

aan Voorzitter stuurgroep Veilig Data Delen
 cc CSB-j, CPO, privacycoördinatoren, hoofd CAD
 van FG en plvFG

onderwerp Annex voorbeelden bij Advies over teruglevering microdata
 datum 14 december 2023

Deze annex bij het Advies over teruglevering microdata van 28 september 2023 bevat per uitzondering op de hoofdregel een fictief voorbeeld. Het is uitgebracht op verzoek van de Stuurgroep Veilig Data Delen. Voor de leesbaarheid is de tekst van het oorspronkelijke advies, waarop deze annex een aanvulling vormt, integraal overgenomen.

Er komen bij ons regelmatig vragen over casuïstiek rondom de teruglevering van microdata. Het betreft uiteenlopende zaken, en er zijn al enkele adviezen van de FG in omloop die de desbetreffende problematiek adresseren. Bij het CBS bestaat behoefte aan een zo veel mogelijk omvattend advies waarin de verschillende vormen van teruglevering worden geadresseerd en wordt aangegeven wat vanuit de AVG en de CBS-wet wel en niet mogelijk is. Dit advies ziet *niet* op doorlevering, waarbij microdata naar een derde partij worden gestuurd. Het gaat in dit advies dus uitsluitend om de situatie waarin het CBS op welke manier dan ook microdata over of van partij X of een selectie daarvan levert aan partij X.

Dit advies is gebaseerd op eerdere FG-adviezen.¹ Volledigheidshalve zij opgemerkt dat de hoofdregel voor microdata ('gegevens' in de wet) volgt uit art. 37 CBS-wet lid 2 en 3, namelijk dat microdata niet worden verstrekt aan andere partijen dan die belast zijn met de wettelijke taak van het CBS. Deze hoofdregel geldt ook ten aanzien van bronhouders en respondenten: er mag in beginsel geen microdata worden teruggeleverd. De uitzonderingen op deze hoofdregel worden in dit advies besproken.

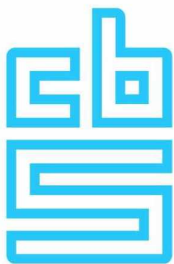
Vormen en functies van teruglevering

Wij onderscheiden zeven vormen van teruglevering met bijbehorende functies:

1) Controlevragen

Controlevragen dienen om er (met redelijke mate) zeker van te zijn dat een respondent de juiste persoon is die in een steekproef is geselecteerd. Voor de hand liggende vragen betreffen naam en geboortedatum. Het is echter niet nodig om dergelijke kenmerken al 'weg te geven'. Voldoende is in principe om er naar te vragen en door middel van een 'blinde' controle in een vragenlijst de juistheid te testen. Er zijn daarbij ook manieren om om te gaan met eventuele spelfouten,

¹ Kwaliteitscontrole externe bronbestanden en de AVG/CBS-wet, d.d. 20 februari 2020, Basisprincipes bij verwerkingen van microgegevens bij werk voor derden, d.d. 16 januari 2018, Advies FG inzake incidentele teruglevering van responsdata d.d. 8 februari 2022.



diacrieten en onbekenden (niet altijd is een geboortedatum even duidelijk). Vanwege het risico dat combinaties van persoonsgegevens bij de verkeerde respondent belanden is het van belang deze zo min mogelijk te vermelden in aanschrijfbrieven en enquêtes.

Voorbeeld: voor een bepaalde schriftelijke enquête is het belangrijk dat deze door de aangeschrevene wordt ingevuld, en niet door een huisgenoot. Het is dan toelaatbaar om zonder een identificerend kenmerk te noemen een controlevraag te stellen waarvan het juiste antwoord de kans vergroot dat de antwoorden door de aangeschrevene zijn gegeven.

