing van het gegevensbeschermingskader was toen, en is vandaag de dag nog steeds, of de gegevens die worden verwerkt informatie van een identificeerbaar individu (natuurlijke persoon) betreffen.

Hoewel een dergelijke vaststelling in de jaren 70 van de vorige eeuw relatief eenvoudig was, is die in de loop van de tijd steeds complexer geworden, vooral in het licht van technologische ontwikkelingen, de algemene toegankelijkheid van technologieën en het streven naar meer open data. Deze ontwikkelingen hebben tot gevolg dat het steeds makkelijker is om persoonsgegevens af te leiden uit datasets die dergelijke gegevens op het eerste gezicht niet lijken te bevatten. Ze hebben ook tot gevolg dat de juridische status van data steeds meer fluïde wordt: doordat data worden gedeeld tussen partijen en de verwerkingen van datasets aanzienlijk verschillen, kan dezelfde dataset het ene moment worden gekwalificeerd als persoonsgegevens en het andere moment niet, of als persoonsgegevens in handen van partij A maar tegelijkertijd als geen persoonsgegevens in handen van partij B.

Daarom is in de loop der tijd in het wettelijke kader het begrip persoonsgegevens uitgebreid. Met name in 1995 breidde de voorloper van de Algemene Verordening Gegevensbescherming, de Richtlijn Gegevensbescherming, de reikwijdte van dit begrip aanzienlijk uit en daarmee ook het aantal datasets dat onder het bereik van het gegevensbeschermingsregime viel. Bij persoonsgegevens gaat het niet alleen om directe maar ook om indirecte informatie, dat wil zeggen gegevens, zoals beschrijvingen, waaruit de identiteit van een persoon kan worden afgeleid. Bij persoonsgegevens gaat het niet alleen om identificerende gegevens, dat wil zeggen gegevens die op dit moment tot een bepaalde persoon kunnen leiden, maar ook identificeerbare gegevens, of met andere woorden gegevens die op dit moment niet tot een bepaalde persoon leiden, maar in de toekomst mogelijk wel. Om te bepalen of een dataset identificeerbare gegevens bevat, moet rekening worden gehouden met alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruikt om gegevens aan een persoon te koppelen. Ten slotte is het niet nodig om de identiteit van een persoon te kennen; als gegevens worden gebruikt om een beslissing te nemen over een specifieke persoon wiens identiteit onbekend is, is het gegevensbeschermingsregime ook van toepassing.

Deze wetswijzigingen hebben geleid tot een substantiële uitbreiding van het toepassingsgebied van het gegevensbeschermingsregime. Tegelijkertijd blijft het begrip persoonsgegevens de bepalende factor bij de beslissing of de gegevensbeschermingsregels van toepassing zijn. In tegenstelling tot het restrictieve regime voor de verwerking van persoonsgegevens heeft de Europese Unie (EU) een ander kader vastgesteld voor de verwerking van niet-persoonsgegevens. De Verordening betreffende het vrije verkeer van niet-persoonsgegevens houdt in wezen in dat er geen beperkingen mogen worden gesteld, noch door de publieke sector, noch door de private sector, met betrekking tot het vrije verkeer van niet-persoonsgegevens. De juridische kwalificatie of een dataset al dan niet persoonsgegevens bevat betekent dus dat er een reguleringskader van bijna 180 graden verschil van toepassing is (hoewel de voorgestelde ‘Data Governance Act’ de zaken nog ingewikkelder kan maken).

Er zijn ook belangrijke technologische en maatschappelijke ontwikkelingen. Big Data, Kunstmatige Intelligentie, Quantum Computing en andere technieken maken het nog gemakkelijker om persoonsgegevens af te leiden uit geaggregeerde, geanonimiseerde of versleutelde datasets; de algemene toegankelijkheid van technologieën maakt het nog moeilijker om de toekomstige status van een dataset te bepalen; en het voortdurende streven naar open data en het hergebruik van overheidsinformatie

betekent dat de juridische status van data nog meer fluïde zal worden. In het licht van deze nieuwe uitdagingen is het de vraag hoe het juridische regime hierop moet reageren. Moet het begrip persoonsgegevens verder worden opgerekt? Zo ja, zou dat in de praktijk niet betekenen dat alle gegevens als persoonsgegevens worden aangemerkt? Moet het huidige onderscheid tussen persoonsgegevens en niet-persoonsgegevens behouden blijven, of moet er een restrictiever regime komen voor niet-persoonsgegevens? En wat betekenen deze ontwikkelingen voor andere gegevenscategorieën in de Algemene Verordening Gegevensbescherming, zoals pseudonieme gegevens en gevoelige (bijzondere) persoonsgegevens?

Tegen deze achtergrond is de onderzoeksvraag voor dit onderzoek: *Welk effect hebben huidige en toekomstige technische ontwikkelingen op het gebied van anonimisering, pseudonimisering, aggregatie en identificatie van gegevens, op het gegevensbeschermingskader en de bescherming van de verschillende soorten gegevens?*

De deelvragen, die helpen bij het beantwoorden van deze onderzoeksvraag, zijn:

Identificeerbaarheid van gegevens

1. Welke (technische) middelen zijn er om (anonieme) data terug te koppelen aan individuen, en in hoeverre speelt de beschikbaarheid van andere (bijvoorbeeld open source) data een rol?
2. Welke (technische) ontwikkelingen worden de komende jaren verwacht met betrekking tot de middelen om gegevens (al dan niet opzettelijk) terug te koppelen aan personen?

Anonimisering en pseudonimisering van gegevens

3. Welke huidige en voorzienbare technische ontwikkelingen kunnen worden gebruikt voor het anonimiseren of pseudonimiseren van persoonsgegevens en welke factoren zijn daarbij bepalend?
4. Welke technische ontwikkelingen op het gebied van anonimisering en pseudonimisering van persoonsgegevens zijn de komende jaren te verwachten?

Identificeerbaarheid in relatie tot anonimisering en pseudonimisering

5. Wat kan er vanuit een juridisch en technisch perspectief worden gezegd over de invulling van het begrip ‘alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruikt’? Welke middelen kunnen redelijkerwijs worden gebruikt en welke factoren spelen daarbij een rol?
6. Hoe verhoudt het antwoord op vraag 5 zich tot ontwikkelingen in huidige en toekomstige anonimiserings- en pseudonimiseringstechnieken?
7. Wanneer is het redelijk om te zeggen dat gegevens niet meer terug te koppelen zijn aan een persoon en dat de dataset waarvan ze deel uitmaken als anoniem kan worden beschouwd?
8. In hoeverre is de test op indirecte identificeerbaarheid objectiveerbaar?

Gevolgen van identificeerbaarheid en anonimisering en pseudonimisering

9. In hoeverre en in welke gevallen kan er sprake zijn van onderregulering wanneer gegevens door anonimisering niet meer aan personen kunnen worden gekoppeld en dus niet binnen de reikwijdte van de Algemene Verordening Gegevensbescherming vallen?
10. In welke mate en in welke gevallen kan er sprake zijn van overregulering wanneer steeds meer gegevens eenvoudig aan individuen kunnen worden gekoppeld door middel van nieuwe technieken (het ongedaan maken van anonimisering en pseudonimisering)?

Overkoepelende analyse

11. Hoe zullen de huidige en toekomstige technische ontwikkelingen de komende periode van invloed zijn op de AVG en rechtsbescherming in brede zin?

Voor het beantwoorden van deelvragen 1-8 zijn verschillende aspecten relevant:

- de verschillende juridische begrippen en de criteria in de definities en afbakeningen;
- de beschikbaarheid van (open access) data en van dataverwerkingstechnologieën; in dit opzicht is het streven van de Europese Unie naar open data en hergebruik van (overheids)data relevant;
- de huidige en toekomstige technologische middelen voor het anonimiseren en de-

anonimiseren, aggregeren en de-aggregeren, pseudonimiseren en de-pseudonimiseren van gegevens; en
- de impact van de ontwikkelende technologische mogelijkheden en het uitbreidende datalandschap op de houdbaarheid van huidige juridische concepten en afbakeningen.

Om deelvragen 9-11 te beantwoorden, zijn verschillende aspecten relevant:
- de reguleringsdoelstelling van het gegevensbeschermingskader en daarmee het licht waarin het gevaar van zowel onder- als overregulering beoordeeld moet worden;
- de lacunes in de regelgeving die voortkomen uit de kloof tussen het juridische en het technologische domein; en
- de alternatieven voor het huidige wettelijke kader die uit voorgangers van de Europese wetgeving en uit wetsvoorstellen, literatuur en interviews kunnen worden gedistilleerd.

Bij de beantwoording van de vragen 9-11, en om te bepalen of er sprake is van onder- en/of overregulering, moet worden bepaald wat het reguleringsdoeleinde van de AVG is en zou moeten zijn. Daarbij moeten twee zaken worden onderzocht. Enerzijds is het de vraag of het gegevensbeschermingsrecht als enige of belangrijkste doel heeft om natuurlijke personen te beschermen. Verschillende auteurs wijzen erop dat de wetgeving inzake gegevensbescherming, althans aanvankelijk, vooral gericht was op de bescherming van objectieve rechtsbeginselen en algemene belangen. Anderzijds wordt in de juridische literatuur bediscussieerd in hoeverre de bescherming van natuurlijke personen de beste basis is voor toekomstige regelgeving en of deze bescherming niet moet worden uitgebreid naar groepen of de samenleving als geheel.

Voor dit onderzoek worden drie methoden ingezet:

1. Doctrinaire en juridische analyse: vier juridische onderscheiden tussen gegevens staan centraal in dit onderzoek, namelijk het onderscheid tussen: anonieme gegevens en persoonsgegevens, geaggregeerde of statistische gegevens en persoonsgegevens, pseudonieme en niet-pseudonieme persoonsgegevens en niet-gevoelige en gevoelige persoonsgegevens. Hiervoor worden de wetten van de EU en de Raad van Europa (RvE), hun wetsgeschiedenis en juridische interpretatie bestudeerd.
2. Literatuuroverzicht.
    a. Beschrijvende literatuur: technische literatuur over (de-)identificatietechnologieën en privacy/gegevensbescherming verbeterende technieken wordt bekeken.
    b. Normatieve literatuur: juridische en reguleringsliteratuur wordt bestudeerd die de uitdagingen van elke categorie gegevens beschrijft en/of nieuwe definities, perspectieven of benaderingen voor de verschillende soorten gegevens voorstelt.
3. Kwalitatieve onderzoeksmethoden.
    a. Interviews: er zijn interviews gehouden met experts met verschillende achtergronden en expertisegebieden.
    b. Workshop: aan het begin van dit onderzoek is een workshop gehouden om problemen en mismatches tussen het juridische en beleidsdomein enerzijds en de technische en praktische realiteit anderzijds te identificeren.

Het onderzoek liep langs de volgende lijnen.

Het wettelijk regime is beoordeeld op drie punten:

(1) Het huidige wettelijke regime en de bestaande definities en uitleg daarvan in literatuur of gezaghebbende adviezen zijn geanalyseerd om te bepalen hoe het bestaande wetgevend kader gegevensverwerking beoordeelt.
(2) De geschiedenis van het juridisch regime vanuit het oogpunt van de definities is om drie redenen geëvalueerd. Ten eerste laat het zien hoe het kader voor gegevensbescherming in de loop van de tijd is gewijzigd in reactie op maatschappelijke en technologische veranderingen.

Ten tweede geeft het inzicht in de logica en beweegredenen achter de huidige definities en categorisering: waarom zijn de definities zoals ze zijn en welk doel wordt nagestreefd. Meer in het algemeen werd aandacht besteed aan de discussie over de achterliggende gedachte van het gegevensbeschermingskader, aangezien dit relevant is met het oog op mogelijke toekomstige wijzigingen in het gegevensbeschermingskader. Ten derde kunnen door de verschillende definities en afbakeningen van de gegevenscategorieën en vooral de variaties die in de wetsgeschiedenis zijn besproken en overwogen, maar werden verworpen, alternatieve manieren worden gevonden om de regulering van gegevens aan te pakken.

(3) De potentiële toekomst van het gegevensbeschermingskader werd beoordeeld. De in dit onderzoek besproken technologische en maatschappelijke ontwikkelingen hebben grote invloed op de invulling en effecten van het huidige regulerende kader. Daarom wordt een overzicht gegeven van de belangrijkste ideeën voor mogelijkheden om het huidige regulerende kader te wijzigen.

Het technologische domein werd beoordeeld op drie punten.

(1) Om het beeld te schetsen van een veld dat voortdurend in beweging is, is een kort overzicht gegeven van de technologische ontwikkelingen na de Tweede Wereldoorlog. Deze beschrijving geeft de achtergrond waartegen het wettelijk kader in de loop van de tijd is gewijzigd.

(2) Het onderzoek heeft de huidige technologieën beoordeeld, met name in het licht van de verschillende juridische gegevenscategorieën en de grenzen daartussen. Deze beschrijving laat zien dat het steeds beter mogelijk wordt een dataset te de-anonimiseren en (gevoelige) persoonsgegevens af te leiden uit één of meer geaggregeerde datasets.

(3) Het onderzoek beschrijft technologische ontwikkelingen die het landschap in de toekomst mogelijk nog verder zullen veranderen. Hieruit blijkt dat de scheidslijnen tussen de verschillende juridische gegevenscategorieën zo mogelijk nog meer zullen vervagen.

Ook is er aandacht besteed aan twee maatschappelijke ontwikkelingen (hoewel deze zowel door juridische als technologische ontwikkelingen zijn ingegeven):

(1) De studie beschrijft hoe technologieën in de loop van de tijd algemeen beschikbaar zijn geworden. Hierdoor beschikken steeds meer overheidsorganisaties, bedrijven en zelfs burgers over zeer geavanceerde technologische middelen. Het gevolg van deze trend is dat als data tussen verschillende partijen worden gedeeld of openbaar worden gemaakt, het steeds waarschijnlijker wordt dat er een partij is die de juridische status van de dataset verandert.

(2) Het onderzoek verwijst kort naar de juridische en maatschappelijke druk om gegevens openbaar te maken. Dit betreft voornamelijk statistische gegevens, overheidsinformatie en niet-persoonsgegevens. Meestal zullen deze datasets op zichzelf geen persoonsgegevens bevatten, maar in combinatie met andere datasets kunnen ze worden gebruikt om (gevoelige) persoonsgegevens te genereren. Bovendien is het, gezien de vooruitgang en de algemene toegankelijkheid van technologieën, steeds waarschijnlijker dat er een partij zal zijn die voldoende middelen zal investeren om een dataset te de-anonimiseren of opnieuw te identificeren.

