

aan Hoofddirecteur ITC (vz IV-raad)  
 cc Directeur BSB, hoofd BIA, CPO, CISO, hoofd SSC  
 van FG en plvFG

onderwerp Advies FG inzake toetsing testen met productiedata vanuit AVG  
 datum 26 juni 2025

De FGs ontvangen regelmatig vragen over de mogelijkheden voor het testen met productiedata. Dit advies beoogt richting te geven aan praktische oplossingen die recht doen aan het gegevensbeschermingsrecht. Testen met productiedata is onwenselijk, omdat testomgevingen vaak meer risico lopen op het optreden van een 'datalek' (inbreuk in verband met persoonsgegevens in de zin van art. 33 AVG).

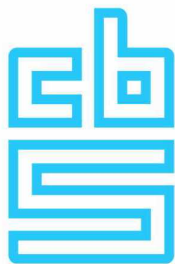
In de courante OTAP-opzet zijn de ontwikkel- en testtrappen toegankelijk voor de ontwikkelaars, en de acceptatie- en productietrappen voor de interne cliënten. Dat weerhoudt de ontwikkelaars van tests op een omgeving die lijkt op de productieomgeving (namelijk de acceptatieomgeving). Voor het testen van eenvoudige software-componenten is dat geen probleem; voor integrale tests is dat een beletsel. De interne cliënt moet dan zorgdragen voor de integrale tests en voldoende informatie terugleveren aan de ontwikkelaars over gevonden softwarefouten ('bugs'). Voor de reproductie van dergelijke fouten zal in de praktijk vaak ook de data nodig zijn die in acceptatie zijn gebruikt.

Ook in het geval wanneer meerdere statistische deelprocessen worden geïntegreerd in een enkel proces, zal op enig moment de werking van het gehele proces moeten worden getest op plausibiliteit van de uitkomst. De vraag is nu onder welke omstandigheden microdata voor dergelijke tests mogen worden gebruikt, en in welke omgeving. Daarnaast is het de vraag of er mitigerende maatregelen genomen kunnen worden om de risico's voor betrokkenen te verkleinen.

*Persoons- en bedrijfsgegevens: pseudonimisering, synthetisering en anonimisering*

Om te beoordelen of een gegeven kwalificeert als microdata (persoonsgegeven of bedrijfsgegeven), is het niet voldoende om enkel de informatie per veld te beschouwen. Neem bijvoorbeeld een record over een persoon dat bestaat uit tien velden. Het eerste daarvan is een identificerend nummer. Ook na onomkeerbare encryptie of verwijdering van dat nummer kwalificeert het record als geheel als persoonsgegeven. Dat komt doordat de negen resterende velden in gezamenlijkheid de persoon al snel uniek identificeerbaar maken.

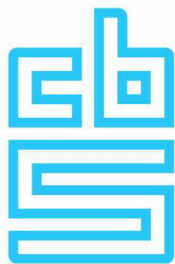
Binnen het CBS is bij de verwerking van microdata in principe nooit sprake van anonimisering (in de gebruikscontext niet tot een persoon herleidbaar), en het is dus onverstandig deze term te gebruiken voor processen die in feite neerkomen op pseudonimisering. Pseudonimisering is nuttig, en het is in principe verplicht vanuit art. 89 lid 1 AVG voor statistische verwerking. Pseudonimisering biedt bovendien een waarborg voor de bescherming van de rechten en vrijheden van betrokkenen. Toch blijven gepseudonimiseerde gegevens in juridische zin kwalificeren als persoons- of bedrijfsgegevens.



Een bijzondere vorm van pseudonimisering of anonimisering is synthetisering. Het gaat hierbij om technieken waarbij met behulp van bestaande data (microdata of geaggregeerde data) een model wordt gegenereerd. Dit model genereert op zijn beurt een synthetische dataset. De nieuwe data behouden sommige kenmerken van de oorspronkelijke dataset, bijvoorbeeld bepaalde patronen of relaties. Synthetische data die voldoende waarde behoudt voor het uitvoeren van statistische analyses zal echter doorgaans gevoelig blijven voor pogingen tot heridentificatie van individuen (in samenhang met andere bronnen), en dus niet kwalificeren als anonieme data.