2) Ingevulde vragenlijsten

Er zijn drie doelen waarvoor vragenlijsten vooringevuld worden aangeboden aan respondenten, alle gericht op een hogere respons op enquêtes en een lagere lastendruk voor de respondent:

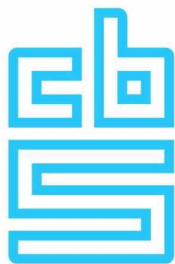
- Om met een al eerder geopende en deels ingevulde vragenlijst op een later tijdstip verder te kunnen gaan.
- Om hulp te bieden bij het invullen van een nieuwe vragenlijst door een vorige versie aan te bieden.
- Om vragen te kunnen stellen over specifieke rechtspersonen of bezittingen.

Indien er voldoende waarborgen zijn ingebouwd om te voorkomen dat anderen dan de beoogde respondent toegang krijgen tot zelf verstrekte microdata, kunnen deze doelen legitiem zijn. Er is immers in beginsel geen sprake van openbaarmaking in de zin van art. 37 lid 3 CBS-wet, en de drie doelen in lijn zijn met de wettelijke taak van het CBS. Er dient echter terughoudend met deze teruglevering omgesprongen te worden, vooral bij de verstrekking van bijzondere persoonsgegevens. Steeds is het nodig om de noodzaak van de verstrekking (dient de ondersteuning bij het invullen van een vragenlijst een statistisch doel) af te wegen bij het risico op verstrekking van gegevens aan de verkeerde persoon (een incident in de zin van art. 33 AVG, in de volksmond een 'datalek').

Voorbeeld: voor een enquête over gezondheid binnen een gezin is het belangrijk om de antwoorden op specifieke vragen over elk van de kinderen te koppelen aan het juiste kind. Het kan dan opportuun zijn om een duidelijk identificerend kenmerk te noemen voor elk kind om het vragencluster aan te koppelen. Dat identificerende kenmerk mag geen bijzonder persoonsgegeven of BSN zijn. Verder mag het op geen enkele manier de veiligheid van eventueel niet-thuiswonende kinderen in gevaar brengen.

3) Communicatie over ingevulde vragenlijsten

Er zijn met name bij bedrijfsenquêtes situaties waarin er onduidelijkheden bestaan over geleverde data en antwoorden. Ook komt het voor dat de uploadportal bepaalde databestanden niet accepteert vanwege validatiefouten. Dat noopt tot communicatie tussen CBS en respondent. Op dit moment wordt daarvoor vaak gebruik gemaakt van e-mail of mondelinge communicatie via de telefoon. Dit is een onwenselijke situatie omdat e-mail en telefoon kunnen leiden tot het verstrekken van microdata aan de verkeerde partij. Verder is gewone e-mail niet in alle gevallen een voldoende veilig communicatiemiddel.



Wij erkennen de noodzaak om communicatie tussen CBS en respondent mogelijk te maken. Het CBS zal zich moeten richten op een methode die hiervoor voldoende veilig is en waarbij voldoende waarborgen zijn getroffen om de identiteit van de respondent vast te stellen. Verder dient de verstrekking van microdata beperkt te blijven tot het minimale dat noodzakelijk is om de vervolgvragen vanuit het CBS te beantwoorden.

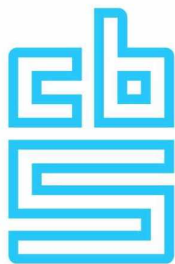
Voorbeeld: voor een enquête wordt gebruik gemaakt van bestanden uit boekhoudprogramma's. Een bedrijf uploadt een gevraagd exportbestand via de uploadportal van het CBS maar krijgt een foutmelding. De contactpersoon belt het CBS met de vraag wat de foutmelding betekent. De CBS-medewerker moet nu een weg vinden om de fout te identificeren en op te lossen. Daartoe moeten – op een zo veilig mogelijke wijze – gegevens worden uitgewisseld.