Dit onderzoek buigt zich over vier juridische gegevenscategorieën die zijn verankerd in de Algemene Verordening Gegevensbescherming, naast persoonsgegevens zijn er: geanonimiseerde gegevens, geaggregeerde of statistische gegevens, gepseudonimiseerde persoonsgegevens en gevoelige persoonsgegevens. Hieronder wordt een samenvatting gegeven van de belangrijkste bevindingen op de volgende punten: (1) de huidige regulering van de verschillende gegevenscategorieën; (2) de twee, soms tegenstrijdige, benaderingen van gegevensregulering die door het kader voor gegevensbescherming lopen; (3) de algemene toegankelijkheid van technologieën en het streven naar open data en het hergebruik van overheidsinformatie; (4) de impact van het veranderende technologische landschap op de regulering van data; (5) de lacunes die bestaan tussen het huidige reguleringsstelsel en de veranderende technologische realiteit; (6) de alternatieven voor het huidige reguleringsregime die in de

literatuur en elders worden gesuggereerd om deze lacunes te dichten; (7) de overkoepelende reguleringsdoelstelling van het gegevensbeschermingskader in het licht waarvan mogelijke wijzigingen moeten worden beoordeeld; (8) de gevaren van over- en onderregulering veroorzaakt door de mismatch tussen het juridische en het technologische domein; (9) en de mogelijke manieren om de bestaande hiaten tussen de twee domeinen op te lossen. Tot slot (10) worden de onderzoeksvraag en deelvragen beantwoord.

1. Juridische categorieën en de elementen daarvan

Dit onderzoek heeft zich gericht op vier gegevenscategorieën onder het gegevensbeschermingsregime. Naast persoonsgegevens beschouwt het onderzoek anonieme gegevens, geaggregeerde of statistische gegevens, pseudonieme persoonsgegevens en gevoelige persoonsgegevens.

Anonieme gegevens

In dit onderzoek is de scheidsgrens tussen persoonsgegevens en anonieme gegevens onderzocht. Anonimiseren betekent het wegnemen van direct of indirect geïdentificeerde of identificeerbare data in data. Als gegevens correct geanonimiseerd zijn is de AVG niet, maar de Verordening vrij verkeer van niet-persoonsgegevens wel van toepassing. Uit de formele definitie van persoonsgegevens (artikel 4 lid 1 AVG), de relevante overwegingen (14, 26, 27 en 30), en de interpretatie door het Hof van Justitie van de Europese Unie (HvJ EU) en de Artikel 29 Werkgroep, vallen minstens vier punten op te maken:

1. Niet alleen direct identificerende gegevens maar ook indirect identificerende gegevens, en niet alleen identificerende gegevens maar ook identificeerbare gegevens, moeten als persoonsgegevens worden aangemerkt. Dat laatste betekent dat de huidige status van gegevens niet bepalend is; om de juridische categorisering te bepalen (zijn gegevens juridisch gezien persoonsgegevens of niet?), moet rekening worden gehouden met de waarschijnlijke toekomstige status ervan. Zoals de Artikel 29 Werkgroep heeft benadrukt moet de verwerkingsverantwoordelijke rekening houden met de mogelijkheid van identificatie die zich ook over 9 jaar kan voordoen. Dit heeft grote gevolgen, met name voor open data, die permanent online blijft en door verschillende partijen zal worden gebruikt.
2. Om vast te stellen of gegevens persoonsgegevens zijn, dient rekening te worden gehouden met alle middelen die redelijkerwijs voor identificatie kunnen worden gebruikt. Om de waarschijnlijkheid van identificatie vast te stellen, moet worden gekeken naar de kosten en de hoeveelheid tijd die nodig is voor identificatie, de beschikbare technologie op het moment van de verwerking, en toekomstige technologische ontwikkelingen. Hoewel dit op zichzelf objectief verifieerbare criteria zijn, hangt de interpretatie ervan, zoals zowel de Artikel 29 Werkgroep als het HvJ EU keer op keer hebben benadrukt, af van de context.
3. De vraag is niet alleen of de verwerkingsverantwoordelijke zelf nu of in de toekomst persoonsgegevens kan verkrijgen uit een dataset, maar ook of een partij die toegang heeft tot de gegevens dat kan. Dit is wederom met name van belang wanneer gegevens online beschikbaar worden gesteld of met meerdere partijen worden gedeeld. Hoe meer partijen toegang hebben tot een database, hoe groter de kans dat iemand persoonsgegevens uit de dataset afleidt, terwijl tegelijkertijd de middelen om te verifiëren of iemand dit heeft gedaan, wanneer en waarom, afnemen.
4. Identificatie is niet vereist; het kunnen uitlichten van een persoon is voldoende. Als een internetbedrijf niet weet wie een persoon is, maar wel gepersonaliseerde advertenties kan tonen aan account 87&^%11!, dan is dat in principe voldoende om de gegevens juridisch te kwalificeren als persoonsgegevens. Evenzo is dit het geval wanneer een verzekeraar aanvragen afwijst van een persoon (zonder zijn naam te kennen) uit een gebied met een specifieke postcode. Op een vergelijkbare manier heeft de Artikel 29 Werkgroep benadrukt dat gegevens kunnen worden beschouwd als "betrekking hebbend op" een persoon, omdat het gebruik ervan waarschijnlijk gevolgen zal hebben voor de rechten en belangen van een bepaalde persoon,

rekening houdend met alle omstandigheden van het concrete geval. De Artikel 29 Werkgroep benadrukte dat het niet nodig is dat het potentiële resultaat een grote impact heeft.

Geaggregeerde gegevens

Door aggregatie kunnen gegevens geanonimiseerd worden door de gegevens niet langer te behandelen op het niveau van n = 1, maar op het niveau van n = 20, n = 100, enz. De analyse van geaggregeerde gegevens kan leiden tot informatie zoals - in zeer basale termen - van de 100.000 mensen met een groene auto heeft 34% een witte bank in de woonkamer. Deze gegevens worden in principe niet als persoonsgegevens beschouwd. Wanneer partijen echter op basis van geaggregeerde gegevens handelen op een manier die directe gevolgen heeft voor natuurlijke personen, dan kan dat wel, bijvoorbeeld wanneer een autobedrijf advertenties voor witte banken stuurt naar alle mensen die een groene auto hebben gekocht. Door de hele AVG heen zijn er verwijzingen naar statistische gegevens en geaggregeerde gegevens bedoeld voor onderzoeksdoeleinden. De AVG onderschrijft de publieke belangen die met statistische analyse gediend kunnen worden (statistische analyse door het Centraal Bureau voor de Statistiek is bijvoorbeeld essentieel voor op informatie gebaseerde beleidsvorming door de overheid). Als gegevens zodanig worden geaggregeerd dat er geen individuele gegevens kunnen worden geëxtraheerd noch worden gebruikt op een manier die directe gevolgen heeft voor concrete personen, is de AVG niet van toepassing. In dat geval kunnen de regels voor het verwerken van statistische gegevens gelden, die normen inhouden voor onder meer vertrouwelijkheid en veiligheid.

Wanneer persoonsgegevens worden gebruikt voor statistische verwerkingen, is de AVG van toepassing, maar laat deze ruimte voor uitzonderingen op nationaal niveau. Artikel 85 AVG maakt uitzonderingen mogelijk voor de verwerking van persoonsgegevens in het kader van de vrijheid van meningsuiting; artikel 86 AVG bepaalt dat persoonsgegevens in officiële documenten die in het bezit zijn van een (semi) publieke instelling, door die instelling mogen worden bekendgemaakt in overeenstemming met wet- en regelgeving om de toegang van het publiek tot officiële documenten in overeenstemming te brengen met het recht op bescherming van persoonsgegevens; en artikel 89 AVG bepaalt dat de lidstaten vrijstellingen kunnen aannemen, met name ten aanzien van de rechten van betrokkenen, wanneer persoonsgegevens worden verwerkt voor archiveringsdoeleinden in het algemeen belang, wetenschappelijke of historische onderzoeksdoeleinden of statistische doeleinden. Er zijn geen factoren uiteengezet om te bepalen wanneer een database zodanig wordt geaggregeerd dat deze kwalificeert als niet-persoonsgegeven. Dit is afhankelijk van de omstandigheden van het geval, rekening houdend met de eerder besproken algemene elementen.

Pseudonieme gegevens

De AVG is van toepassing op gepseudonimiseerde gegevens, maar er zijn enkele uitzonderingen van toepassing op de verplichtingen van verwerkingsverantwoordelijken wanneer zij gegevens hebben gepseudonimiseerd. Bovendien wordt pseudonimisering beschouwd als een manier om technische en organisatorische veiligheidsstandaarden te implementeren, specifieke verplichtingen die zijn vastgelegd in het kader voor gegevensbescherming. De AVG (Artikel 4 lid 5) definieert 'pseudonimisering' als het verwerken van persoonsgegevens op een zodanige manier dat de persoonsgegevens niet meer aan een specifieke betrokkene kunnen worden toegeschreven zonder het gebruik van aanvullende informatie, op voorwaarde dat dergelijke aanvullende informatie afzonderlijk wordt bewaard en is onderworpen aan technische en organisatorische maatregelen om ervoor te zorgen dat de persoonsgegevens niet worden toegeschreven aan een geïdentificeerde of identificeerbare natuurlijke persoon. Overweging 28 maakt duidelijk dat pseudonimisering van persoonsgegevens de risico's voor de betrokkenen kan verkleinen. Daarom wordt, zoals uiteengezet in overweging 29, pseudonimisering gestimuleerd door de AVG.

Het begrip pseudonimisering is nieuw in de AVG; het speelde geen rol in eerdere

gegevensbeschermingsregelingen. Hoewel de Verordening benadrukt dat andere technieken voor een veilige verwerking van persoonsgegevens niet worden uitgesloten door het feit dat pseudonieme gegevens apart worden gedefinieerd, wordt aan deze techniek wel een bijzondere status toegekend. Wat de juiste interpretatie van dit juridische begrip complex maakt is dat de AVG vaak pseudonimisering in één adem noemt met encryptie, een term die niet apart wordt gedefinieerd. De Artikel 29 Werkgroep steunt het toegenomen gebruik van pseudonimiseringstechnieken, waarvan zij vijf belangrijke technieken onderscheidt, waaronder encryptie, en beschouwt (bepaalde vormen van) encryptie als een subset van pseudonimiseringstechnieken.

Gevoelige persoonsgegevens

Gevoelige persoonsgegevens worden onder het gegevensbeschermingsregime apart gedefinieerd ten opzichte van 'gewone' persoonsgegevens. Bijzondere of gevoelige persoonsgegevens zijn duidelijk omschreven en afgebakend; de verwerking geldt per definitie als potentieel schadelijk voor de belangen van natuurlijke personen. Het verwerken van gevoelige gegevens is in principe verboden, al geldt er een groot aantal uitzonderingen op dat verbod. Gevoelige gegevens worden gedefinieerd als persoonsgegevens waaruit ras of etnische afkomst, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen of lidmaatschap van een vakbond blijken, en de verwerking van genetische gegevens, biometrische gegevens met het oog op de unieke identificatie van een natuurlijke persoon, gegevens over gezondheid of gegevens over het seksleven of de seksuele geaardheid van een natuurlijk persoon. De verwerking van strafrechtelijke gegevens door rechtshandhavingsinstanties valt onder de zogenoemde Politierichtlijn.

Het HvJ EU heeft een ruime interpretatie gegeven aan wat als gevoelige persoonsgegevens moet worden beschouwd. In de *Lindqvist*-zaak had iemand bijvoorbeeld op een blog geschreven dat een collega vanwege medische redenen deeltijd werkte omdat ze een voetblessure had opgelopen. De vraag of informatie over het hebben van een voetblessure reeds kwalificeert als 'medische gegevens' werd door de rechter slechts kort, staccato en bevestigend beantwoord. Een andere zaak, die van *V.*, betrof de overdracht van een medisch dossier in het kader van de arbeidsverhouding. De rechter wees erop dat medische gegevens bijzonder gevoelige gegevens zijn, waardoor schijnbaar een hiërarchie ontstaat tussen verschillende categorieën gevoelige persoonsgegevens en medische gegevens bovenaan komen te staan.

2. Juridisch regime: de categorale en de contextuele benadering

Er is een spanning tussen twee reguleringsbenaderingen op het gebied van gegevensbescherming: een contextuele en een categorale benadering, een benadering die rekening houdt met de omstandigheden van het geval en een benadering die is gebaseerd op vaste definities en duidelijke regels die aan de definities zijn gekoppeld. Elk van deze benaderingen heeft duidelijke voor- en nadelen. De eerste benadering kan per scenario met alle relevante aspecten rekening houden, past zich beter aan, aan veranderende omstandigheden en loopt dus niet het risico achterhaald te zijn of omzeild te worden. Fluïde en contextuele reguleringsbenaderingen hebben echter het nadeel dat ze vaag zijn en weinig rechtszekerheid bieden, zowel voor de verwerkingsverantwoordelijke als voor de betrokkene. De tweede benadering lost dit probleem op: het geeft een duidelijke reeks definities en categorieën en koppelt daaraan een duidelijke reeks regels. Maar het nadeel is ook duidelijk, namelijk het risico omzeild te worden en verouderd te raken; ook is deze benadering minder granulair dan een contextuele benadering.

Er is een diepe ambivalentie in de reguleringsbenadering op dit punt.

Op het eerste gezicht is de categorale benadering het duidelijkst. Zo had de ontkoppeling van het recht op gegevensbescherming van het recht op privacy te maken met een de-contextualisering van het recht.

In het mensenrechtenkader wordt een claim beoordeeld op zowel de ratione materiae (valt de klacht onder de materiële reikwijdte van het ingeroepen artikel?) als de ratione personae (kan de verzoeker bewijzen in aanmerkelijke mate getroffen te zijn?). Wat dat tweede beginsel betreft, geldt een aanzienlijke drempel, aangezien verzoekers moeten kunnen aantonen dat zij directe, individualiseerbare en substantiële schade hebben geleden. In het kader voor gegevensbescherming worden beide principes samengevoegd. Dit betekent dat elke verwerking van persoonsgegevens, hoe alledaags en betekenisloos ook, wordt beschouwd als verwerking van persoonsgegevens, waarop de AVG van toepassing is. Bijgevolg wordt het contextuele of op schade gebaseerde element dat essentieel is voor evaluaties van mensenrechtenvraagstukken weggelaten uit de gegevensbeschermingsregeling. De toepassing van het gegevensbeschermingsregime, anders dan bijvoorbeeld het recht op privacy, is niet afhankelijk van de vraag of er schade is toegebracht aan een eiser of rechthebbende.

Daarnaast is het duidelijk dat het gegevensbeschermingskader werkt met een binair onderscheid tussen persoonsgegevens en niet-persoonsgegevens. De EU heeft persoonsgegevens voorzien van de hoogste vorm van rechtsbescherming ter wereld, via de AVG en de Politierichtlijn, terwijl de EU met betrekking tot de verwerking van niet-persoonsgegevens expliciet beperkingen die zijn opgelegd door de private en publieke sector organisaties ontmoedigt door middel van de Verordening over het vrije verkeer van niet-persoonsgegevens. Omdat het onderscheid tussen persoonsgegevens en niet-persoonsgegevens binair is (alhoewel de voorgestelde Data Governance Act dit beeld misschien zal compliceren), zal de vraag of een dataset als een van beide wordt gecategoriseerd, een reguleringsverschil van 180 graden betekenen. Ook ten aanzien van zowel pseudonieme gegevens als gevoelige persoonsgegevens is sprake van een binaire benadering: gegevens zijn pseudoniem of niet, persoonsgegevens zijn gevoelig of niet. Met betrekking tot het laatste type gegevens is de categorale benadering nog duidelijker. De AVG bevat een beperkte en uitputtende lijst van soorten gegevens die als gevoelig worden beschouwd. De verwerking van dergelijke gegevens is in principe verboden.

