

aan DG

cc CPO, directeur CSB, hoofd CAD, prico's, CSB-j, directeur CSB

van FG

onderwerp Wat zijn persoonsgegevens?

datum 7 februari 2023

Er bestaan – ook binnen het CBS – veel misverstanden over wat kwalificeert als persoonsgegeven. Daarom heb ik besloten een verduidelijking te geven van het begrip binnen de context van de AVG. Daarbij geef ik ook aan wat de relatie is tussen het wettelijke begrip persoonsgegeven en de binnen het CBS courante begrippen 'direct en indirect identificerende gegevens' met betrekking tot personen.

### 1. Wat zegt de AVG?

Persoonsgegevens zijn volgens artikel 4 sub 1 AVG "alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon („de betrokkene”); als identificeerbaar wordt beschouwd een natuurlijke persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identicator zoals een naam, een identificatienummer, locatiegegevens, een online identicator of van een of meer elementen die kenmerkend zijn voor de fysieke, fysiologische, genetische, psychische, economische, culturele of sociale identiteit van die natuurlijke persoon."

Hiermee geeft de definitie van persoonsgegeven al direct een aantal belangrijke kenmerken:

- Een gegeven kwalificeert als persoonsgegeven, ook als het gaat over een natuurlijke persoon die nog niet geïdentificeerd is. Daarmee zijn ook attributen die potentieel identificerend zijn wel degelijk persoonsgegevens.

*Voorbeeld 1: een voldoende specifiek Nederlands bruto inkomen kan identificerend zijn, omdat de groep voor wie het 't juiste inkomen is, heel klein is. Het is dan een persoonsgegeven dat (vrijwel) direct identificerend is.*

*Voorbeeld 2: een IP-adres is vaak (binnen voorwaarden) identificerend voor een kleine groep mensen, bijvoorbeeld een gezin. Het is dan een persoonsgegeven dat (vrijwel) direct identificerend is.*

- Een persoonsgegeven gaat over een natuurlijke persoon. Een natuurlijke persoon is een levende persoon. De gegevens van overleden personen zijn geen persoonsgegevens in de zin van de AVG. Dat wil echter niet zeggen dat zij niet beschermenswaardig zijn.

*Let op: een attribuut dat direct of indirect identificerend is voor een overleden persoon is geen persoonsgegeven voor die overledene, maar kan dat, bijvoorbeeld door erfelijkheid, wel zijn voor een natuurlijke persoon, en daarmee alsnog gelden als persoonsgegeven, echter niet m.b.t. de overledene.*



- Ook pseudoniemen zijn persoonsgegevens. Een RIN of PRIN kwalificeert in elk geval binnen het CBS als persoonsgegeven. Attributen die zijn gekoppeld aan een RIN of PRIN zijn eveneens persoonsgegevens, omdat ze – voorzover ze zelf niet identificerend zijn – via de RIN of PRIN alsnog als attribuut van een identificeerbare natuurlijke persoon worden gezien.

*Voorbeeld: geslacht is zelfstandig een te generiek attribuut om een persoon te identificeren. Samen met een pseudoniem of bijvoorbeeld NAW-gegevens is het echter wel degelijk een persoonsgegeven.*

- De stapeling van attributen leidt vaak tot identificeerbaarheid. Een verzameling attributen die elk niet-identificerend zijn, kan als verzameling wel degelijk identificerend zijn. In hun context (deze verzameling) is elk van de attributen dan eveneens een persoonsgegeven.

*Voorbeeld: de combinatie van enkele vrij generieke attributen zoals geslacht, hoogst genoten opleiding, inkomensklasse, beroep, woonplaats en eigenwoningbezit zal vaak al naar een zeer beperkt aantal individuen wijzen, of naar een enkel individu.*

## 2. Jurisprudentie

De volgende kenmerken van persoonsgegevens zijn gebaseerd op de jurisprudentie van het Europese Hof van Justitie, de hoogste rechter in de EU:

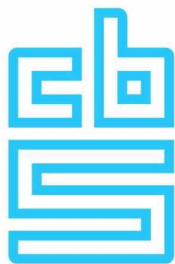
Een gegeven kwalificeert als persoonsgegeven als de partij die over het gegeven beschikt de wettelijke mogelijkheden heeft om de betrokkene te identificeren.<sup>1</sup> Dit bevestigt de contextgebondenheid van de kwalificatie van een gegeven als persoonsgegeven. Wat voor de ene partij een persoonsgegeven is, is dat voor de andere niet. Zo valt te argumenteren dat een kenteken voor een willekeurige burger geen persoonsgegeven is, maar voor een opsporingsambtenaar wel. Dit is overigens discutabel: zo is het kenteken van de auto van de buurman wel een persoonsgegeven, en het kenteken van een auto die betrokken is bij een aanrijding met je eigen auto ook (via de aangifte en verzekeringsclaim).