4) Verbeteren datakwaliteit en vaststellen betekenis van data ten behoeve van de statistiek

Vroeg in het statistisch proces is het onvermijdelijk dat er tussen bronhouder of dataleverancier en CBS communicatie is over de aard en inhoud van geleverde data. Daarbij gaat het bijvoorbeeld om ontbrekende data, om kennelijk foutieve of ongeschoonde data. Het is daarbij noodzakelijk te kunnen benoemen om welke rijen ('records') en welke velden het probleem draait. Dat mag alleen op basis van fouten die in de data zelf, zonder koppeling van andere CBS-data, worden geconstateerd. Een deskundige bronhouder/dataleverancier had deze zelf ook kunnen constateren. Hoewel in een ideale wereld die kwaliteitscontroles bij de bronhouder gebeuren, zal het in de praktijk onvermijdelijk zijn dat het CBS hier een rol in speelt. Deze feedback op record-niveau dient zo vroeg mogelijk vervangen te worden door reguliere kwaliteitsrapportages met geaggregeerde uitkomsten.

Reguliere kwaliteitsrapportages dienen onder andere als verantwoording voor geconstateerde kwaliteitsissues in registers en geleverde databestanden richting bronhouder/dataleverancier. Het is niet de bedoeling dat in dergelijke kwaliteitsrapportages op recordniveau wordt gerapporteerd. Ook algemene uitspraken over de dataset mogen niet tot een vaste set records te herleiden zijn (een query met een vaste, voorspelbare uitkomst uit de dataset is vergelijkbaar met rapporteren op recordniveau). Bij het opstellen van reguliere kwaliteitsrapportages kan wel van andere dan geleverde microdata gebruik gemaakt worden, maar mag de uitkomst ervan alleen als aggregaat worden teruggeleverd. Er mag dus bij reguliere kwaliteitsrapportages geen sprake zijn van de teruglevering van microdata of herleidbaarheid tot een concreet begrensde set microdata.

*Voorbeeld: bij een bronhouder van een **nieuw toe te voegen** registratie zijn vragen gerezen over het formaat waarin gegevens moeten worden aangeleverd. Zij willen graag een beperkt aantal records met echte data aanleveren via een uploadportaal van het CBS. Het CBS kan aan de hand daarvan vaststellen of de geleverde gegevens voldoen aan de kwaliteitseisen van het CBS. Om de fouten in de levering duidelijk te communiceren, mag het CBS voor deze nieuwe registratie via een beveiligd kanaal een versie met correcties van microdata terugleveren.*



5) Verbetering datakwaliteit van registers van bronhouders

Hoewel kwaliteitsverbetering van registers gehouden bij bronhouders leidt tot een verbetering van de statistiek, is dit op zichzelf geen reden waarom het CBS mag bijdragen aan een dergelijke kwaliteitsverbetering. In de praktijk kan het CBS vaak prima een schaduwregistratie aanhouden waarin het de kwaliteit van de geleverde data zelf verbetert. Dat is mogelijk door de statistische kennis die bij het CBS berust, en door de koppeling met andere bronnen. De taak van het CBS is het verrichten van statistisch onderzoek, het bevorderen van een statistische informatievoorziening van overheidswege, en de bevordering van de nauwkeurigheid en volledigheid van de van overheidswege openbaar te maken statistieken.

Dat laatste facet (art. 3 lid 2 onder b CBS-wet) laat enige ruimte voor bevordering van de kwaliteit van overheidsregisters, voorzover daarbij geen sprake is van strijd met art. 37 lid 3 (onthullingsverbod) en art. 3 lid 1 CBS-wet (openbaarmakingsverplichting). Dit betekent dat werkzaamheden kunnen worden verricht die facilitair zijn aan de verbetering van de inhoud van een register. Daarbij dient het CBS zich altijd rekenschap te geven van het feit dat de doelen waarmee een register wordt bijgehouden niet noodzakelijk in lijn zijn met de statistische doelen van het CBS.

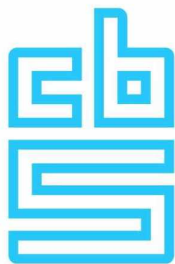
Het CBS mag geen rol spelen in het faciliteren van rechtsgevolgen voor burgers en bedrijven via op kwaliteitscriteria gebaseerde registerwijzigingen. Ook hier geldt dat het belangrijk is om de uitkomsten of verbeteringen niet op recordniveau terug te koppelen naar de bronhouder. Het is van belang te beseffen dat statistische variabelen zoals de SBI door derde partijen gaandeweg gebruikt kunnen gaan worden voor het vaststellen van rechtsgevolgen. De herkomst van die variabele als statistische classificatie doet niet af aan het terugleveringsverbod.