Een laatste punt dat moet worden benadrukt, is dat het kader voor gegevensbescherming als geheel gebaseerd is op binaire onderscheiden en wordt gekenmerkt door een categorale benadering. Zo worden duidelijke verschillen tussen verschillende actoren, zoals de verwerkingsverantwoordelijke, de gegevensverwerker en de betrokkene, vastgelegd. Elk van deze actoren heeft een duidelijk omschreven rol, een reeks verplichtingen, rechten en regelgevende verantwoordelijkheden. Een partij kan niet tegelijkertijd gegevensverwerker en verwerkingsverantwoordelijke zijn met betrekking tot dezelfde gegevensverwerking: het is een kwestie van of/of. Evenzo is bij een gegevensverwerking een partij ofwel een (mede)verantwoordelijke ofwel een betrokkene.

Anderzijds is een contextuele benadering zichtbaar. Hoewel het onderscheid tussen persoonsgegevens en niet-persoonsgegevens bijvoorbeeld binair is, omvat de definitie van persoonsgegevens een contextueel aspect. Het begrip ‘identificeerbaar’ houdt in dat gegevens die op dit moment geen identificatie van een persoon mogelijk maken, maar dit in de toekomst wel mogelijk zullen maken, nu al als persoonsgegevens worden aangemerkt.

Bovendien, hoewel de categorie van pseudonieme gegevens op zichzelf binair is - gegevens zijn pseudoniem of niet - wordt deze categorie door velen gezien als een tussencategorie tussen persoons- en niet-persoonsgegevens. Pseudonieme gegevens zijn niet anoniem en daarom is de AVG van toepassing, maar ze zijn niet zo eenvoudig te koppelen aan een geïdentificeerde persoon. Daarom staat de AVG een aantal uitzonderingen toe wanneer gegevens worden gepseudonimiseerd. Evenzo, hoewel het onderscheid tussen niet-gevoelige en gevoelige gegevens vaak als absoluut wordt gepresenteerd, zijn alle verschillende rechten en plichten van toepassing op zowel de verwerking van gevoelige als niet-gevoelige persoonsgegevens. Het enige verschil is de legitieme grond voor het verwerken van de gegevens (artikel 6 en artikel 9) en hoewel artikel 9 als uitgangspunt neemt dat het verwerken van gevoelige gegevens verboden is, somt het een groot aantal uitzonderingen op dit verbod op, waardoor het verschil tussen de verwerking van gevoelige en niet-gevoelige gegevens minder binair is dan op het

eerste gezicht lijkt.

Wanneer het gegevensbeschermingsregime van toepassing is, zijn de meeste verplichtingen en vereisten contextafhankelijk, wat in het algemeen betekent dat hoe meer gegevens worden verzameld, hoe gevoeliger die gegevens zijn, hoe hoger het risico van gegevensverwerking of hoe meer partijen erbij betrokken zijn, hoe strenger de regels en verplichtingen moeten worden geïnterpreteerd. Deze contextuele benadering is van toepassing op de verplichting om een gegevensbeschermingsbeleid te implementeren, technische en organisatorische beveiligingsmaatregelen te nemen en om aan 'data protection by design' of 'by default' te doen. Ook kennen de verplichtingen en vereisten uit de AVG bepaalde contextafhankelijke beperkingen en uitzonderingen. Zo geldt het documentatievereiste niet voor kleine organisaties die zich niet bezighouden met risicovolle verwerkingen; een gegevensbeschermingseffectbeoordeling hoeft alleen te worden uitgevoerd wanneer potentiële schade waarschijnlijk is; en organisaties uit de particuliere sector hoeven alleen een functionaris voor gegevensbescherming aan te stellen wanneer hun kernactiviteiten bestaan uit het regelmatig en systematisch monitoren van betrokkenen op grote schaal of ze grootschalige verwerking van gevoelige gegevens verrichten en de melding van datalekken is afhankelijk van de schade die waarschijnlijk uit de inbreuk voortvloeit. De kernregels uit het gegevensbeschermingskader zijn dan ook zeer contextueel.

Ten slotte moet worden benadrukt dat hoewel de Europese benadering van privacy en gegevensbescherming vaak in contrast wordt gebracht met de Amerikaanse (de eerste een omnibusbenadering, de tweede een sectorale benadering), het contrast minder scherp is dan vaak wordt gedacht. De EU maakt expliciet onderscheid tussen twee contexten wanneer zij gegevensbeschermingsregels toepast: de algemene context, die onder de AVG valt, en de wetshandhavingscontext waarop de Politierichtlijn van toepassing is. Daarnaast bevordert de AVG het gebruik van gedragscodes, waarmee sectoren hun eigen invulling en specificatie aan het gegevensbeschermingsregime kunnen geven. Dat deze mogelijkheid nauwelijks wordt gebruikt omdat sectoren vrezen voor de administratieve lasten van het uitvoeren van toezicht en het afhandelen van klachten, betekent niet dat dit niet wordt gestimuleerd door de AVG.

3. De impact van de beschikbaarheid van data en datatechnologieën op de wettelijke regulering van data

De beschikbaarheid van data groeit exponentieel. Sinds de jaren zeventig, toen de eerste grotere databases ontstonden, en nu, vijftig jaar later, is het datalandschap ingrijpend veranderd. Niet alleen worden er meer data verzameld en beschikbaar gesteld, maar fundamenteler, de samenleving is veranderd van een analoge naar een gedataficeerde samenleving, waarin vrijwel alle aspecten van het leven worden gevolgd met sensoren, cookies, camera's en satellieten. Niet alleen overheden gebruiken monitoringstechnieken om de verschillende aspecten van het leven te monitoren, ook grote internetbedrijven en steeds meer data-gedreven bedrijven doen dit. Ook burgers hebben toegang tot allerhande spyware, drones en andere sensorische producten om gegevens over zichzelf en anderen te verzamelen. Deze gegevens worden gedeeld via intermediaire platforms, opgeslagen in de 'cloud' en beschikbaar gesteld op besloten of open platforms. Een andere trend is dat, met Web 2.0, door gebruikers gegenereerde inhoud (sociale netwerken) is geëxplodeerd, en daarom zijn gebruikers zelf een belangrijke bron van persoonsgegevens (van zichzelf en hun vrienden) geworden.

Er is ook een juridisch streven om gegevens vrij te geven. In de Europese Unie zijn er verschillende wetten die partijen verplichten zich open te stellen. Zo stelt de Open Access-richtlijn voor dat de lidstaten zoveel mogelijk overheidsinformatie gratis, in open access en herbruikbaar formaat openbaar maken. De richtlijn vennootschapsrecht verplicht de lidstaten om de nodige maatregelen te nemen om te zorgen voor verplichte openbaarmaking door vennootschappen van onder meer de oprichtingsakte, de statuten, de benoeming en de beëindiging van hun ambt. De verordening betreffende het vrije verkeer van niet-persoonsgegevens, om een laatste voorbeeld te geven, ontmoedigt zowel organisaties in de publieke sector als de particuliere sector om niet-persoonsgegevens te privatiseren.

Op het gebied van open data hebben de afgelopen jaren drie belangrijke ontwikkelingen plaatsgevonden:

1. Digitalisering: overheidsdocumenten lagen vroeger in archieven, bibliotheken of speciaal daarvoor bestemde documentatiecentra. Tegenwoordig worden steeds meer documenten online beschikbaar gesteld. Dit heeft een belangrijk effect op de zogenaamde 'practical obscurity'. Het feit dat men zich in het verleden de moeite moest getroosten om naar de plaats te gaan waar de documenten waren opgeslagen, ze op te vragen en in te zien, betekende dat in de praktijk slechts een beperkt aantal mensen de informatie zou raadplegen. In grote lijnen waren dat journalisten, historici, kritische burgers die de overheid op de voet volgden en amateurhistorici die hun stamboom onderzochten. Door de documenten openbaar te maken op het internet en geen toegangsbarrières op te werpen kan iedereen deze documenten gemakkelijk bekijken.
2. Actieve openbaarmaking: in het pre-digitale tijdperk werden de meeste documenten 'passief openbaar gemaakt'; burgers, journalisten en anderen kregen op verzoek toegang tot bepaalde documenten. Ze moesten dus al een globaal idee hebben van wat ze zochten, de openbaarmaking van documenten vereiste hun initiatief en de documenten werden meestal slechts voor een bepaalde periode beschikbaar gesteld. Momenteel worden documenten steeds vaker actief openbaar gemaakt; de overheid publiceert documenten niet op verzoek, maar op eigen initiatief. Dit betekent dat er geen specifieke reden meer is waarom een document beschikbaar wordt gesteld. Iedereen heeft er toegang toe en op elk moment.
3. Technologieën: de technische mogelijkheden om dergelijke documenten te doorzoeken zijn aanzienlijk toegenomen. Deze omvatten algoritmen en kunstmatige intelligentie die teksten kunnen analyseren op woorden, correlaties en onderwerpen. Waar het voorheen vooral individuen waren die toegang zochten tot overheidsdocumenten, zijn het momenteel technologiebedrijven die de beste uitgangspositie hebben om de miljoenen overheidsdocumenten die online verschijnen te scannen en te analyseren.

Daarnaast heeft er, gezien de algemene beschikbaarheid van data en datatechnologieën, het gemak van dataverzameling en -verwerking en de lagere kosten, een belangrijke verschuiving plaatsgevonden in het type dataverwerking. Gezien de kosten en praktische en technologische beperkingen voor het verzamelen van gegevens, waren veel gegevensoperaties, zelfs tot 20 jaar geleden, heel doelgericht. Er was een specifiek en vooraf vastgesteld doel waarvoor specifieke gegevens over specifieke entiteiten werden verzameld. Momenteel hebben echter veel, zo niet de meeste, gegevensverwerkingsoperaties betrekking op structurele en systemische gegevensverzamelingen, zoals camera's en sensoren die iedereen, overal in het publieke domein permanent bewaken en alomtegenwoordige online tracking. Deze verschuiving betekent dat de verzamelde gegevens vaak niet betrekking hebben op vooraf geïdentificeerde individuen, maar op groepen, categorieën of de gehele bevolking. Dit heeft op zijn beurt een verschuiving in gang gezet van de analyse van individuele gegevens naar die van statistische en geaggregeerde gegevens, van directe naar afgeleide gegevens en van zekere naar probabilistische informatie.

Deze ontwikkelingen hebben effect op de manier waarop de huidige wetgeving inzake gegevensbescherming is ingericht en met name op de categorale aanpak.

1. Werken met afgebakende definities van verschillende soorten gegevens gaat alleen als een 'datum' op een relatief stabiele manier in één categorie valt. Dit is steeds minder het geval. De aard van de data in ‘Big Data’-processen is niet stabiel, maar veranderend. Een dataset met gewone persoonsgegevens kan worden gekoppeld aan, en verrijkt met, een andere dataset om gevoelige gegevens af te leiden; de gegevens kunnen vervolgens worden geaggregeerd of ontdaan van identificatiegegevens; vervolgens kunnen de gegevens worden gedeanonimiseerd of geïntegreerd in een andere dataset om persoonsgegevens te creëren. Dit alles kan in een fractie van een seconde gebeuren. De vraag is dus of het zinvol is om met goed gedefinieerde categorieën te werken als dezelfde 'datum' of dataset letterlijk van seconde tot seconde in een

andere categorie kan vallen.

2. Ook wordt het steeds moeilijker om de status van gegevens precies te bepalen. De beoordeling of de gegevens de identificatie van een persoon mogelijk maken en of de informatie al dan niet als anoniem kan worden beschouwd, hangt af van de omstandigheden van het geval. Om de huidige status van een datum of dataset te bepalen, moet daarom rekening worden gehouden met de verwachte toekomstige status van de gegevens. Gezien de algemene beschikbaarheid van technologieën en de minimale investering die nodig is, wordt het steeds waarschijnlijker dat wanneer een database wordt gedeeld of anderszins beschikbaar wordt gesteld, er een partij is die deze gaat verrijken met andere gegevens. Zo wordt het steeds waarschijnlijker dat als een geanonimiseerde dataset openbaar wordt gemaakt, er een partij is die deze de-anonimiseert of combineert met andere data om persoonlijke profielen te maken; dat als een set persoonsgegevens wordt gedeeld, er een partij is die de gegevens zodanig gaat gebruiken om te komen tot een dataset met gevoelige persoonsgegevens; enzovoort. Aan de andere kant zullen er andere partijen zijn die toegang hebben tot die gegevens, maar zich niet bezighouden met dergelijke activiteiten; partijen die de gegevens niet zullen gebruiken, gebruiken zoals deze worden verstrekt of zelfs een database met persoonsgegevens de-identificeren. Wie wat doet is vooraf niet duidelijk. De juridische categorie waartoe de gegevens behoren is dus niet langer een kwaliteit van de gegevens zelf, maar een product van de inspanningen en investeringen van een verwerkingsverantwoordelijke.
3. De vraag is of het onderscheid tussen verschillende categorieën gegevens nog relevant is. De achterliggende gedachte is dat de verwerking van persoonsgegevens gevolgen heeft voor natuurlijke personen, terwijl de verwerking van niet-persoonsgegevens dat niet heeft en dat de verwerking van gevoelige persoonsgegevens zeer grote gevolgen kan hebben (groter dan de verwerking van ‘gewone’ persoonsgegevens gegevens normaal gesproken heeft), zodat dit onder het strengste regime valt, persoonsgegevens onder het 'normale' beschermingsregime vallen en de verwerking van niet-persoonsgegevens aan geen enkele beperking onderworpen is. De vraag is in hoeverre deze aanname nog houdbaar is in de 21e eeuw. Moderne gegevensverwerking op basis van geaggregeerde gegevens kan grote individuele en maatschappelijke gevolgen hebben. Profilering van groepen in plaats van individuen betekent voorts dat de gevolgen aanzienlijk kunnen zijn, maar niet altijd direct te relateren zijn aan individuen.

4. De impact van huidige en toekomstige datatechnologieën op de juridische categorieën

Dit onderzoek richtte zich op de technologische ontwikkelingen met betrekking tot vier toepassingsgebieden, namelijk anonimisering, aggregatie, pseudonimisering en het afleiden van gevoelige gegevens uit niet-gevoelige (persoons)gegevens. De bevindingen met betrekking tot elk van deze toepassingsgebieden zullen hieronder worden samengevat.

