

aan DG

cc Hoofddirecteur SER, hoofddirecteur BVS, directeur BSB, prico SER, CMO, CPO
van FG en plvFG

onderwerp Advies over publicatiebestand PIAAC

datum 2 september 2024

De DG heeft ons gevraagd advies uit te brengen over de mogelijkheid van het samenstellen van een publicatiebestand van het PIAAC-onderzoek in Nederland. Vanuit methodologische hoek is reeds geadviseerd over de mogelijke scenario's. Onze toetsing is op basis van respectievelijk de CBS-wet en de AVG.

PIAAC is een internationaal onderzoek dat in Nederland wordt uitgevoerd op basis van een steekproef van het CBS van 10.000 personen. Het onderzoek brengt de kennis en vaardigheden van de Nederlandse bevolking in kaart. Het Kohnstamm Instituut van de Universiteit van Amsterdam en onderzoeksbureau Kantar voeren het onderzoek uit. De opdrachtgevers zijn de ministeries van OCenW en EZK.

De informatiebrochure over het onderzoek voor respondenten vermeldt met betrekking tot de behandeling van de gegevens van betrokkenen: "PIAAC voldoet aan de AVG-wet. Alle informatie wordt strikt vertrouwelijk behandeld. Na het interview worden uw naam- en adresgegevens vernietigd. De resultaten van het PIAAC-onderzoek die worden gepubliceerd, zijn nooit te herleiden naar een persoon."

Voordat we overgaan tot de toetsing van de vier geschetste scenario's, zetten we nog even de belangrijkste wettelijke condities op een rij. De CBS-wet legt in art. 37 vast dat in het kader van zijn taken ontvangen gegevens uitsluitend voor statistische doeleinden mogen worden gebruikt, dat deze gegevens niet aan anderen mogen worden verstrekt dan degenen die met de taak van het CBS belast zijn, en dat in openbaar gemaakte gegevens geen individuele huishoudens of personen mogen worden herkend.

Op het verstrektingsverbod bestaat alleen een uitzondering vanuit art. 41 CBS-wet. Deze uitzondering geldt voor bepaalde onderzoeksinstellingen en –enclaves, en onder de conditie dat passende maatregelen zijn genomen om herkenning van individuele personen en huishoudens te voorkomen. Deze conditie geldt (uiteraard) ook na het uitvoeren van een onderzoek. Zo worden tabellen die het resultaat zijn van onderzoek op de Remote Access-voorziening van het CBS elk op onthulling gecontroleerd.

We toetsen nu achtereenvolgens de scenario's die zijn opgesteld in het document 'Scenario's voor een PIAAC-publicatiebestand'.



1. Het CBS volgt de eigen regels voor een publicatiebestand om de privacy van de respondenten te kunnen garanderen. Dit is het CBS-scenario dat voldoet aan de CBS-beveiligingsregels. In het bestand zullen maximaal 15 variabelen worden opgenomen die ingedikt worden of waarbij scores worden onderdrukt als sprake is van zeldzame (combinaties van) variabelen.

Dit scenario volgt het CBS-beleid, inclusief de standaarden voor statistische beveiliging die het CBS zichzelf (en degenen die gebruik maken van RA) oplegt. Dit scenario geeft daarmee uitvoering aan de verplichtingen uit art. 37 CBS-wet.

2. Als scenario 1, maar daarbovenop worden ook de 2300 overige (inhoudelijke) variabelen opgenomen. Gevoelige informatie van respondenten is dan niet adequaat beschermd (zelfs met slechts 15 identificerende variabelen zullen er unieke personen zijn en waarbij de overige inhoudelijke variabelen ook nog identificerende informatie bevatten), maar onderzoekers kunnen deze 2300 variabelen analyseren in samenhang met de maximaal 15 identificerende variabelen zoals die door het CBS in het bestand worden opgenomen.

In dit scenario wordt duidelijk niet voldaan aan de verplichtingen uit art.37 CBS-wet. Ook gaat dit scenario voorbij aan de regels uit de AVG. Immers, statistische verwerkingen zijn gebonden aan de eisen uit art. 89 lid 1 AVG, te weten passende waarborgen om minimale verwerking van persoonsgegevens te garanderen. Het zal – en niet slechts in theoretische zin – mogelijk zijn om individuen te identificeren met behulp van kennis van betrokkenen en/of toegang tot aanvullende (openbare) bronnen. Daarmee zijn alle identificerende en inhoudelijke variabelen in het publicatiebestand alsnog effectief persoonsgegevens.

3. Als scenario 2, maar daarbovenop worden niet slechts de maximaal 15 identificerende variabelen in de microdata opgenomen, maar 225 (de helft) van alle 450 identificerende variabelen zoals het Kohnstamm Instituut die wil leveren aan de OESO. Het wordt dan gemakkelijker respondenten te identificeren en gevoelige informatie te onthullen, maar voor alle respondenten kunnen dan de 2300 overige (inhoudelijke) variabelen worden gedesaggregeerd naar categorieën van 225 identificerende variabelen.