Voorbeeld: de constatering van het CBS met behulp van achtergrondkenmerken dat veel wellness-gelegenheden zich met een suboptimale of zelfs incorrecte SBI-code identificeren kan niet op recordniveau aan de KVK worden gecommuniceerd. Een algemene constatering dat in een bepaalde sector veel suboptimale SBI-codes voorkomen, kan echter wel worden gegeven.

6) Teruglevering met toegevoegde waarde: selectie zonder gebruik van achtergrondkenmerken

In dit geval is sprake van teruglevering van microdata zonder nieuwe gegevens, met een selectie die de leverende partij zelf ook zelf had kunnen bepalen, maar waarvoor die partij zelf de expertise niet in huis heeft. Het maken van de selectie moet ook een statistisch doel hebben. Omdat de bronhouder alle informatie in huis heeft om deze selectie zelf te maken en er een statistisch doel is, is in dit geval teruglevering toegestaan. Let op: ook in dit geval mag er 'achter de schermen' bij het CBS geen sprake zijn van verrijking aan de hand van andere data dan die al berust bij de bronhouder.

Voorbeeld: uit een set van databestanden die bij elke provincie aanwezig is, kan het CBS afleiden welke landbouwgrond het meest geschikt is voor het vestigen van zonneparken. De provincies zouden dat zelf ook kunnen maar hebben de expertise niet. Het CBS kan voor alle provincies deze geschikte percelen terugleveren en tegelijkertijd een aggregaat publiceren over de hoeveelheid geschikte grond (en daarmee de verwachte opwekcapaciteit) per provincie.



7) Teruglevering met toegevoegde waarde: selectie met gebruik van achtergrondkenmerken

Er zijn omstandigheden waarin het CBS een populatiebestand verzorgt met als doel het laten verrijken van dit populatiebestand met extra gegevens door een externe bronhouder. Om te voorkomen dat onnodige informatie van deze bronhouder opgevraagd wordt, kan het CBS juist in deze situatie toegevoegde waarde leveren door een doelpopulatie terug te leveren aan de bronhouder. Deze doelpopulatie kan worden bepaald op basis van een koppeling met andere bronnen met achtergrondkenmerken.

Dergelijke gevallen zijn een uitzondering op de regel dat een teruglevering niet verrijkt mag worden. De verrijking mag echter niet op recordniveau herkenbaar zijn voor de bronhouder. Daartoe wordt de populatie uitgebreid door middel van oversampling, in principe minimaal met een factor drie, van individuen die het te onderzoeken kenmerk of de te onderzoeken kenmerken niet hebben maar wel onderdeel uitmaken van het register van de bronhouder. Dat betekent dat uit de teruggeleverde populatiebestand op individueel niveau niet op te maken is of een individu het te onderzoeken kenmerk heeft.

De levering van een populatiebestand moet ten dienste staan van het statistisch oogmerk. In sommige gevallen kan de levering bijdragen aan dataminimalisatie ex art. 5 lid 1 onder c doordat een bronhouder niet langer genoodzaakt is een integraal register aan het CBS te leveren en toch een onderzoek uit kan (laten) voeren met gebruikmaking van achtergrondkenmerken die bij het CBS bekend zijn. Het meeleveren van een recordvolgnummer dat latere koppeling van kenmerken mogelijk maakt moet niet worden beschouwd als levering van microdata. Elke mogelijkheid tot gebruik van dat recordvolgnummer voor een andere koppeling dient daartoe wel te worden voorkomen. Het mag dus niet in een andere context herbruikbaar zijn.