*Voorbeeld: de echtgenoot van een aan een Nederlandse onderzoeksinstelling werkzame onderzoeker maakt een foto van de kamer waarin de onderzoeker aan het werk is achter haar laptop. Op de foto blijkt een rijtje RINs te staan. Aangenomen dat de onderzoeker zich verder aan alle regels voor RA-gebruik houdt, zijn deze RINs voor de echtgenoot geen persoonsgegevens omdat deze geen middelen heeft om betrokkenen te identificeren.*

De antwoorden van een kandidaat op een examen en de opmerkingen die daarbij door de corrector zijn geplaatst vormen persoonsgegevens in de zin van de AVG. Dit geldt niet voor de examenvragen

---

<sup>1</sup> Arrest van 19 oktober 2016, Breyer (C-582/14, EU:C:2016:779)



zelf.<sup>2</sup> Ik noem dit de olievlekwerking van persoonsgegevens. Het bijhouden van bijv. medische dossiers en personeelsdossiers met observaties van artsen of leidinggevendenden heeft consequenties voor de betrokkene. Hoewel diagnoses en observaties ‘secundair’ zijn (ze zijn opgesteld door een derde over betrokkene, daarmee een geconstrueerd, ‘nieuw’ attribuut), gaan ze wel degelijk over betrokkene, en zijn het dus ook persoonsgegevens. Voor dergelijke gegevens kunnen soms beperkingen worden gesteld aan bijvoorbeeld inzage, maar dat het persoonsgegevens betreft, staat buiten kijf.

*Voorbeeld: het samenstellen van indicatoren op individueel niveau, bijvoorbeeld het berekenen van een SES-score of een vatbaarheid voor bepaalde ziekten, leidt tot nieuwe persoonsgegevens. Ook de toepassing van een algoritme op bijvoorbeeld aggregaten gerelateerd aan postcodegebieden kan leiden tot nieuwe persoonsgegevens. Zo is een postcodegebaseerde kredietscore effectief een persoonsgegeven, net als een kansindicator voor hartafwijkingen op basis van aggregaten die wordt toegepast op individuen.*

### 3. Consequenties

Met betrekking tot *synthetische data* wordt soms geclaimd dat het niet om persoonsgegevens gaat. In bepaalde gevallen is dat terecht, bijvoorbeeld bij synthetische data op basis van aggregaten. Er wordt bij synthetische data gebaseerd op persoonsgegevens echter niet altijd helder onderscheiden tussen het niveau van individuele attributen en records. Hoewel een voor alle attributen gesynthetiseerde dataset geen records zal bevatten die *gezamenlijk* een individu identificeren, kan dat wel degelijk gelden voor de individuele attributen. Indien een attribuut voldoende specifiek en daardoor identificerend is, of wanneer met elders toegankelijke data alsnog identificatie mogelijk is, dan is sprake van een persoonsgegeven.

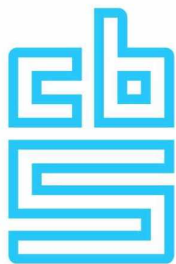
De toepassing van privacy enhancing en privacy preserving technieken leidt vaak tot een hoger beschermingsniveau, maar de gegevens die hier onderwerp van zijn, blijven persoonsgegevens, en dus onderhevig aan het regime van de AVG. Claims van leveranciers en belanghebbenden met betrekking tot het ‘privacybeschermende’ karakter van technieken en technologieën dienen altijd met een flinke korrel zout te worden genomen. In uitzonderlijke gevallen kan de status van persoonsgegevens wel vervallen, maar daarvoor zijn de drempels hoog. Daarnaast wijs ik graag op de noodzaak van outputcontrole. Ook als een product geen persoonsgegevens bevat of lijkt te bevatten, kan er nog steeds het risico van statistische onthulling zijn.<sup>3</sup>

Aggregaten en algoritmen die het product zijn van statistisch onderzoek kunnen, ook zonder dat zij zelf nog persoonsgegevens bevatten, wel degelijk worden ingezet om persoonsgegevens te genereren. Die mogelijkheid lijkt vaak over het hoofd te worden gezien. De onzorgvuldige toepassing van de output van statistisch onderzoek kan ernstig consequenties hebben voor degenen op wie die output wordt toegepast. Dat gold al voor de stigmatiseringsrisico’s van statistieken en voor beleidsbeslissingen die op basis van statistiek worden genomen, maar kan in het tijdperk van

---

<sup>2</sup> Arrest van 20 december 2017, Nowak (C-434/16, EU:C:2017:994)

<sup>3</sup> Wolf, P.-P. de en B. van den Berg (2021), Privacy Preserving Techniques and Statistical Disclosure Control, Discussion paper.



grootschalige (semi)geautomatiseerde oordeels- en besluitvorming tot ernstige consequenties voor betrokkenen leiden.