Anonimisering

De meest relevante anonimiseringstechnieken in het kader van dit onderzoek zijn:

1. Maskeren: beoogt een relatie te genereren tussen de oorspronkelijke set X en de gegenereerde set Y, zodat de indirecte identifiers worden gemaskeerd.
   1.1 Niet-perturbatieve maskering: gedeeltelijke onderdrukking of reductie van detail of verruwing van de originele dataset X. Hierdoor is dataset Y niet per se een verstoorde dataset, maar eerder een gereduceerde versie van de dataset X. Niet-perturbatieve maskering omvat onder andere:
      1.1.1 Sampling: vrijgeven van een sample S van de originele dataset X. Sampling is geschikt voor kwalitatieve identifiers waarop geen rekenkundige bewerkingen kunnen worden uitgevoerd, zoals de oogkleur van een persoon of de maanden van het jaar;

1.1.2 Generalisatie: reductie van data granulariteit zodat dataset Y minder nauwkeurig is dan dataset X. Deze techniek is geschikt voor kwalitatieve identifiers, omdat het de maskering van bestanden met ongebruikelijke combinaties ondersteunt;
1.1.3 'Top- en bottom'-codering: een speciaal geval van generalisatie waarbij top-codes of bottom-codes worden ingesteld vanuit de originele identifiers van dataset X;
1.1.4 Onderdrukking: verwijdering van de gehele of van bepaalde identifiers in dataset Y vóór de vrijgave ervan. Aangezien het herstellen van informatie niet mogelijk is, wordt onderdrukking beschouwd als de sterkste anonimiseringstechniek.

1.2 Perturbatieve maskering: de vervorming of verstoring van microdata zodat de statistische eigenschappen van de oorspronkelijke dataset X behouden blijven in dataset Y. Perturbatieve maskering omvat onder meer:
1.2.1 Ruistoevoeging: maskering van identifiers door willekeurige ruis toe te voegen;
1.2.2 Data swapping: het uitwisselen van identifiers tussen individuele records;
1.2.3 Microaggregatie: clustering van records van dataset X in kleine aggregaten of groepen van k elementen, waarbij het gemiddelde van de waarden van de groep waartoe het record behoort, wordt gepubliceerd in dataset Y.

2. Synthetische data: heeft tot doel een dataset Y te creëren die bestaat uit willekeurig gesimuleerde bestanden die niet direct uit de dataset X zijn afgeleid, met behoud van de statistische eigenschappen van de oorspronkelijke dataset X. Als zodanig kunnen standaarddeviaties, medianen, lineaire regressie of andere statistische technieken worden gebruikt om synthetische gegevens te genereren.

Manieren om anonimiteit vanuit een technisch perspectief te definiëren omvatten, maar zijn niet beperkt tot:

1. k-anonimiteit: probeert de her-identificatie van bestanden te voorkomen op basis van een vooraf gedefinieerde set van indirecte identifiers. Een cel in een database verwijst in ieder geval naar k individuen;
2. l-diversiteit: heeft tot doel ervoor te zorgen dat elke groep gevoelige identifiers verschillende waarden bevat en dat geen van deze waarden domineert in frequentie;
3. t-closeness: stelt het gebruik van een relatief instrument voor om de variabiliteit van de waarden van de gevoelige identifiers te meten, waardoor de informatiewinst over de betrokkenen wordt beperkt. Alle waarden die door het sensitieve attribuut worden aangenomen, worden als even gevoelig beschouwd;
4. ε-differentiële privacy: de gegevensbeheerder genereert geanonimiseerde weergaven van een dataset met behoud van een kopie van de originele gegevens. Die weergaven of subsets zijn dus anoniem, maar de gegevensbeheerder heeft vaak nog steeds identificerende informatie.

Hoewel elk van deze technieken waardevol is, kan geen van deze technieken absolute anonimiteit garanderen. Met voldoende tijd, middelen en adequate technologie kunnen vrijwel alle geanonimiseerde gegevens worden ge-deanonimiseerd. Al in 2009 concludeerde Paul Ohm dat data ofwel waardevol ofwel perfect anoniem kunnen zijn, maar nooit beide. Technische literatuur onderstreept dat dit punt nu meer dan ooit waar is. Technische experts verwachten, zoals blijkt uit de interviews, geen revolutionaire nieuwe ontwikkelingen op het gebied van anonimisering of de-anonimisering, maar menen over het algemeen dat volledige anonimisering, zeker in juridische zin, steeds moeilijker zal worden gezien de algemene beschikbaarheid van technologieën en de algemene beschikbaarheid van gegevens.

Aggregatie

Door aggregatie worden de gegevens in een dataset niet op individueel niveau (n = 1) gepresenteerd, maar op geaggregeerd niveau (n = 10; n = 100; n = 1000). Hoe hoger het aggregatieniveau, hoe waarschijnlijker het is dat de dataset juridisch gezien geen persoonsgegevens bevat, hoewel een

dergelijke beoordeling altijd afhankelijk is van de omstandigheden van het geval. De meest relevante aggregatietechnieken in het kader van dit onderzoek zijn:

1. Aggregatie op basis van derden: vertrouwde derden kunnen onbewerkte gegevens verzamelen, deze gegevens aggregeren en de resulterende gegevens overdragen aan geautoriseerde ontvangers. Op deze manier hebben de ontvangers alleen de geaggregeerde gegevens. Dit is echter mogelijk niet het geval voor een vertrouwde derde partij.
2. Aggregatie op basis van gegevensverstoring: willekeurige ruis wordt toegevoegd aan de verzamelde gegevens zodat de oorspronkelijke gegevens niet traceerbaar zijn, maar geaggregeerde waarden kunnen nog steeds worden berekend met een kleine of verwaarloosbare fout. Het nadeel van gegevensverstoring is het verschil tussen de oorspronkelijke gegevens en de verstoorde gegevens, wat in bepaalde gevallen kan leiden tot ongelijkheden in de berekening.
3. Aggregatie op basis van cryptografie: cryptografische primitieven kunnen worden gebruikt om de nadelen van de vorige methoden weg te nemen. Volledig homomorfe encryptie is een encryptietechnologie waarmee analyses in de cijfertekst op dezelfde manier als in de leesbare tekst kunnen worden uitgevoerd zonder de geheime sleutel te delen. Dit houdt in dat de berekening wordt uitgevoerd over de versleutelde gegevens zonder de noodzaak om deze te ontsleutelen, waardoor het delen van gegevens met derden mogelijk wordt. De resultaten van de berekening zijn gelijkelijk versleuteld, zodat alleen de data-exporteurs de gegevens kunnen ontsleutelen.
4. Statistical Disclosure Control: misschien wel de belangrijkste techniek voor het verzamelen van gegevens, vooral in het licht van het openbaar maken van de gegevens, is Statistical Disclosure Control (SDC). SDC heeft als doel om, zowel direct als indirect, identificerende informatie in een dataset te elimineren, met zoveel mogelijk behoud van de datakwaliteit. De specialist die verantwoordelijk is voor het beschermen van de gegevens moet verschillende methoden van openbaarmakingscontrole gebruiken, zodanig dat het minimaal vereiste beschermingsniveau wordt bereikt en dat het informatieverlies zo klein mogelijk is, wat per situatie zal verschillen. Wat informatieverlies is, kan niet als zodanig worden bepaald, omdat informatie een subjectief begrip is dat door elke gebruiker anders kan worden gedefinieerd.

Hoewel de technische mogelijkheden voor het anonimiseren van gegevens in geaggregeerde datasets groot zijn, en in het algemeen groter dan wanneer gegevens niet worden geaggregeerd, doet zich een nieuw probleem voor, dat in de technische literatuur wordt aangeduid als het samenstellingsprobleem. Dit betekent dat uit de combinatie van twee of meer datasets die zelf geen persoonsgegevens bevatten, persoonsgegevens kunnen worden afgeleid. Het kan gaan om gegevens over geïdentificeerde personen die vroeger in die databases zaten, maar het kan ook om andere personen gaan. Bovendien moet worden benadrukt dat als een partij algemene informatie zou gebruiken om beslissingen te nemen die van invloed zijn op personen, dit op juridisch gebied als persoonsgegevens zou worden gekwalificeerd. Uiteraard is het vooraf moeilijk in te schatten welke partij welke geaggregeerde data zal gebruiken voor welk type besluitvorming.

Hoewel anonimisering van geaggregeerde gegevens in isolatie potentieel mogelijk is, als bijvoorbeeld alleen de dataset als een relevante bron voor identificatiedoeleinden wordt beschouwd, zijn zowel de literatuur als de voor dit onderzoek geïnterviewde experts het erover eens dat dit steeds minder bepalend zal zijn. Dat heeft niet zozeer te maken met ontwikkelende technieken, maar met het groeiende datalandschap en de beschikbaarheid van open data. Omdat het waarschijnlijk is dat bijna elke geaggregeerde dataset op termijn zal worden gebruikt voor gevolgtrekkingen op persoonlijk niveau, voor samenstellingsactiviteiten en/of voor het ontwikkelen van besluitvormingsbeleid dat gevolgen heeft voor mensen, kan, vanuit juridisch perspectief, geen enkele geaggregeerde dataset worden aangemerkt als absoluut buiten het gegevensbeschermingsregime vallend.

Pseudonimisering

De meest relevante pseudonimiseringstechnieken in het kader van dit onderzoek zijn:

1. Hashing is een techniek waarmee pseudoniemen kunnen worden afgeleid. In een notendop zijn hashfuncties functies die een invoer van willekeurige lengte comprimeren tot een resultaat met een vaste lengte. Deze uitvoer met een vaste grootte wordt een berichtsamenvatting, hash-waarde, hash-code of gewoon hash genoemd. Als een identifier *m* wordt gebruikt als invoer in de hash-functie *h*, zal de functie een pseudoniem met vaste grootte *h(m)* teruggeven.
2. Hashing met een sleutel of keyed hashing bouwt voort op conventionele hashing door een geheime sleutel toe te voegen die de uitvoer van de functie *h* verandert. Hashing met sleutel kan verschillende pseudoniemen produceren voor dezelfde invoer, afhankelijk van de keuze van de specifieke sleutel.
3. 'Salted' hashing is een variant van keyed hashing, waarbij gebruik wordt gemaakt van een conventionele hashfunctie in combinatie met een zogenaamde 'salt', of aanvullende willekeurig uitziende data. Net als keyed hashing, produceert hashing met salt verschillende pseudoniemen voor dezelfde initiële identifier. Daarom heeft salted hashing dezelfde eigenschappen als keyed hashing, zolang het 'salt' op de juiste manier is beveiligd en derden er geen kennis van hebben.
4. Peppered hashing bestaat uit het toevoegen van een geheim aan het 'salt' tijdens het hashen en het apart opslaan van salts en pseudoniemen in een ander medium, bijvoorbeeld in een hardware beveiligingsmodule. De 'pepper' deelt daarom bepaalde eigenschappen met salt omdat het een willekeurige waarde is en vergelijkbaar met een coderingssleutel omdat het geheim moet worden gehouden.
5. Tokenisatie bestaat uit het vervangen van identifiers door willekeurig gegenereerde waarden, ook wel tokens genoemd, zonder enige wiskundige relatie en zonder het type of de lengte van de gegevens te veranderen. Dit is een belangrijk verschil met encryptie. In tegenstelling tot de laatstgenoemde, voorkomt de onveranderlijkheid van gegevenstypen en lengtes bij tokenisatie elke onbegrijpelijkheid van informatie door verwerking in tussenliggende systemen. Tegelijkertijd betekent dit ook een afname van de rekenkracht die nodig is om de tokens te verwerken. Aangezien er geen sleutels of algoritmen zijn gebruikt om de oorspronkelijke identifier uit het token af te leiden, impliceert de kennis van een token niet de openbaarmaking van persoonsgegevens.

De meest relevante encryptietechnieken in het kader van dit onderzoek zijn:

1. Symmetrische encryptie bestaat uit het gebruik van één geheime sleutel om elektronische informatie zowel te versleutelen als te ontsleutelen. Partijen die vertrouwen op symmetrische versleuteling moeten de geheime sleutel delen om het ontsleutelingsproces mogelijk te maken. Symmetrische encryptie transformeert de initiële identifier (maar ook de volledige dataset) in een pseudoniem (of cijfertekst), die vervolgens wordt gedecodeerd om de initiële identifier te onthullen.
2. Asymmetrische encryptie bestaat uit het gebruik van twee sleutels, een openbare en een privésleutel, om elektronische informatie zowel te versleutelen als te ontsleutelen. Partijen die vertrouwen op asymmetrische versleuteling moeten vertrouwen op de openbare sleutel om de gegevens te versleutelen en op de privésleutel om deze te ontsleutelen. Openbare en privésleutels zijn wiskundig gerelateerd, maar worden op passende wijze onderscheiden door de introductie van willekeur in het coderingsproces om te voorkomen dat de privésleutel kan worden vastgesteld.
3. Homomorfe encryptie maakt berekeningen op versleutelde gegevens mogelijk. Berekenen op versleutelde gegevens verwijst naar het feit dat een partij *Pn* die de initiële identifiers of invoer *mn* heeft en de functie *f* wil berekenen om *f*(*m1*,…,*mn*) te verkrijgen, in plaats daarvan de versleuteling of pseudoniemen van de invoer *cn* kan berekenen om *f'*(*c1*,…,*cn*) te verkrijgen, die ontcijferd kan worden tot *f*(*m1*,…,*mn*). Het voordeel van homomorfe encryptie is dat persoonsgegevens vertrouwelijk blijven terwijl ze worden geanalyseerd of 'gemined' zonder dat ze moeten worden ontsleuteld en de uitvoer in gevaar komt.
4. Multiparty computation (MPC) verschilt van de drie eerder besproken technieken, hoewel het

gerelateerd is aan homomorfe encryptie. MPC is een techniek die zich toelegt op protocollen waarmee een reeks partijen gezamenlijk een functie van hun invoer of identificatiegegevens kan berekenen, terwijl wordt vermeden dat iets anders wordt onthuld dan de uitvoer van de genoemde functie. MPC zorgt ervoor dat de input van de partijen in het algemeen geheim blijft tijdens de gehele verwerking van data-aggregatie, en wordt dus beschouwd als een geavanceerd privacy-behoudend instrument voor pseudonimisering. Het kan worden gebruikt als een encryptietechniek, maar is veel breder in termen van mogelijke toepassingen.

In de technische literatuur worden encryptie- en pseudonimiseringstechnieken meestal beschreven als vormen van privacyverhogende of -behoudende technologieën. Welke techniek het meest geschikt is, hangt af van de context, het type gegevens, de betrokken actoren en andere aanwezige waarborgen. Dat is de reden waarom, hoewel sommige technieken over het algemeen als zwakker worden beschouwd dan andere, van geen enkele techniek kan worden gezegd dat deze de voorkeur heeft, en geen enkele techniek categorisch kan worden uitgesloten. Sommige pseudonimiserings- of encryptietechnieken, vooral wanneer ze worden toegepast in combinatie met andere privacyverhogende technologieën, kunnen zo sterk zijn dat ze de belangen van betrokkenen beter kunnen beschermen dan bepaalde anonimiseringstechnieken.

