

aan Voorzitter stuurgroep Veilig Data Delen

cc CPO, prico SER, prico DRI

van FG en plvFG

onderwerp Advies FG inzake DANS-bestanden

datum 1 juli 2024

De voorzitter van de stuurgroep Veilig Data Delen heeft ons gevraagd advies uit te brengen over de omgang met DANS-bestanden. Wij geven in het voorliggend advies antwoord op de volgende vragen:

- Kwalificeren DANS-bestanden als microdata?
- Voldoet de omgang met DANS-bestanden aan de vereisten zoals neergelegd in art. 41 lid 1, te weten het treffen van passende maatregelen door het CBS om herkenning van afzonderlijke personen e.d. te voorkomen?
- Voldoet de omgang met DANS-bestanden aan de vereisten zoals neergelegd in art. 42 CBS-wet, te weten het treffen van voldoende maatregelen door verzoeker om te voorkomen dat de verzameling van gegevens voor andere doelen dan statistisch of wetenschappelijk onderzoek wordt gebruikt?

We hebben kennisgenomen van de notitie Levering DANS-bestanden.¹ We hebben ons daarnaast laten voorlichten over de wijze waarop DANS-bestanden statistisch worden beveiligd, en over de inhoud ervan. De meeste DANS-bestanden zijn enquêtebestanden, en bevatten dus de resultaten van antwoorden op vragen aan respondenten uit een steekproef. Om een DANS-bestand beschikbaar te kunnen stellen voor onderzoek buiten het CBS moeten de volgende beveiligingsregels gehanteerd worden om de kans op onthulling zo minimaal mogelijk te houden:

- Het bestand mag geen directe identificatoren bevatten (bijv. een BSN of telefoonnummer).
- In het bestand mogen geen regels voorkomen met combinaties van drie variabelen die minder dan 100 keer in een doelpopulatie voorkomen (bijv. een vrouw uit Zierikzee met een knieprothese).

Als niet aan deze laatste regel wordt voldaan, worden voor het bestand met name de volgende twee mechanismen gebruikt om het bestand in te dikken:

- Hercodering. Voor bepaalde variabelen worden nieuwe (bredere) categorieën geïntroduceerd zodat er meer eenheden onder een categorie vallen. Hercodering vindt bijv. plaats door een plaatsnaam te vervangen door een provincienaam, of een specifiek beroep door een beroepsklasse of sector.
- Lokale onderdrukking. Onderdrukking vindt bijvoorbeeld door in plaats van een gegeven antwoord gebruik te maken van de eigenschap 'geen antwoord gegeven'. Een (te)

¹ 5.1.2.e en 5.1.2.e Levering DANS-bestanden, notitie voor Stuurgroep Veilig Data Delen d.d. 30 april 2024.



specifieke waarde van een variabele verdwijnt hiermee, en daarmee wordt eveneens de groep in de doelpopulatie groter.

De beveiliging van DANS-bestanden beoogt het voorkomen van spontane zichtherkenning. Dat is de situatie waarin een onderzoeker bij het bekijken van de microdata daarin een haar of hem bekend individu zou kunnen herkennen. De beveiliging van DANS-bestanden is echter niet voldoende om misbruik te voorkomen. Misbruik zou bijvoorbeeld kunnen plaatsvinden door het profileren van individuen in een combinatie van een DANS-bestand met een ander bestand met deels overlappende profielen. Te denken valt hierbij aan een profielendatabase van een social media provider die wordt gekoppeld met het DANS-bestand.

Volgens vaste jurisprudentie (Breyer²) kwalificeert een gegeven als een persoonsgegeven indien een partij beschikt over (extra) wettige middelen waarmee hij de betrokken persoon mede aan de hand van dat gegeven kan identificeren. Het is in een onderzoekscontext vrij waarschijnlijk dat een onderzoeker beschikt over meerdere bestanden die in combinatie met elkaar extra informatie kunnen opleveren, waaronder de identificeerbaarheid van individuen.