Voorbeeld: een academisch ziekenhuis heeft een beperkte populatie personen met een zeldzame aandoening onder behandeling. Het ziekenhuis is geïnteresseerd in comorbiditeiten in het DSM-5-spectrum maar beschikt niet over deze informatie. Bij het CBS is wel bekend of individuen al dan niet in behandeling zijn voor een aandoening in het DSM-5-spectrum bij een derde partij, maar niet voor welke aandoening. Het ziekenhuis levert een populatiebestand aan het CBS. Het CBS matcht de populatie op behandeling door de derde partij en levert het gematchte populatiebestand, drie maal oversampled met individuen zonder deze behandelrelatie, terug aan het academisch ziekenhuis. Het academisch ziekenhuis stuurt het populatiebestand aan de derde partij. De derde partij levert geanonimiseerde DSM-5-diagnosegegevens voor het populatiebestand aan het CBS voor opname in de RA-omgeving. Hierna kan het ziekenhuis onderzoek doen naar de comorbiditeiten zonder dat wordt prijsgegeven welke patiënten het betreft.



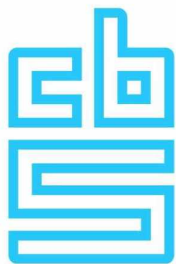
Samenvatting vormen functies van teruglevering

| Nr | Vorm van teruglevering | Levering aan | Procesfase | Primaire functie |
|----|--|--------------|---|---|
| 1 | Controle vragen | Respondent | Primaire waarneming | Verifiëren primaire waarneming |
| 2 | Inge vulde vragenlijsten | Respondent | Primaire waarneming | Faciliteren primaire waarneming |
| 3 | Communicatie over ingevulde vragenlijsten | Respondent | Analyse en bewerking | Kwaliteitsverhoging resultaten primaire waarneming |
| 4 | Verbetering data kwaliteit | Bronhouder | Bij aanvang levering | Kwaliteitsverhoging secundaire waarneming |
| 5 | Verbetering en kwaliteit registers | Bronhouder | Periodiek, per levering of per herziening | Kwaliteitsverhoging secundaire waarneming |
| 6 | Teruglevering met toegevoegde waarde: selectie zonder gebruik van achtergrondkenmerken | Bronhouder | Analyse | Verhogen kwaliteit en eenduidigheid statistiek |
| 7 | Teruglevering met toegevoegde waarde: selectie met gebruik van achtergrondkenmerken | Bronhouder | Populatiebestand | Dataminimalisatie binnen het statistisch proces en wetenschappelijk onderzoek |

Samenvatting randvoorwaarden teruglevering

De hoofdregel is dat teruglevering van microdata niet is toegestaan (art. 37 CBS-wet). Bij teruglevering in uitzonderingsgevallen gelden de volgende randvoorwaarden:

- De teruglevering moet ten dienst staan van het statistisch oogmerk (zie overweging 162 van de AVG).
- Een teruglevering door het CBS mag niet leiden tot (potentiële) rechtsgevolgen voor individuele betrokkenen of bedrijven (uitgezonderd de verstrekking van microdata noodzakelijk voor de handhaving op grond van par. 4 van hst 5 CBS-wet; dit valt onder het statistisch oogmerk).
- Er mogen nooit meer microdata worden teruggeleverd dan oorspronkelijk door de respondent, bronhouder of dataleverancier is leverd. Een door het CBS gegenereerd projectspecifiek recordvolnummer (geen BSN, geen RIN) geldt in dit kader niet als microdata.
- Er mag geen selectie van microdata worden teruggeleverd die gebaseerd is op enige vorm van koppeling of verrijking met andere microdata van het CBS, tenzij (1) de microdata de respondent zelf betreffen en de teruglevering ten dienste staat van de primaire waarneming of (2) sprake is van levering van een oversampled populatiebestand met uitsluitend records die al in het register van de bronhouder voorkwamen.



- Er moeten adequate technische en organisatorische maatregelen worden getroffen om ervoor te zorgen dat de teruglevering plaatsvindt naar dezelfde respondent, bronhouder of dataleverancier, en dat deze teruglevering op veilige wijze geschiedt.
- Als een bronhouder of dataleverancier in staat is of zou moeten zijn om zelf de kwaliteit van de data te verbeteren, dan is er voor het CBS in dezen geen rol.
- Bij elke teruglevering moet het beginsel van dataminimalisatie worden toegepast: er mogen niet meer microdata worden geleverd dan noodzakelijk voor het vastgestelde doel.