Inferentie van gevoelige gegevens

Uit de technische literatuur blijkt duidelijk dat het steeds gemakkelijker wordt om persoonsgegevens af te leiden uit geaggregeerde gegevens en gevoelige persoonsgegevens uit persoonsgegevens of niet-persoonsgegevens. Statistische organisaties en volkstellingsbureaus publiceren bijvoorbeeld vaak geaggregeerde datasets, die volgens hen geen persoonlijke informatie bevatten. Maar door middel van zogenaamde databasereconstructie-aanvallen is het vaak mogelijk om het geslacht, de leeftijd, het ras, de etniciteit en gedetailleerde geografische locatie te reconstrueren die is vastgelegd voor ongeveer de helft van de bevolking in de dataset. Bovendien kunnen door het combineren van twee datasets die zelf geen persoonsgegevens bevatten, persoonsgegevens worden verkregen en zelfs gevoelige persoonsgegevens worden afgeleid. Bijgevolg kunnen zowel door de beschikbaarheid van open data als door de toegenomen technologische capaciteiten om data af te leiden, gevoelige gegevens worden gedestilleerd uit zowel ‘gewone’ persoonsgegevens als uit niet-persoonsgegevens. Zowel wetenschappelijke literatuur als de voor dit onderzoek geïnterviewde experts benadrukken dat deze trend in de loop der tijd alleen maar zal toenemen.

5. De kloof tussen het wettelijke regime en de technologische realiteit

Er zijn verschillende spanningsvelden tussen het technologische domein en de manier waarop het wettelijk kader is opgesteld. De in het kader van dit onderzoek belangrijksten zullen hieronder worden toegelicht.

Anonieme gegevens

1. Hoewel het wettelijk kader onderscheid maakt tussen anonieme en pseudonieme gegevens, is dit onderscheid voor technische experts niet onomstreden. Vanuit technisch oogpunt zouden data anoniem genoemd kunnen worden wanneer een aantal relevante variabelen worden verwijderd. Veel technische experts gaan uit van niveaus van anonimiteit. Er is een schaal van volledige anonimiteit naar directe identificeerbaarheid in plaats van een binair onderscheid, zoals is vervat in de AVG.
2. Het feit dat de AVG geen tijdslimiet stelt aan wanneer gegevens opnieuw kunnen worden geïdentificeerd of geanonimiseerd, betekent dat het zeer waarschijnlijk is dat de gegevens op een bepaald moment aan een natuurlijke persoon zullen worden gekoppeld en dus als persoonsgegevens dienen te worden aangemerkt.

3. Een aantal technische experts vraagt zich af of de wettelijke definitie van anonieme gegevens in de 21e eeuw kan worden gehandhaafd, aangezien het steeds moeilijker zal worden om aan de wettelijke drempel te voldoen. Vanuit technologisch oogpunt is het bijna onmogelijk om over echt anonieme gegevens te beschikken. Met name wanneer geanonimiseerde datasets worden gedeeld of online beschikbaar worden gesteld, is de kans groot dat er een partij is die de data her-identificeert of samenvoegt met andere datasets om tot persoonsgegevens te komen.
4. Sommige auteurs concluderen dat de stand van de techniek die verband houdt met de technieken die door de Artikel 29 Werkgroep zijn opgesomd, bevestigt dat anonimiseringsmethoden voor grote uitdagingen staan met betrekking tot de originele gegevens en dat dit niet langer vanuit een statisch perspectief kan worden beschouwd, maar een dynamisch perspectief vereist.
5. Veel technische experts vinden de juridische definitie van anonimisering onduidelijk en vaag. Een bijzonder punt van aandacht is de term 'redelijkerwijs valt te verwachten'.

Geaggregeerde gegevens

1. Wettelijke regelgeving behandelt (micro)data en geaggregeerde (macro)data hetzelfde, terwijl er wel duidelijk verschillende risico's verbonden zijn aan het openbaar maken van micro- en macrodata. Op datasetniveau vereist het spreken van absoluut anonieme gegevens in het algemeen dat de dataset zodanig wordt gestript dat er vrijwel geen relevante informatie overblijft, terwijl er op geaggregeerd niveau veel meer mogelijkheden zijn om individuen te beschermen tegen identificatie. Op hun beurt kunnen geaggregeerde gegevens op andere manieren aan personen worden gekoppeld, met name wanneer deze online beschikbaar worden gesteld.
2. De toegenomen beschikbaarheid van open data maakt het moeilijk om een goede inschatting te maken van de risico's die gepaard gaan met het vrijgeven van geaggregeerde gegevens door statistische organisaties of andere partijen.
3. Statistische gegevens worden gebruikt om kennis te genereren door middel van analyse van bestaande data om zo aannames over individuen te doen, bijvoorbeeld, door het in kaart brengen van ervaringen uit het verleden en het vastleggen van correlaties tussen bepaalde karakteristieken, bepaalde uitkomsten, of bepaald gedrag. Met AI- en Big Data-analyse kunnen mensen op bruikbare manieren worden geprofileerd zonder persoonlijk of individueel te worden geïdentificeerd. Aangezien de stand van de technologie het mogelijk maakt om meer informatie uit niet-persoonsgegevens te halen, wordt een grotere rol toegekend aan het gebruik van dergelijke gegevens. Deze trend kan niet adequaat worden aangepakt door het gegevensbeschermingskader, dat sterk is gebaseerd op de notie van de identificatie van individuele natuurlijke personen.
4. Voor veel technische experts en professionals geeft het juridische regime tegenstrijdige signalen. Enerzijds worden open data, hergebruik van overheidsinformatie en dataportabiliteit bevorderd; anderzijds wordt de nadruk gelegd op privacy, geheimhouding en gegevensbescherming. Wat deze spanning complexer maakt, is dat wetgevers en rechters geen uniform beeld geven van in hoeverre het verzamelen en gebruiken van geaggregeerde gegevens gereguleerd moet worden of in hoeverre geaggregeerde gegevens ook persoonsgegevens kunnen zijn.

Pseudonieme gegevens

1. Technische experts gaan er traditioneel van uit dat pseudonimisering betekent dat een of meer identifiers worden vervangen door een pseudoniem. De AVG definieert pseudonimisering echter als verwerking waarbij aanvullende informatie, die her-identificatie mogelijk maakt, op een andere plaats wordt opgeslagen; er kunnen pseudonieme gegevens zijn zonder een expliciet pseudoniem. De twee definities liggen zeer dicht bij elkaar, maar zijn niet volledig identiek.
2. Technisch gezien is het niet duidelijk waarom pseudonimisering, als vorm van risicopreventie, een bijzondere status zou moeten hebben binnen het wettelijk kader, omdat er meerdere

manieren en technieken zijn om dit te bewerkstelligen. De voorkeur geven aan deze ene techniek lijkt in schril contrast te staan met de veronderstelde technologische neutraliteit van het gegevensbeschermingskader.
3. Niet alle vormen van pseudonimisering zijn even veilig. Het wettelijke regime geeft geen richtlijnen over welk type techniek het meest geschikt is voor welk type context.

Gevoelige persoonsgegevens

1. Experts zetten vraagtekens bij de vaste categorieën gevoelige gegevens die in de AVG worden gebruikt. Zo zouden de financiële positie, de sociaaleconomische achtergrond of inkomen in veel gevallen als gevoelige gegevens kunnen worden behandeld, omdat de potentieel nadelige effecten van het verwerken van dergelijke gegevens op zijn minst even ernstig kunnen zijn als het verwerken van, bijvoorbeeld, het lidmaatschap van een politieke organisatie of een vakbond. Zij wijzen erop dat wat als gevoelige persoonsgegevens moet of kan worden beschouwd, verschilt per regio of land. Daarom kan het werken met één vaste lijst van soorten gevoelige gegevens voor alle EU-landen bijzonder uitdagend zijn.
2. Het wettelijk regime richt zich op vaste categorieën gegevens, terwijl wat technisch wel of niet gevoelig is, niet afhankelijk is van het type gegevens. Gegevensverwerking kan gevoelig en schadelijk zijn, zelfs zonder de verwerking van de categorieën gegevens die in de AVG worden vermeld, of kan niet-schadelijk zijn, zelfs als een of meer van de soorten gegevens die als gevoelig zijn gecategoriseerd, worden verwerkt.
3. Veel technische experts wijzen op het feit dat gevoelige informatie vaak kan worden afgeleid uit niet-gevoelige persoonsgegevens en zelfs uit niet-persoonsgegevens. Het in het wettelijke regime gehanteerde binaire onderscheid tussen gevoelige en niet-gevoelige gegevens houdt onvoldoende rekening met de technologische complexiteit en realiteit op dit punt.

6. Reguleringsalternatieven gevonden in wet en literatuur

Er zijn alternatieven voorgesteld voor het huidige regulerende regime. De meest relevante voor de doeleinden van dit onderzoek zijn:

Anonieme gegevens

1. Maak een einde aan het onderscheid tussen anonieme en niet-anonieme gegevens. Als het steeds waarschijnlijker wordt dat gegevens worden gedeanonimiseerd en als niet-persoonsgegevens kunnen worden gebruikt voor ingrijpende gegevensprocessen, kan de keuze om anonieme of niet-persoonsgegevens buiten de reikwijdte van de gegevensbeschermingswetgeving te plaatsen overbodig worden.
2. Gebruik een minder breed concept dan de huidige definitie van persoonsgegevens om een duidelijker onderscheid te maken tussen persoonsgegevens en geanonimiseerde gegevens. Het begrip ‘identificeerbaar’ zou bijvoorbeeld kunnen worden geschrapt, of er kan een specifieke horizon of tijdslimiet aan worden toegevoegd.
3. Creëer verschillende niveaus van identificeerbaarheid, wat betekent dat de toepassing van het gegevensbeschermingskader geen zwart-witkwestie is, maar een geleidelijke schaal, bijvoorbeeld des te meer gegevens worden geanonimiseerd des te minder normen voor gegevensbescherming van toepassing zijn.
4. In plaats van te werken met het zeer contextuele 'alle middelen die redelijkerwijs door de verwerkingsverantwoordelijke of door een andere persoon kunnen worden gebruikt', zou kunnen worden overwogen om de AVG aan te passen met een zin die werd voorgesteld in het wetgevingsproces van de Richtlijn bescherming persoonsgegevens, namelijk 'ten koste van een buitensporige inspanning'.

Geaggregeerde gegevens

1. In plaats van een compromis te zoeken tussen open data en gegevensbescherming, zou een optie kunnen zijn om de gegevensbeschermingsregeling te laten prevaleren of het kader te laten bieden voor het gebruiken, delen en openbaar maken van statistische en geaggregeerde gegevens. Dit is in wezen wat het HvJ EU heeft gedaan in *Latvijas Republikas Saeima*.
2. Een nog verdergaande optie zou kunnen zijn om terug te keren naar een van de eerdere regels over statistische gegevens, namelijk dat 'statistische gegevens alleen in geaggregeerde vorm mogen worden vrijgegeven als het onmogelijk is om de informatie aan een bepaalde persoon te koppelen.'
3. Er zou een uitgebreid kader kunnen komen om de behoefte aan open data en het verwerken van statistische gegevens enerzijds te verzoenen met de behoefte aan privacy en gegevensbescherming anderzijds. Een dergelijk kader zou op EU-niveau moeten worden aangenomen en niet aan de lidstaten moeten worden overgelaten, en zou in detail moeten specificeren hoe deze twee beginselen in concrete situaties met elkaar kunnen worden verenigd.
4. Het gegevensbeschermingskader zou explicieter kunnen zijn in termen van een drempel of grens van gegevensanonimiteit wanneer gegevens worden geaggregeerd of in termen van de technische normen die moeten worden toegepast bij het vrijgeven van geaggregeerde gegevenssets.
5. Een radicaler alternatief zou kunnen zijn om manieren te vinden om de regelgeving inzake privacy en gegevensbescherming te baseren op andere concepten dan identificeerbaarheid. Zo hebben sommige auteurs voorgesteld om de focus op individuele privacy en identificeerbaarheid van natuurlijke personen los te laten en in plaats daarvan of in aanvulling daarop meer nadruk te leggen op de bescherming van groepen, categorieën en datacollectieven.
6. Er kunnen concretere regels voor het openbaar maken van geaggregeerde gegevens worden ontwikkeld, zoals het hebben van een minimum *n* per cel met een frequentieteltabel of regels over dominantie met kwantitatieve magnitudetabellen. Ook zouden controles voor het vrijgeven van groepsdata kunnen worden gegeven.

Pseudonieme gegevens

1. De AVG of het Europees Comité voor gegevensbescherming (European Data Protection Board - EDPB) zou meer richtsnoeren kunnen geven ten aanzien van welke soorten pseudonimiseringstechnieken het meest geschikt worden geacht voor welke context.
2. Overwogen kan worden om pseudonimisering af te stemmen op het concept van gegevensbeheer. Een gegevensbeheerder kan fungeren als een pseudonimiseringsentiteit die verantwoordelijk is voor het verwerken van pseudoniemen, die onder specifieke voorwaarden gegevenstoegang kan verlenen aan onderzoekers of bedrijven en gegevens kan afschermen tegen ongewenste of onrechtmatige toegang.
3. Sommige deskundigen stellen voor om de specifieke verwijzing naar pseudonimisering uit de AVG te schrappen, zowel omdat het als te vaag wordt beschouwd als omdat er geen reden is om deze techniek te verkiezen boven andere risicomijdingstechnieken.
4. Anderen daarentegen hebben gesuggereerd om de categorie pseudonieme gegevens een nog prominentere rol te geven, waardoor het een officiële tussencategorie wordt tussen anonieme gegevens en persoonsgegevens.

Gevoelige persoonsgegevens

1. Verschillende auteurs hebben gesuggereerd dat de gevoeligheid van gegevensverwerking niet langer afhangt van het type gegevens dat wordt verwerkt, maar veeleer van de verwerkingstechnologieën en het gebruik ervan. Daarom hebben zij voorgesteld om de bijzondere regeling voor gevoelige persoonsgegevens uit het gegevensbeschermingskader weg

te laten.

2. Verruim de lijst met gevoelige gegevens en neem daarin o.a. financiële gegevens op, zoals bij het opstellen van de AVG werd gesuggereerd maar uiteindelijk werd afgewezen.
3. Als alternatief is voorgesteld om te werken met een lijst met voorbeelden in plaats van met vaste categorieën, wat de oorspronkelijke benadering was voor de regulering van gevoelige persoonsgegevens. Ook zou kunnen worden overwogen om een restcategorie in te voeren, vergelijkbaar met de verwijzing naar 'of andere status' in artikel 14 van het Europees Verdrag voor de Rechten van de Mens (EVRM).
4. Er zou kunnen worden overwogen om onderscheid te maken tussen de verschillende categorieën gevoelige persoonsgegevens, een benadering die lijkt te worden gevolgd door het HvJ EU, waarbij gezondheidsgegevens in de meest gevoelige categorie worden geplaatst, terwijl andere gegevens als minder gevoelig kunnen worden beschouwd.
5. Hoewel de meeste zorgen gaan over de vraag of de AVG strikt genoeg is voor speciale categorieën gegevens, zijn er ook tegengestelde argumenten. Er is een voortdurende discussie in hoeverre het mogelijk is om gevoelige persoonsgegevens te verwerken om discriminatie te voorkomen, bijvoorbeeld in KI-systemen. Het gebruik van gevoelige persoonsgegevens kan nodig zijn om discriminatie te voorkomen, vooral als het gaat om data gedreven besluitvorming. Om een van de onderliggende grondgedachten van de categorie bijzondere gegevens te bevorderen, namelijk het voorkomen van discriminerende praktijken, kan het dus nodig zijn om meer gevoelige persoonsgegevens te verwerken in plaats van minder.