De kans op identificeerbaarheid wordt groter door (a) de omvang van het DANS-bestand (het bestand betreft de complete respons op de steekproef), (b) de vrij ruime mogelijkheden om data onder de doelstelling van wetenschappelijk of statistisch onderzoek te combineren en (b) de langdurige beschikbaarheid van het bestand in combinatie met het voortschrijden van technologische mogelijkheden.

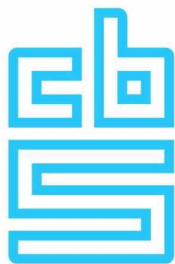
DANS-bestanden zijn ook bekend als microdata onder contract. Zij worden verstrekt onder specifieke aanvullende voorwaarden die bepalen wie de bestanden mogen gebruiken en voor welk doel. De vorm waarin zij worden verstrekt, namelijk als fysiek (downloadbaar) bestand, maakt het echter voor het CBS onmogelijk om zicht te houden op de mate waarin aan de contractuele condities wordt voldaan.

Daarmee ontstaan risico's van onrechtmatig gebruik, niet-vernietiging, doorverstrekking aan onbevoegden en verlies. Dat deze risico's niet denkbeeldig zijn, merken we uit de – overigens spaarzame – meldingen van oneigenlijk doorverstrekking van microdatabestanden. We moeten er echter vanuit gaan dat we daarin slechts het topje van de ijsberg zien omdat niet alle doorverstrekingen worden gemeld.

Contractuele regeling en feitelijke gang van zaken toegang DANS-bestanden

Wij hebben informatie ingewonnen bij CBS account management, microdata services en het bureau van DANS om na te gaan wat de praktische werkwijze is rondom de verkrijging van toegang tot beveiligde DANS-bestanden:

² HvJ EU 19 oktober 2016, ECLI:EU:C:2016:779, Breyer t. Bondsrepubliek Duitsland.



- Bij de aanvraag van een CBS-DANS bestand gaat een gebruiker akkoord met de licentievoorwaarden van DANS.
- Er wordt na de aanvraag automatisch een mail verstuurd naar het CBS met de gegevens van de aanvrager.
- Het CBS toetst of de betreffende onderzoeker werkt bij een instelling met een instellingsmachtiging van het CBS. Gaat het om een student, dan moet de student een geheimhoudingsverklaring ondertekenen.
- Als aan alle voorwaarden is gedaan, wordt het bestand ter beschikking gesteld aan de aanvrager.

De licentievoorwaarden van DANS³ leggen de verantwoordelijkheid om te voldoen aan wetgeving bij de onderzoeker (art. 1). Dit wordt herhaald in art. 6 specifiek voor de omgang met persoonsgegevens conform bijv. de AVG, met de aanvulling dat de rechthebbende nadere condities kan stellen op dit gebied. Verder bepaalt de licentie dat voor reproductie of openbaarmaking van de dataset of een substantieel deel daarvan toestemming van de rechthebbende nodig is (art. 3). Dit artikel zegt dus niets over de reproductie of openbaarmaking van een of enkele records uit een dataset.

De geheimhoudingsverklaring van het CBS ziet onder meer op de geheimhouding van elk gegeven ten aanzien van een individu, onderneming of instelling dat bij de uitvoering van werkzaamheden met de dataset bekend wordt. Ook de begeleider moet tekenen, en wel voor de verantwoordelijkheid van de naleving van de condities uit de geheimhoudingsverklaring door de student.

Deze procedure is in lijn met de Samenwerkingsovereenkomst tussen de KNAW en het CBS.⁴ Art. 3 van deze overeenkomst stelt dat DANS de CBS-bestanden aan alle instellingen kan verstrekken die een geldige machtiging voor toegang tot microdata van het CBS hebben, en aan masterstudenten die een aanvraag hebben die is getekend door een bevoegd begeleider van een gemachtigde instelling.

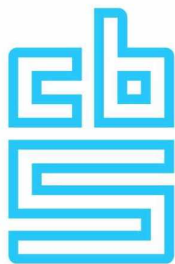
Kwalificeren DANS-bestanden als microdata?