In de tabel hieronder zijn niet-gepseudonimiseerde, gepseudonimiseerde, gesynthetiseerde en fictieve gegevens gegroepeerd. Deze zijn te onderscheiden van fictieve gegevens en aggregaten. We veronderstellen dat geanonimiseerde microdata niet bestaan. Van belang is elke set testdata van voldoende metadata over de aard van individuele velden en de aard van de data te voorzien (bijv.: is de data 'echt' of gegenereerd?). Zo is in het verleden bij de (onrechtmatige) opslag van microdata in een code repository niet altijd te achterhalen geweest of het om microdata of fictieve gegevens ging.

Microdata	Fictieve gegevens	Aggregaten
<p><b>Niet-gepseudonimiseerde microdata:</b> Persoons- en bedrijfsgegevens met direct identificerende gegevens</p> <p><b>Gepseudonimiseerde microdata:</b> persoons- en bedrijfsgegevens waaruit direct identificerende kenmerken zijn verwijderd, en eventueel een nieuwe, betekenisloze identifier is toegevoegd om koppeling of heridentificatie mogelijk te maken</p> <p><b>Gesynthetiseerde microdata:</b> resultaat van modelmatig genereren van nieuwe data met achtergrondgebruik van microdata; heridentificatie is niet meer per definitie mogelijk, maar er blijft een restrisico</p>	<p>Op microdata gelijkende, verzonden persoons- en bedrijfsgegevens, al dan niet gegenereerd met behulp van technieken voor het produceren van synthetische data. Deze data zijn anoniem.</p>	<p>Geanonimiseerde gegevens: indien aggregaten conform het Handboek statistische beveiliging afdoende statistisch beveiligd zijn, mogen we ervan uitgaan dat deze gelden als anoniem, omdat er afdoende waarborgen zijn dat er uit de aggregaten geen kenmerken van individuele personen of bedrijven afgeleid kunnen worden.</p>



#### *Noodzakelijkheidstoets*

Het is van belang steeds voor ogen te houden dat een minder veilige wijze van testen voldoet aan het noodzakelijkheids criterium. Zodra het mogelijk is om te testen in een beter beveiligde omgeving (bijv. zonder 'poorten' naar buiten via internet en e-mail), met gepseudonimiseerde (waaronder gesynthetiseerde) data of met geanonimiseerde data, dan moet die weg worden gekozen. Voor bijv. unit testing kunnen vaak prima kleine fictieve datasets worden gemaakt die de datavalidatie en -verwerking voldoende kunnen ondersteunen. In zo'n geval is er dus geen reden om productiedata (ook al is die gepseudonimiseerd of geanonimiseerd) te gebruiken.

Er bestaan uiteraard situaties waarin fictieve datasets niet inzetbaar zijn. Dit kan het geval zijn als de plausibiliteit van de uitkomst van een statistisch proces dient te worden getoetst, bijv. tegen een bestaande productiestraat. In zo'n geval kan worden gekozen voor een test in de acceptatie-omgeving, aangenomen dat deze beter beveiligd is dan de O- en T-trappen. Steeds is de overweging: als het veiliger *kan*, dan *moet* het ook veiliger.

#### *Maatregelen ter mitigering van het risico*

De risico's van testen met gegevens zijn het laagste bij fictieve data, aggregaten en synthetische data op basis van aggregaten. Het risico wordt hoger, in oplopende volgorde, bij het gebruik van:

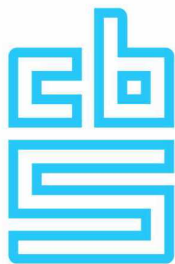
- Gesynthetiseerde data gemaakt op basis van microdata
- Gepseudonimiseerde microdata
- Niet-gepseudonimiseerde microdata

Als er een zwaarwegend belang is om in een testomgeving te testen met productiedata, dan dienen de daaruit voortvloeiende risico's te worden gemitigeerd. Deels is dat mogelijk met maatregelen die ook voor acceptatie- en productie-omgevingen gelden (de opsomming is niet uitputtend):