7. Reguleringsdoelstelling van de gegevensbeschermingsregeling

Om te beoordelen of er lacunes in de regelgeving zijn en, zo ja, welke, is het nodig om te beoordelen wat de reguleringsdoelstelling van het privacy- en gegevensbeschermingsregime eigenlijk is. Dit is een punt van discussie: moet gegevensbescherming worden gezien als een regime dat grenzen en beperkingen oplegt aan verwerkingsverantwoordelijken of gaat het primair om de controlerechten van betrokkenen?

Enerzijds kan worden verwezen naar artikel 5 van de Algemene Verordening Gegevensbescherming, dat als de ruggengraat van deze wet wordt gezien. Het stelt dat persoonsgegevens rechtmatig, behoorlijk en op transparante wijze moeten worden verwerkt, moeten worden verzameld voor gespecificeerde, expliciete en legitieme doeleinden en niet verder moeten worden verwerkt op een manier die onverenigbaar is met die doeleinden, adequaat, relevant en beperkt moeten zijn tot wat nodig is met betrekking tot de doeleinden waarvoor ze worden verwerkt, nauwkeurig en, waar nodig, actueel moeten zijn, niet langer dan nodig worden bewaard in een vorm die identificatie van betrokkenen mogelijk maakt, en verwerkt op een wijze die zorgt voor een passende beveiliging van de persoonsgegevens. Dit zijn allemaal verplichtingen die op de verwerkingsverantwoordelijke rusten en die van toepassing zijn onafhankelijk van eventuele rechten die door betrokkenen worden ingeroepen. Aan de andere kant worden er in het gegevensbeschermingsregime steeds meer rechten toegekend aan betrokkenen. Bovendien is, vooral door Duitse invloed, het begrip informatieve zelfbeschikking steeds populairder geworden. Daarom stellen sommigen dat, in plaats van verplichtingen die aan de verwerkingsverantwoordelijken worden opgelegd, de rechten van de betrokkenen de kern vormen van het gegevensbeschermingsregime.

Wat de beoordeling van deze kwestie bemoeilijkt, is het feit dat het gegevensbeschermingsregime niet alleen een beschermend doel heeft, maar dat artikel 1 AVG ook als doel erkent om de verwerking van persoonsgegevens in de EU te faciliteren. Een van de expliciete doelstellingen van het gegevensbeschermingskader van 1995 was het wegnemen van belemmeringen voor de doorgifte van persoonsgegevens binnen de Unie door één gemeenschappelijk niveau van gegevensbescherming vast te stellen. Vóór de richtlijn had elk land eigen normen voor gegevensbescherming, wat het gebruik en de doorvoer van persoonsgegevens belemmerde. Het aannemen van één EU-breed kader voor

gegevensbescherming loste dit probleem op. De regels in de AVG verbieden zelden specifieke gegevensverwerkingen. In de meeste gevallen bevatten ze procedurele waarborgen en beginselen die een correcte gegevensverwerking garanderen. Gegevensbescherming kan dus worden gezien als het bevorderen van gegevensverwerking door het bieden van een algemeen kader.

Tot slot bevat de AVG veel expliciete uitzonderingen voor specifieke verwerkingen. De belangrijkste in het kader van dit onderzoek zijn die met betrekking tot de vrijheid van meningsuiting, archivering, statistisch onderzoek, open overheid en het hergebruik van overheidsinformatie. De EU heeft de keuze gemaakt om verder te gaan dan het bevorderen van openheid en transparantie ten aanzien van overheidspraktijken; het heeft het hergebruik van overheidsinformatie gestimuleerd. Toch maakt de Open Data Richtlijn duidelijk dat deze geen invloed heeft op de AVG; wat dit in de praktijk betekent wordt opengelaten.

Binnen de EU bestaat onduidelijkheid over hoe om te gaan met de conflicten tussen de verschillende regimes en de beschermingsdoeleinden die eraan ten grondslag liggen. Over het algemeen neemt de EU-regelgever instrumenten aan die zijn gebaseerd op duidelijke en afgebakende gegevenscategorieën, terwijl er in de rechtspraak wordt gekozen voor een contextuele benadering. Adviesorganen zoals de Artikel 29-Werkgroep en de EDPB propageren ook een flexibele benadering en hebben in de loop der tijd de reikwijdte van onder meer persoonsgegevens verruimd. Gerechtshoven hebben duidelijke grenzen gesteld wanneer regelgevers onderscheid tussen typen data gebruiken om lagere beschermingsniveaus in te voeren, zoals toen het HvJ EU de EU dataretentie-richtlijn nietig verklaarde. Een vergelijkbare benadering is waar te nemen met betrekking tot de verschuiving naar open data en het hergebruik van overheidsinformatie. Hoewel dit sterk wordt gestimuleerd door de EU-wetgever, zijn rechters terughoudender. Zo vroeg het HvJ zich af of het voor het beschermen of verbeteren van de verkeersveiligheid noodzakelijk was om toegang te verlenen tot gegevens over verkeersovertredingen. Het stelde vast dat het regime derden toegang gaf tot de informatie, zelfs als die derden andere doeleinden hadden dan die welke verband hielden met het vergroten van de verkeersveiligheid, wat niet was toegestaan.

8. Gevaren van over- en onderregulering

De moeilijkheid bij het beoordelen van het bestaan van lacunes in de regelgeving en de wenselijkheid van reguleringsalternatieven is dat eerst de discussie over de reguleringsdoel(en) van het privacy- en gegevensbeschermingskader moet worden beslecht, terwijl dat een punt van discussie blijft. Bovendien is er geen duidelijke voorkeur in reguleringsbenadering: een categorale, een contextuele of een hybride. Elk heeft zijn eigen voor- en nadelen. Het is dus zowel een kwestie van smaak of er lacunes in de regelgeving zijn en zo ja, wat dat in de toekomst betekent voor de rechtsbescherming van gegevens in brede zin. Bovendien houdt de keuze tussen verschillende reguleringsopties een keuze in waar het reguleringsprerogatief wordt geplaatst. Hoe meer duidelijkheid er wordt verschaft in het wettelijke regime, hoe meer het prerogatief bij de wetgevende macht wordt gelegd. Hoe meer een contextuele benadering wordt gevolgd, hoe meer de rechterlijke en/of de uitvoerende macht de juiste interpretatie van de regels per context moet geven. Het eerste heeft het voordeel van democratische legitimiteit, het tweede van praktische toepasbaarheid. Het eerste heeft het voordeel dat het rechtszekerheid biedt door voor alle situaties één benadering in te voeren; het tweede heeft het voordeel dat het in staat is om granulariteit te bieden in regulering.

Om een voorbeeld te geven, misschien is de essentiële vraag die deze studie oproept wel of het begrip 'persoonsgegevens' en de subcriteria 'identificeerbaarheid' en ‘middelen redelijkerwijs valt te verwachten’ moeten worden behouden, of dat niet-persoonsgegevens moeten worden beschermd, bijvoorbeeld onder een AVG-light regime, of dat er een geleidelijke schaal naar identificatie moet worden ingevoerd. Die vraag hangt af van wat de reguleringsgrondgedachte van het gegevensbeschermingskader wordt geacht te zijn. Als het de individuele belangen van natuurlijke

personen beschermt, is er geen directe noodzaak om ook de verwerking van geaggregeerde of anonieme gegevens te regelen. Om mogelijke schade aan te pakken die voortvloeit uit beleid en acties op basis van groepsprofielen kan het aan de rechterlijke macht worden gelaten om het regulerende regime zo te interpreteren dat deze schade wordt gedekt, hetzij op basis van de AVG, hetzij op grond van artikel 8 EVRM. Als het doel van de regelgeving is om de datamacht van organisaties in de publieke en private sector in te perken, dan is het logisch om ook beperkingen en eisen te stellen aan de verwerking van niet-persoonsgegevens, en zou het geen probleem zijn om het gegevensbeschermingsregime uit te breiden om ook de verwerking van niet-persoonsgegevens te dekken en de materiële reikwijdte ervan los te koppelen van de identificeerbaarheid van een natuurlijke persoon. Beide keuzes roepen bovendien de vraag op van de specificiteit van de regelgeving. De toezichthouder handhaaft tot nu toe een strikt reguleringsonderscheid tussen niet-persoonsgegevens en persoonsgegevens, maar in de praktijk is dit onderscheid moeilijk te maken. Rechters hebben bijgevolg de definitie van persoonsgegevens uitgebreid tot gegevens die steeds meer perifeer zijn aan natuurlijke persoon, terwijl de verwerkingsverantwoordelijken om meer aanwijzingen vragen om dat onderscheid te maken. Het gevaar van het intact laten van de huidige aanpak is dat verantwoordelijke data-organisaties aan de veilige kant blijven, terwijl anderen de grenzen van de wet oprekken. Bovendien geldt dat hoe minder duidelijkheid in de regelgeving wordt gegeven, hoe moeilijker het zal zijn om de regels te handhaven, omdat elke gegevensverwerking mogelijk een eigen rechtmatigheidsbeoordeling vereist. Wanneer de keuze wordt gemaakt om het huidige regulerende regime intact te laten, is het dus nog steeds de vraag of er meer reguleringsrichtsnoeren moeten worden gegeven aan verwerkingsverantwoordelijken over het maken van onderscheid tussen gegevenscategorieën.

Bovendien, wanneer de keuze wordt gemaakt om niet-persoonsgegevens te onderwerpen aan wetgeving kunnen opnieuw twee verschillende benaderingen worden gevolgd: een categorale en een contextuele. Ofwel handhaaft het reguleringsregime een onderscheid tussen niet-persoonsgegevens en persoonsgegevens, maar hecht het een ander reguleringsregime aan niet-persoonsgegevens, ofwel heft het deze differentiatie en mogelijk andere gegevensonderscheiden op, en maakt het regime het type regels en de regeldruk voor de verwerkingsverantwoordelijken afhankelijk van de beoordeling van de betrokken risico's per geval (danwel gerelateerd aan individuele, groeps- en/of maatschappelijke belangen).

Voor de kwestie van overregulering is het van belang in hoeverre het stimuleren van gegevensverwerkingen op dezelfde voet wordt geplaatst als het beschermingsdoel van het gegevensbeschermingsregime en hoe het doel om open data-omgevingen en het hergebruik van overheidsinformatie te bevorderen wordt beoordeeld. Moet het bevorderingsdoel worden gezien als een even belangrijke grondgedachte als het beschermingsdoel, of kan deze grondgedachte alleen worden bevorderd binnen de grenzen die voortvloeien uit het beschermingsdoel? Als dat laatste het geval is, is overregulering geen wezenlijk risico, terwijl het voorkomen van onderregulering het hoofddoel is. Als beide doelen echter op dezelfde voet worden geplaatst, heeft het bevorderen van het ene doel bijna per definitie gevolgen voor het andere. Vervolgens rijst de vraag welk type regulering het meest effectief is. Hoewel een contextueel kader de meeste ruimte lijkt te laten voor data-innovatie, pleiten verwerkingsverantwoordelijken op het eerste gezicht vaak expliciet voor meer duidelijkheid en zekerheid op het gebied van regelgeving, omdat ze bang zijn voor terugslag en investeringen die niet lonend zijn.

Een soortgelijk punt moet worden opgemerkt met betrekking tot de beschermende grondgedachte. Experts hebben benadrukt dat de benadering van de AVG, waarbij het verwerken van gevoelige gegevens in principe verboden is, steeds meer het doel mist dat het beoogt, namelijk het beschermen van individuen tegen schade. Om discriminerende praktijken in KI-systemen te voorkomen, kan het nodig zijn om gevoelige persoonsgegevens te verwerken. Anderen hebben benadrukt dat het zelfs in bredere zin nodig kan zijn om niet te focussen op gegevensminimalisatie maar op *gegevensminimumisatie*, op de eis dat een minimumniveau van gegevens wordt verzameld,

geanalyseerd en opgeslagen. Het is dus zelfs een kwestie van debat of het beschermingsdoel uit de AVG het best gediend is door beperkingen op dataprocessen op te leggen.

9. Hoe zullen de huidige en toekomstige technische ontwikkelingen de komende periode van invloed zijn op de AVG en rechtsbescherming in brede zin?

Het is duidelijk dat de technologische ontwikkelingen en de algemene beschikbaarheid van data nu en in de toekomst tot gevolg hebben dat anonimisering steeds moeilijker wordt. De status van data wordt steeds volatieler en wordt steeds minder een kenmerk van data en datasets zelf en steeds meer een effect van de inspanningen van de verwerkingsverantwoordelijke. De juridische categorieën zullen steeds meer fluïde en minder stabiel worden en één database kan juridisch verschillend zijn per partij die er toegang toe heeft. Een database die op zichzelf alleen niet-persoonsgegevens bevat, kan worden omgezet in een database met persoonsgegevens door deze te combineren met een andere database, kan worden gebruikt om gevoelige persoonsgegevens te verkrijgen, om het volgende moment weer te worden geaggregeerd en geanonimiseerd. Gezien deze trends en gezien de begrippen 'identificeerbaarheid' en 'alle middelen die redelijkerwijs kunnen worden gebruikt', zullen steeds meer gegevens, zo niet alle, onder het gegevensbeschermingskader vallen en moeten ze mogelijk zelfs worden behandeld als (potentiële) gevoelige persoonsgegevens, waardoor het strengste van alle regimes zou gelden.