Deze vraag moet positief beantwoord worden. De rijen in een DANS-bestand representeren elk een individu. Ook een voor spontane zichtherkenning beveiligd bestand kwalificeert nog steeds als microdata.

Voldoet de omgang met DANS-bestanden aan de vereisten zoals neergelegd in art. 41 lid 1, te weten het treffen van passende maatregelen door het CBS om herkenning van afzonderlijke personen e.d. te voorkomen?

³ DANS-licentie van kracht op 9 januari 2020.

⁴ Samenwerkingsovereenkomst tussen De Koninklijke Nederlandse Akademie van Wetenschappen ten behoeve van Data Archiving and Networked Services en het Centraal Bureau voor de Statistiek d.d. 3 januari 2017.



De toegepaste regels voor statistische beveiliging zijn voldoende om spontane zichtherkenning te voorkomen. Dit laat echter onverlet dat bij combinatie met andere bestanden wel degelijk de mogelijkheid bestaat dat individuen geïdentificeerd kunnen worden. Naast het indikken van de bestanden en de eisen zoals gesteld voor het ontvangen en behouden van een instellingsmachtiging, gelden voor DANS-bestanden slechts contractuele voorwaarden voor gebruik. Het CBS heeft op de werkzaamheid daarvan geen of onvoldoende zicht.

Voldoet de omgang met DANS-bestanden aan de vereisten zoals neergelegd in art. 42 CBS-wet, te weten het treffen van voldoende maatregelen door verzoeker om te voorkomen dat de verzameling van gegevens voor andere doelen dan statistisch of wetenschappelijk onderzoek wordt gebruikt?
De DANS-licentieovereenkomst wijst niet op de specifieke risico's van onthulling van identificerende gegevens over personen, huishoudens, bedrijven en instellingen. Wij constateren dat er bij gebruikers van microdatabestanden sowieso veel onduidelijkheid is over of deze kwalificeren als te beschermen persoonsgegevens of –bedrijfsgegevens onder de AVG en de CBS-wet. Het CBS mag er dus niet vanuit gaan dat gebruikers van DANS-bestanden uit de licentietekst begrijpen wat het te beschermen belang is.

Volgens onze informatie komt de bepaling over gegevens van individuen uit de geheimhoudingsovereenkomst een reguliere onderzoeker zelfs niet onder ogen (tenzij deze begeleider is van een student-aanvrager). Daarnaast is onduidelijk hoe de begeleider van een student verantwoordelijk kan worden gehouden voor het handelen van een student. De begeleider staat in een arbeidsverhouding tot de instelling die beschikt over een instellingsmachtiging, en het is dan ook hoogstens de instelling zelf die hiertoe gehouden kan worden, al kan de begeleider hier – bijvoorbeeld met voorlichting en onderwijs – wel een rol in spelen.

Wij zijn van mening dat de contractuele afspraken een onvoldoende maatregel vormen om te voorkomen dat de verzameling van gegevens voor andere doelen dan statistiek en wetenschap wordt gebruikt aan de zijde van de ontvanger. Er zijn weinig handhavingsmogelijkheden. Hoewel er bepaalde sancties zijn gedefinieerd in de DANS-licentieovereenkomst is onduidelijk hoe misbruik aan het licht kan komen. Meer specifiek: wij hebben geen bewijzen gezien dat de ontvangende instanties in control zijn van deze afspraken, en dat zij dit aantonen aan het CBS.

Advies

Contractuele bepalingen vormen zonder aanvullende maatregelen een zwakke constructie voor beperking van onthullingsrisico van microdatabestanden zoals DANS-bestanden. Wij adviseren dan ook terughoudend te zijn met de voortzetting van DANS-bestandsleveringen en deze waar mogelijk te vervangen door een beter controleerbaar alternatief. In elk geval lijkt het noodzakelijk de ontvangers van DANS-bestanden te informeren dat zij kwalificeren als persoons- en/of bedrijfsgegevens (ofwel microdata) en beter toe te lichten wat er wel en niet met de bestanden gedaan mag worden.