- De beperking van het aantal personen met toegang tot de omgeving
- Beperking van de rechten van deze personen op de omgeving (bijv. alleen leesrechten)
- Beperking van bepaalde mogelijkheden, zoals gebruik van e-mail en bestandsoverdracht
- Pseudonimisering van persoons- en bedrijfsgegevens
- Logging en monitoring op het gebruik van de omgeving

Sommige van deze maatregelen zijn uit hun aard lastig te combineren met ontwikkel- en testomgeving, omdat ontwikkelaars meer armslag nodig hebben om hun taken te verrichten. Er kan ook aan maatregelen worden gedacht die in een productieomgeving niet zouden werken, maar voor bepaalde tests wel geschikt zijn, zoals:

- Beperking van de omvang van de set productiedata 'in de lengte', bijv. door middel van een steekproef
- Beperking van de omvang van de set productiedata tot de velden die daadwerkelijk nodig zijn voor de test; vullen van de velden die niet nodig zijn met fictieve data



### *Rechtmatigheid*

Testen is een verwerking van persoonsgegevens (voor zover met persoonsgegevens in al dan niet gepseudonimiseerde of gesynthetiseerde vorm wordt getest). Voor elke verwerking van persoonsgegevens is een grondslag vereist. De grondslagen zijn opgenomen in art. 6 lid 1 van de AVG. Voor de rechtmatigheid is het in principe niet van belang in welke omgeving wordt getest indien de beginselen van art. 5 AVG in acht worden genomen en technische en organisatorische beveiligingsmaatregelen worden getroffen. Voor testen bij het CBS kan worden gekozen uit de volgende grondslagen:

- De verwerking is noodzakelijk voor de uitoefening van een taak van algemeen belang (het doortesten van een statistisch (primaire) proces is noodzakelijk om de betrouwbaarheid van dat proces te borgen)
- De verwerking is noodzakelijk voor de behartiging van de gerechtvaardigde belangen van verwerkingsverantwoordelijke (het doortesten van een secundaire, ondersteunend proces is noodzakelijk om de betrouwbaarheid van dat proces te borgen)

Belangrijk is dat de gegevens die voor tests worden gebruikt niet langer worden bewaard dan noodzakelijk is voor het uitvoeren van de tests. Een set microdata mag dus alleen worden bewaard in een testomgeving zolang er regelmatig gebruikt wordt gemaakt van die set om tests uit te voeren. Zodra die set daartoe niet meer voldoet, bijvoorbeeld omdat er een nieuwe testset is opgesteld of omdat de testwerkzaamheden zijn afgerond, moet de oude worden verwijderd.

De verwerking van bijzondere categorieën van persoonsgegevens is verboden in art. 9 lid 1 AVG. Daarop is een uitzondering gemaakt voor de verwerking voor o.m. statistische doeleinden. Deze uitzondering is te vinden in art. 9 lid 2 sub j AVG. De meeste testwerkzaamheden voor secundaire processen zullen vallen onder de uitzondering in het kader van arbeidsrechtelijke verplichtingen ex art. 9 lid 2 sub b AVG. In al deze gevallen dienen de testwerkzaamheden gezien te worden als verlengde verwerking t.o.v. het oorspronkelijke doel. De noodzakelijkheid dient dan ook onbetwist te zijn.

### *Conclusie*

Vanuit het gegevensbeschermingsrecht is er geen absoluut, doorslaggevend bezwaar tegen het gebruik van productiedata in de ontwikkel- en testtrappen van een OTAP-systeem (een combinatie van een ontwikkel-, test-, acceptatie- en productieomgeving). Echter, de risico's die het werken met productiedata oproept dienen te allen tijde serieus genomen te worden en te worden gemitigeerd met technische en organisatorische maatregelen. Bij de afweging van welke omgeving met welke data moet worden gebruikt, dient altijd de noodzaak van afwijking van de volgende twee testscenario's te worden aangetoond: testen op acceptatie met productiedata, of testen op de testomgeving met fictieve data. Dit advies doet niet af aan eventueel doorslaggevende bezwaren tegen gebruik van de testtrap voor testen met productiedata vanuit de hoek van de informatiebeveiliging.