Of dit als problematisch wordt beschouwd, staat ter discussie en hangt af van wat de grondgedachte van het gegevensbeschermingskader is en welke effecten van onder- en overregulering het meest waarschijnlijk zijn. In dit onderzoek zijn geen verschillende scenario's gevonden voor hoe het technologische domein en de beschikbaarheid van open data zich in de loop van de tijd zullen ontwikkelen. Literatuur, geïnterviewde experts en experts die zijn uitgenodigd voor de workshop die voor dit onderzoek is gehouden, wijzen allemaal in dezelfde richting: anonimisering wordt steeds moeilijker, juridische categorisering zal steeds moeilijker worden en de status van gegevens zal steeds meer een effect zijn van de inspanningen van de verwerkingsverantwoordelijke. Er zijn wel meerdere scenario's gevonden voor hoe het wettelijk regime zou kunnen reageren op de toegenomen beschikbaarheid van open data en de algemene toegankelijkheid van dataverwerkingstechnologieën. Uit de suggesties kunnen vijf strategieën worden afgeleid, die als volgt kunnen worden samengevat:

1. **Het gegevensbeschermingskader laten zoals het is:** Het gegevensbeschermingskader wordt beschouwd als een perfect evenwicht tussen de beschermende grondgedachte en de bevorderende grondgedachte, tussen de keuze voor een categorale en een contextuele reguleringsbenadering en tussen het overlaten van het reguleringsprerogatief aan de wetgever en tegelijkertijd de rechterlijke en uitvoerende autoriteiten in staat te stellen concepten en regels in de praktijk te verfijnen, met het oog op specifieke contexten en situaties. Hoewel kan worden gezegd dat de technologische praktijk afwijkt van het reguleringsregime en dat deze afwijking in de loop der jaren heel goed steeds groter zou kunnen worden, betekent dit niet dat de regels moeten veranderen. Er moet veeleer meer worden geïnvesteerd om ervoor te zorgen dat de praktijk in overeenstemming blijft met de regels. Voor zover de verwerking van niet-persoonsgegevens een belangrijke impact heeft, wordt deze al gedekt door de AVG wanneer besluiten worden genomen waarin een individu wordt geïdentificeerd of op een persoonlijke manier wordt getroffen, of door artikel 8 EVRM, wanneer beleid van invloed is op het zeer brede begrip van privéleven. Het EHRM is bereid geweest om een regime te ontwikkelen voor het verzamelen van metadata, om een reguleringslacune te voorkomen, en heeft claims geaccepteerd waarin de eiser geen persoonlijk nadeel heeft geleden, maar heeft zich gericht op de maatschappelijke effecten van grootschalige gegevensverwerking. De wetgeving inzake gegevensbescherming hoeft bovendien niet alle problemen van de datagedreven omgeving op te lossen.
2. **Handhaving van het gegevensbeschermingskader en investeren in nauwkeurigere definities:** De hoofdlijnen en contouren van het huidige reguleringsstelsel worden geschikt

geacht voor de 21e eeuw, terwijl de belangrijkste uitdaging op het gebied van regelgeving de behoefte is aan meer duidelijkheid over de definities van de verschillende gegevenscategorieën, de grenzen tussen verschillende categorieën en de regulering van dat soort gegevens. In dit scenario zijn verschillende reguleringsalternatieven mogelijk, zoals het uitvaardigen van meer richtlijnen en het invoeren van een bewijslast voor de verwerkingsverantwoordelijke om aan te tonen dat gegevens anoniem en/of versleuteld zijn. Om meer duidelijkheid te scheppen over het onderscheid tussen niet-persoonsgegevens en persoonsgegevens, zouden de contextuele elementen in de definitie van persoonsgegevens en in de beschrijving van anonimisering kunnen worden verwijderd. Dit zou de vraag of persoonsgegevens worden verwerkt en of het kader voor gegevensbescherming van toepassing is, de-contextualiseren. Ook kan de categorie pseudonieme gegevens worden weggelaten. Deze categorie wordt zowel bekritiseerd vanwege zijn vaagheid als omdat het één privacybeschermende techniek voorrang geeft boven anderen, waarvoor geen duidelijke verklaring bestaat. Ten slotte kan worden overwogen om de lijst met gevoelige persoonsgegevens uit te breiden. Mogelijke aanvullende categorieën die in dit onderzoek zijn geïdentificeerd, zijn onder meer financiële en sociaaleconomische gegevens, gegevens over kinderen, locatiegegevens en metagegevens.

3. **Het gegevensbeschermingskader behouden en investeren in meer contextualiteit:** Als de belangrijkste uitdaging op het gebied van regelgeving wordt beschouwd het gebrek aan contextualiteit en aanpasbaarheid van het huidige reguleringsstelsel. Tijdens dit onderzoek zijn er verschillende reguleringsalternatieven naar voren gekomen, zoals, maar niet beperkt tot, de toevoeging van het contextualiteitsbeginsel aan de lijst van artikel 5 AVG, waarbij de verwerkingsverantwoordelijke wordt verplicht elk principe, elke verplichting en elk vereiste onder het gegevensbeschermingsregime in overweging te nemen in het kader van de context waarin de verwerking plaatsvindt. Als alternatief kan worden overwogen om de lijst van gevoelige gegevens te herformuleren zoals deze oorspronkelijk was, namelijk als voorbeelden in plaats van een uitputtende lijst, of om een restcategorie op te nemen, vergelijkbaar met artikel 14 EVRM. Pseudonieme gegevens zouden een meer prominente plaats kunnen krijgen als tussencategorie tussen niet-persoonsgegevens en persoonsgegevens.
4. **Herziening van het gegevensbeschermingskader met gebruikmaking van duidelijk gedefinieerde gegevenscategorieën:** Strategie 4 is vergelijkbaar met strategie 2, maar in dit scenario is een fundamentele herziening van het huidige reguleringskader noodzakelijk. In dit scenario wordt aangenomen dat het nog steeds mogelijk is om met gegevenscategorieën te werken, zelfs de huidige categorieën, maar in het licht van de technologische ontwikkelingen moet het reguleringsregime dat erop wordt toegepast worden heroverwogen. Een aantal reguleringsalternatieven zou kunnen worden overwogen, zoals het aannemen van een AVG-light regime voor niet-persoonsgegevens; dit zou bijvoorbeeld kunnen betekenen dat alle gegevensverwerkingen in overeenstemming moeten zijn met de beginselen van artikel 5 AVG. Ook zou, in het licht van een beschermend regime voor niet-persoonsgegevens, kunnen worden overwogen om het gegevensverwerkingsregime te structureren rond stadia van gegevensverwerking: het verzamelen en opslaan van gegevens, het analyseren van gegevens en het gebruiken van gegevens of de uitkomsten van gegevensanalyse. Het huidige reguleringsregime richt zich vrijwel uitsluitend op het moment dat gegevens worden verzameld en opgeslagen. Er zijn vrijwel geen regels voor de analyse van gegevens en voor het gebruik van gegevens, misschien met uitzondering van één bepaling over het verbod op geautomatiseerde besluitvorming. Dit wordt als problematisch ervaren omdat de kern van de meeste hedendaagse verwerkingen uit het analyseren van gegevens bestaat. Voor de inkadering van de analyse van gegevens zou inspiratie kunnen worden gezocht in de regels die gelden voor statistische bureaus.
5. **Herziening van het gegevensbeschermingskader, waarbij duidelijk gedefinieerde gegevenscategorieën worden verwijderd:** Strategie 5 is vergelijkbaar met strategie 3, maar in dit scenario is een fundamentele herziening van het huidige reguleringskader noodzakelijk. In dit scenario is het eenvoudigweg onmogelijk om met verschillende gegevensdefinities te werken

en aan elk van deze verschillende niveaus van wettelijke bescherming te koppelen. In plaats daarvan moet een volledig contextuele benadering worden gevolgd, die volledig afhankelijk is van een analyse van geval tot geval van de potentiële schade die het gevolg is van een bepaalde verwerking. Dergelijke schade kan verband houden met individuele belangen en/of maatschappelijke belangen. De meeste van de huidige verplichtingen en vereisten zouden intact kunnen blijven, maar ze zouden afhankelijk worden gemaakt van het risiconiveau. De AVG zou in wezen kunnen worden teruggebracht tot een eenvoudige set regels, namelijk een lijst van principes en verplichtingen voor verwerkingsverantwoordelijken die nu in de verordening staan en daarbij specificeren dat deze op hen van toepassing zijn, rekening houdend met de stand van de techniek, de kosten van uitvoering en de aard, omvang, context en doeleinden van de verwerking, de aard van de gegevens alsmede het risico en de ernst voor de individuele en/of maatschappelijke belangen.

Figuur: Schaal van een volledig categorale benadering (optie 4) naar een volledig contextuele benadering (optie 5)

10. Antwoorden op de onderzoeksvragen

1. Welke middelen zijn er om (anonieme) data terug te koppelen naar individuen en in hoeverre speelt de beschikbaarheid van andere (bijvoorbeeld open source) data een rol?

Er zijn veel middelen beschikbaar om gegevens terug te koppelen naar individuen. Dit onderzoek is niet tot een volledige en uitputtende lijst van mogelijkheden gekomen, maar heeft een aantal gangbare middelen besproken om dit te doen. Voorbeelden zijn databasereconstructieaanvallen (waardoor een geaggregeerde database opnieuw wordt geïdentificeerd), samenstelling (waardoor twee of meer geanonimiseerde datasets samengevoegd kunnen worden tot (gevoelige) persoonsgegevens) en verschillende de-anonimiseringstechnologieën. Uit geanonimiseerde datasets kan informatie worden afgeleid over personen die in eerste instantie niet in de dataset zaten en geaggregeerde data kunnen in het bijzonder worden gebruikt voor besluitvormingsprocessen die een significant effect kunnen hebben op burgers in het algemeen en specifieke groepen in bijzonder. Als dat laatste het geval is, kunnen die gegevens kwalificeren als persoonsgegevens.

Open data speelt hierbij een belangrijke rol, zozeer zelfs dat veel experts erop wijzen dat het weliswaar

mogelijk is om een geïsoleerde dataset te de-individualiseren, maar omdat het mogelijk is om deze te combineren met andere online vrij beschikbare data het nooit kan worden uitgesloten en het integendeel steeds waarschijnlijker zal worden, dat een geanonimiseerde dataset op termijn door een of andere partij wordt ge-deanonimiseerd. Geaggregeerde gegevens kunnen, wanneer ze beschikbaar worden gesteld, worden gebruikt voor besluitvorming die gevolgen heeft voor specifieke geïdentificeerde of niet-geïdentificeerde burgers. Hoe gegevens zullen worden gebruikt, kan vooraf niet met zekerheid worden gecontroleerd of ingeschat. Echter, de kans dat wanneer data online beschikbaar wordt gesteld, deze door een partij worden gebruikt op manieren die effect hebben op concrete individuen, groepen of de samenleving als geheel, wordt steeds groter.

2. Welke (technische) ontwikkelingen worden de komende jaren verwacht met betrekking tot de middelen om gegevens (al dan niet opzettelijk) terug te koppelen aan personen?

Het zal steeds moeilijker worden om de (juridische) anonimiteit van datasets te waarborgen. Experts die voor dit onderzoek zijn geïnterviewd, betwijfelen nu al of het mogelijk is om aan de wettelijke criteria voor anonimiteit te voldoen. Terwijl het wettelijke regime anonimiteit als een binair vraagstuk beschouwt, zien de meeste technische experts het als een schaal. De meeste technologieën en tegen-technologieën zijn verwikkeld in een kat-en-muisspel. Dit wordt ook verondersteld het geval te zijn voor de toekomst van onder meer anonimiserings- en de-anonimiseringstechnieken, aggregatie- en inferentietechnieken en voor encryptie en decryptie. De meest fundamentele verschuiving is de algemene toegankelijkheid van dergelijke technologieën. Dit betekent dat, vooral wanneer gegevens online beschikbaar worden gesteld, het steeds waarschijnlijker wordt dat er wereldwijd enkele partijen zullen zijn die geavanceerde technologieën zullen gebruiken om gegevens te ontsleutelen, opnieuw te identificeren of te de-anonimiseren en de nodige tijd, energie en moeite zullen investeren om dit te doen. Een belangrijke ontwikkeling op het gebied van encryptie is quantum computing.

Quantum computing heeft bepaalde kenmerken die zijn afgeleid van de kwantummechanica en die het mogelijk maken om complexe factorisatieproblemen op te lossen waarmee traditionele computers worstelen. In plaats van met bits te werken, werken kwantumcomputers met kwantumbits of qubits. Qubits kunnen tegelijkertijd een waarde van 0 of 1 aannemen, in tegenstelling tot traditionele bits, die enkel een toestand van ofwel 0 ofwel 1 hebben. Hierdoor kunnen kwantumcomputers meerdere parallelle berekeningen uitvoeren waarvoor conventionele computers niet geschikt zijn. Als gevolg hiervan kan quantum computing mogelijke alle huidige vormen van cryptografie kraken, net zoals de huidige technieken de met Data Encryption Standard (DES) versleutelde berichten van 40 jaar geleden kunnen ontsleutelen.

3. Welke actuele en voorzienbare technische ontwikkelingen kunnen worden gebruikt voor het anonimiseren of pseudonimiseren van persoonsgegevens en welke factoren zijn daarbij bepalend?

Er bestaan verschillende technieken voor zowel anonimisering als pseudonimisering. Voorbeelden van anonimiseringstechnieken omvatten, maar zijn niet beperkt tot: het maskeren en gebruiken van synthetische gegevens. Er zijn verschillende factoren die bepalend zijn, maar veel hangt af van de vraag of er een technische of een juridische benadering wordt gekozen. Ook zijn in de technische literatuur verschillende soorten anonimiteit naar voren gebracht, elk met hun eigen nadruk op verschillende factoren, met als belangrijksten: k-anonimiteit, l-diversiteit, t-nabijheid en ε-differentiële privacy.

Voor aggregatie kan onderscheid worden gemaakt tussen onder meer aggregatie op basis van derden, aggregatie op basis van dataverstoring en aggregatie op basis van cryptografie. Elk daarvan onderstreept verschillende factoren die bepalend worden geacht. Misschien wel de belangrijkste techniek voor het aggregeren van gegevens, vooral in het licht van het vrijgeven van gegevens, is SDC. Er is geen vaste standaard voor SDC; elke organisatie kan zijn eigen factoren, normen en drempels hanteren, rekening houdend met de dataset, de waarde ervan en mogelijke privacy risico's.

Er bestaan verschillende pseudonimiseringstechnieken, de belangrijksten voor de doeleinden van dit onderzoek: hashing, key hashing, salt hashing en pepper hashing. Encryptie wordt juridisch gezien beschouwd als een deelverzameling van pseudonimisering. Er bestaan verschillende encryptietechnieken, de belangrijksten: symmetrische encryptie, asymmetrische encryptie, homomorfe encryptie en multiparty-computation (wat meer is dan enkel een encryptietechniek). De laatste is een techniek die zich bezighoudt met protocollen waarmee een reeks partijen gezamenlijk een functie van hun invoer of identificatiegegevens kan berekenen, terwijl wordt vermeden dat iets anders wordt onthuld dan de uitvoer van die functie.

4. Welke technische ontwikkelingen op het gebied van anonimisering en pseudonimisering van persoonsgegevens zijn de komende jaren te verwachten?

De meeste geïnterviewde experts en de voor dit onderzoek geëvalueerde literatuur verwachten geen technologische revolutie op het gebied van anonimisering en pseudonimisering, maar verwachten dat het kat-en-muisspel de komende jaren zal doorgaan. Door de steeds grotere beschikbaarheid van gegevens en de algemene toegankelijkheid van technologieën kan het echter nog moeilijker worden om tot anonieme of pseudonieme gegevens te komen. Quantum computing kan, zoals gezegd, een belangrijke impact hebben op encryptie. Daarnaast zal ‘deep learning’ naar verwachting de komende jaren nog meer bekendheid krijgen. Beide technologieën kunnen een nadelig effect hebben op privacy, maar ze kunnen ook in het voordeel van privacy worden ingezet. Post-kwantumversleuteling wordt als veel veiliger beschouwd dan de huidige vormen van versleuteling, en momenteel worden reeds deep privacy-tools (privacy-tools op basis van deep learning-modellen) ontwikkeld.

5. Wat kan er vanuit het juridisch en technisch perspectief gezegd worden over de interpretatie van het begrip ‘alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruikt’? Welke middelen zijn redelijkerwijs in te zetten en welke factoren spelen daarbij een rol?