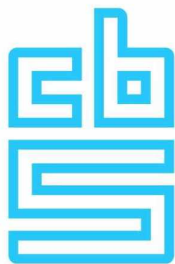
In dit scenario spelen dezelfde problemen als in scenario 2, alleen is het aantal persoonsgegevens kleiner, maar nog steeds aanzienlijk.

4. Als scenario 3, maar in plaats van een selectie van 225 identificerende variabelen worden alle 450 identificerende variabelen zoals het Kohnstamm Instituut heeft voorgesteld (met een zeer beperkt aantal indikkingen en zonder onderdrukkingen) opgenomen. We hebben dan het scenario van het Kohnstamm Instituut waarbij de CBS-regels in zeer ernstige mate worden overschreven en vrijwel alle gevoelige gegevens in het bestand aan unieke personen zijn toe te schrijven.

Het behoeft geen toelichting dat dit scenario ook een zeer groot aantal persoonsgegevens onthult.

De notitie geeft nog twee alternatieven:

- Het maken van een DANS-bestand. Gezien het gebrek aan controle op gebruik en verspreiding van DANS-bestanden achten wij dit geen verantwoorde route. In de huidige



praktijk is het simpelweg te eenvoudig een eenmaal gedownload bestand (bewust of ongewild) verder te verspreiden.

- Het gebruik van de Remote Access-omgeving voor het bestand. Dit beschouwen wij als een acceptabele optie die voldoet aan de eisen die in de AVG en de CBS-wet gesteld worden.

Steeds vaker wordt door bepaalde partijen de status ‘persoonsgegevens’ voor de enqueteresultaten in een gepseudonimiseerd bestand in twijfel getrokken. Wij gaan in deze beweging niet mee.

Volgens vaste jurisprudentie van het Europese Hof¹ is voor elke partij die beschikt over de (wettige) middelen om een individu te identificeren aan de hand van een gegeven of set van gegevens dat gegeven een persoonsgegeven. Er wordt hierbij wel verwezen naar een uitspraak van het Gerecht.² De feitelijke situatie is hier echter een totaal andere dan in het geval van openbaarmaking van een publicatiebestand met een groot aantal identificerende en inhoudelijke variabelen, die zelfstandige identificatie van personen mogelijk maken zonder toegang tot een (ont)pseudonimiseringsleutel.

Conclusie

De openbaarmaking van een PIAAC-publicatiebestand conform scenario 2, 3 of 4 is een ernstige inbreuk op zowel de CBS-wet als op de AVG. We zouden een dergelijke openbaarmaking niet alleen als onrechtmatig kwalificeren, maar ook zien als inbreuk (‘datalek’) in de zin van art. 33 en 34 AVG, meldplichtig richting de Autoriteit Persoonsgegevens en richting betrokkenen. Ook zou deze openbaarmaking rechtstreeks in strijd zijn met de informatie die in het onderzoek aan betrokkenen is verstrekt. Van de twee overige scenario’s zien wij de beschikbaarstelling via Remote Access als enige acceptabele route.

Voor wat betreft de aansprakelijkheid van partijen voor deze inbreuk: het CBS zou op basis van de vermoedelijk gehanteerde figuur van gezamenlijke verwerkingsverantwoordelijkheid mogelijk (mede)aansprakelijk zijn. Los van financiële en reputatierisico’s, willen wij benadrukken dat de mogelijke schade voor betrokkenen door bekendmaking van hun persoonsgegevens aanzienlijk is, **en adviseren wij alles in het werk te stellen de openbaarmaking te voorkomen.**

¹ Arrest van het HvJEU 19 oktober 2016, Breyer (C-582/14, EU:C:2016:779): een dynamisch IP-adres is een persoonsgegeven voor een aanbieder van online mediadiensten. Arrest van het HvJEU van 20 december 2017, Nowak (C-434/16, EU:C:2017:994): niet alle gegevens aan de hand waarvan een persoon kan worden geïdentificeerd hoeven bij een en dezelfde persoon te berusten om gekwalificeerd te worden als persoonsgegevens.

² Arrest van het Gerecht EU 26 april 2023, GAR-EDPS (T-557/20, ECLI:EU:T:2023:219): gegevens die voor de ene partij gepseudonimiseerd zijn, kunnen voor een andere partij gelden als geanonimiseerd. Deze zgn. ‘relatieve leer’ is niet nieuw, wel omstreden, in het licht van bijv. juridische mogelijkheden die later in de tijd ontstaan om alsnog personen te kunnen identificeren, en toegenomen rekenkracht en nieuwe gegevensbronnen die hetzelfde effect hebben. Er is beroep aangetekend deze uitspraak, er volgt dus nog een uitspraak 5.1.2.e van Justitie EU.