Vanuit juridisch oogpunt hebben zowel het HvJ EU als de Artikel 29 Werkgroep keer op keer benadrukt dat de beoordeling van welke middelen redelijkerwijs kunnen worden geacht te worden gebruikt per geval moet worden gemaakt, rekening houdend met alle relevante omstandigheden van het geval en met oog op verschillende relevante, maar niet op zich bepalende factoren, zoals de kosten en de tijd die nodig zijn voor identificatie, de beschikbare technologie op het moment van de verwerking en technologische ontwikkelingen. Hoewel dit op zichzelf objectieve criteria zijn, hangt de interpretatie ervan af van de context. Dus hoewel het onderscheid tussen niet-persoonsgegevens en persoonsgegevens juridisch binair en absoluut is, zijn de criteria om te bepalen of gegevens anoniem zijn zeer contextueel.

Vanuit technisch perspectief is de contextuele benadering het meest voor de hand liggend. De meeste technische experts geloven niet in absolute of volledige anonimiteit, maar wijzen eerder op een schaal van hoe moeilijk het is om een database te de-anonimiseren of opnieuw te identificeren. Omdat de technologische mogelijkheden voor de-anonimisering evolueren, moet een beoordeling van de technische normen om gegevens te anonimiseren mogelijk continue of periodiek gebeuren. Een zwart-wit onderscheid tussen anonieme en niet-anonieme gegevens ligt in dit verband niet voor de hand; vanuit technisch oogpunt zou het eerder passender kunnen zijn om te werken met een schaal waarbij hoe anoniemer gegevens zijn, hoe minder (strenge) gegevensbeschermingsnormen van toepassing zijn. Er is geen uitputtende lijst van factoren vanuit een technologisch perspectief waarmee rekening moet worden gehouden om de redelijk waarschijnlijke middelen te bepalen (een juridisch begrip dat in de meeste technologische discussies niet voorkomt).

6. Hoe verhoudt het antwoord op vraag 5 zich tot ontwikkelingen in huidige en toekomstige anonimiserings- en pseudonimiseringstechnieken?

De algemene beschikbaarheid van open data en de algemene toegankelijkheid van datatechnologieën zullen een drievoudig effect hebben op de mogelijkheden om anonimisering en pseudonimisering te realiseren.

Ten eerste is de aard van de data in Big Data-processen niet stabiel, maar volatiel. Een dataset met gewone persoonsgegevens kan worden gekoppeld aan en verrijkt met een andere dataset om gevoelige gegevens af te leiden; de gegevens kunnen vervolgens worden samengevoegd of ontdaan van identificatiegegevens en niet-persoonlijk worden, zoals geaggregeerde of anonieme gegevens; vervolgens kunnen de gegevens worden gedeanonimiseerd of geïntegreerd in een andere dataset om opnieuw persoonsgegevens te creëren. Dit alles kan in een fractie van een seconde gebeuren. De vraag is daarom of het zin heeft om met afgebakende categorieën te werken als dezelfde 'datum' of dataset letterlijk van de ene seconde op de andere in een andere categorie kan vallen en de volgende seconde in weer een andere.

Ten tweede wordt het als gevolg van het voorgaande steeds moeilijker om de status van gegevens precies te bepalen. Om de huidige status van een datum of dataset te bepalen, moet rekening worden gehouden met de verwachte toekomstige status van de gegevens. Gezien de algemene toegankelijkheid van technologieën en de minimale investering die nodig is, wordt het steeds waarschijnlijker dat wanneer een database wordt gedeeld of anderszins beschikbaar wordt gesteld, er een partij is die deze combineert met andere data, deze verrijkt met data van internet geschraapt of samenvoegt in een bestaande dataset, maar ook dat er andere partijen zijn die dat niet willen. De juridische categorie waartoe de gegevens behoren, is dus niet langer een kwaliteit van de gegevens zelf, maar een product van de inspanningen en investeringen van een verwerkingsverantwoordelijke. Bijgevolg is het de vraag of anonimisering of pseudonimisering kan worden bereikt in een context waarin het bepalen van de status van gegevens nauwelijks haalbaar is.

Ten derde zijn moderne gegevensverwerkingen in toenemende mate gebaseerd op geaggregeerde gegevens, die ook zeer grote individuele en sociale gevolgen kunnen hebben. Het profileren van doelgroepen in plaats van individuen wordt een gangbare verwerkingshandeling in de informatiemaatschappij. De gevolgen van deze activiteiten kunnen negatief zijn voor de groep, zonder dat de schade direct te relateren is aan individuen. Het idee dat hoe gevoeliger de gegevens zijn en hoe directer ze aan een persoon kunnen worden gekoppeld, des te strikter de verwerking ervan moet worden gereguleerd, kan daarom in twijfel worden getrokken. Daarnaast is het de vraag of de focus op de identificeerbaarheid van een individu (natuurlijke persoon) en vervolgens de noties van anonimisering en pseudonimisering die daarop zijn gebaseerd, te handhaven zijn in de 21e eeuw.

7. Wanneer is het redelijk om te zeggen dat gegevens niet meer terug kunnen worden gekoppeld aan een persoon en dat de dataset waarvan ze deel uitmaken als anoniem kan worden beschouwd?

Hoewel er vanuit juridisch oogpunt een verschil is tussen niet-persoonsgegevens en persoonsgegevens, valt dit onderscheid vanuit technisch oogpunt uiteen in ten minste drie relevante subcategorieën:

1. de situatie waarin gegevens nooit persoonlijk waren, maar wel zouden kunnen zijn, zoals wanneer klimaatdata worden gebruikt om beslissingen te nemen over de verzekering van individuele boeren;
2. de situatie waarin gegevens persoonlijk waren, maar de identifiers zijn gestript of gegevens zijn geanonimiseerd op een zodanige manier dat de betrokkene niet kan worden geïdentificeerd of identificeerbaar is. Hierbij bestaat het gevaar dat gegevens opnieuw worden gereïdentificeerd of gedeanonimiseerd;
3. de situatie waarin gegevens worden geaggregeerd. Hierbij bestaat zowel het gevaar dat gegevens

kunnen worden gedeaggregeerd, dat de combinatie van twee geaggregeerde datasets persoonsgegevens kunnen opleveren, en dat geaggregeerde gegevens kunnen worden gebruikt om beslissingen te nemen die van invloed zijn op individuele betrokkenen of om hen eruit te pikken, zonder hun identiteit te kennen.

Voor elk van die scenario's zijn er verschillende bedreigingen. Vanuit het technologische domein is duidelijk dat het bijna nooit aannemelijk is dat data niet meer terug te koppelen zijn aan een individu. Er zijn altijd risico's voor de-anonimisering, er zijn altijd mogelijkheden tot gegevenssamenstelling en het kan nooit worden uitgesloten dat gegevens worden gebruikt om niet-geïdentificeerde individuen te onderscheiden of om beslisbomen te ontwikkelen die een impact hebben op groepen en/of individuen. Daardoor is het steeds moeilijker te bevestigen dat data niet meer terug te koppelen zijn aan een individu en dat de dataset waarvan ze deel uitmaken als anoniem kan worden beschouwd.

8. In hoeverre is de toets op indirecte identificeerbaarheid objectiveerbaar?

Er zijn maar weinig aanwijzingen gevonden om de toets op indirecte identificeerbaarheid meer objectiveerbaar te maken. Het is belangrijk om te onderstrepen dat het objectiveerbaar maken van de test niet het doel was van de EU-regelgever. Integendeel, de huidige open, contextuele en fluïde reeks criteria kreeg de voorkeur boven de meer beperkende criteria die werden overwogen en verworpen. Zo bevatte het oorspronkelijke voorstel voor de richtlijn gegevensbescherming niet het begrip anonimiteit, maar veeleer dat van ‘depersonalisatie’, dat werd opgevat als het zodanig wijzigen van informatie dat het niet langer aan een specifiek individu kon worden gekoppeld. In de memorie van toelichting werd bepaald dat ‘een gegeven kan worden beschouwd als gedepersonaliseerd, zelfs als het theoretisch zou kunnen worden gerepersonaliseerd met behulp van onevenredige technische en financiële middelen’. Tegelijkertijd definieerde de toelichting depersonalisatie als "het zodanig wijzigen van persoonsgegevens dat de informatie die ze bevatten niet langer in verband kan worden gebracht met een specifieke persoon of een persoon die kan worden bepaald, behalve tegen de prijs van een buitensporige inspanning.” Overmatige inspanning is nog steeds contextueel, maar minder dan “alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruik”; ook is de drempel duidelijk anders.

Er zijn in dit onderzoek weinig aanwijzingen gevonden om de toetsing van indirecte identificeerbaarheid objectiever te maken dan het schrappen van het begrip 'identificeerbaarheid', dat oorspronkelijk geen deel uitmaakte van de definitie van persoonsgegevens onder de gegevensbeschermingsregimes van vóór 1995, of het beperken van de lijst van factoren die moeten worden opgenomen om te bepalen welke middelen redelijkerwijs moeten worden gebruikt. Misschien is de enige concrete suggestie die werd geïdentificeerd, het stellen van een tijdsbeperking of een horizon voor de evaluatie van de middelen die redelijkerwijs kunnen worden gebruikt. Het is bijna altijd zeer waarschijnlijk dat over 20 jaar gegevens die nu anoniem zijn, kunnen worden gedeanonimiseerd. Onder het huidige wettelijke regime moet, wanneer gegevens zo lang worden bewaard of als ze openbaar worden gemaakt, rekening worden gehouden met dergelijke middelen die redelijkerwijs kunnen worden gebruikt bij het bepalen of het gegevensbeschermingsregime van toepassing is, terwijl het vrijwel onmogelijk is te voorzien hoe het technologische landschap en de beschikbaarheid van data zich de komende 20 jaar zal ontwikkelen.

9. In hoeverre en in welke gevallen kan er sprake zijn van onderregulering wanneer gegevens niet meer door middel van anonimisering aan personen worden gekoppeld en dus niet onder de AVG vallen?
10. In welke mate en in welke gevallen kan er sprake zijn van overregulering wanneer steeds meer gegevens eenvoudig aan individuen kunnen worden gekoppeld door middel van nieuwe technieken (het ongedaan maken van maatregelen van anonimisering en pseudonimisering)?

Het beantwoorden van de vragen 9 en 10 hangt af van wat wordt beschouwd als de reguleringsdoelstelling van het gegevensbeschermingsregime: moet het gegevensbeschermingskader

worden beschouwd vanuit een beschermend perspectief of vanuit het perspectief van het faciliteren van gegevensverwerking binnen een vastgesteld kader, of als een combinatie tussen beide? Moet het beschermingsdoel worden opgevat als het voornamelijk bieden van bescherming aan individuele belangen of (ook) aan groeps- en maatschappelijke belangen? Moet het gegevensbeschermingsregime worden opgevat als het stellen van beperkingen voor gegevensverwerking of als het bieden van een kader voor het gebruik en het delen van gegevens? Is de beschermende grondgedachte het best gediend door beperkingen, of kan er soms meer gegevensverwerking nodig zijn om de belangen van individuen en/of de samenleving zo goed mogelijk te dienen? Is de grondgedachte van het faciliteren van gegevensgebruik het best gediend met een open en contextueel kader of met het stellen van strikte en duidelijke regels waarbinnen gegevensverwerking als legitiem wordt beschouwd? Dit onderzoek heeft niet ten doel om op deze vragen een definitief antwoord te geven; wel is duidelijk dat afhankelijk van de antwoorden op deze vragen verschillende lacunes in de regelgeving en gevaren voor over- en/of onderregulering zullen worden gevonden.

Of er bijvoorbeeld sprake is van onderregulering omdat ‘persoonsgegevens’ alleen gekoppeld zijn aan de identificeerbaarheid van natuurlijke personen en omdat het gegevensbeschermingskader primair verwijst naar de belangen van de betrokkene, hangt af van wat als het kerndoel van het gegevensbeschermingskader wordt gezien. Als wordt aangenomen dat het gegevensbeschermingskader bescherming biedt of zou moeten bieden aan meer algemene, groeps- of maatschappelijke belangen, dan kan er zeker sprake zijn van onderregulering omdat de verwerking van geaggregeerde en anonieme gegevens niet onder het huidige regime valt. Of de trend van rechters en adviesorganen om de reikwijdte van persoonsgegevens en de materiële reikwijdte van het gegevensbeschermingskader uit te breiden tot overregulering leidt, hangt af van de vraag of de nadruk wordt gelegd op de beschermende grondgedachte van het gegevensbeschermingskader, in welk geval er geen sprake zou zijn van overregulering, maar het juist toe valt te juichen dat de reikwijdte steeds wordt uitgebreid, of dat de nadruk wordt gelegd op het faciliterende doel, in welk geval een te grote reikwijdte van het gegevensbeschermingsregime sneller als overregulering kan worden beschouwd.

11. Hoe zullen de huidige en toekomstige technische ontwikkelingen de komende periode van invloed zijn op de AVG en de rechtsbescherming van gegevens in brede zin?

Het is duidelijk dat de technologische ontwikkelingen en de algemene beschikbaarheid van data nu en in de toekomst tot gevolg hebben dat anonimisering steeds moeilijker wordt. De status van data wordt steeds volatieler en wordt steeds minder een kenmerk van data en datasets zelf en meer en meer een effect van de inspanningen van de verwerkingsverantwoordelijke. De juridische categorieën zullen steeds meer fluïde en minder stabiel worden en één database kan juridisch verschillend worden beoordeeld ten aanzien van verschillende partijen die daar toegang toe hebben. Een database die op zichzelf alleen niet-persoonsgegevens bevat kan worden omgezet in persoonsgegevens door deze het volgende moment te combineren met een andere database, kan vervolgens worden gebruikt om gevoelige persoonsgegevens af te leiden, om het moment daarop weer te worden geaggregeerd en geanonimiseerd. Gezien deze trends en gezien de begrippen ‘identificeerbaarheid’ en ‘alle middelen waarvan redelijkerwijs valt te verwachten dat zij worden gebruik’, zullen steeds meer gegevens, zo niet alle, onder het gegevensbeschermingskader vallen.

In dit onderzoek zijn geen verschillende scenario's gevonden voor hoe de technologische wereld en de beschikbaarheid van open data zich in de loop van de tijd zullen ontwikkelen: literatuur, geïnterviewde experts en experts die zijn uitgenodigd voor de workshop die voor dit onderzoek is gehouden, wijzen allemaal in dezelfde richting. Er zijn echter meerdere scenario's gevonden voor hoe het wettelijk regime zou kunnen reageren op de toegenomen beschikbaarheid van open data en de algemene toegankelijkheid van technologie. Uit de suggesties zijn vijf globale strategieën afgeleid: het huidige gegevensbeschermingskader intact laten, focussen op duidelijkere gegevenscategorieën, meer nadruk leggen op contextualiteit, gebruik maken van verschillende gegevenscategorieën en daaraan gekoppelde

reguleringsregimes, of focussen op een volledig contextueel gegevensbeschermingskader.