



# Advies FG m.b.t. beleid onthullingsgevaar statistische informatie in de nabije toekomst

5.1.2.e

projectnummer    Klik en typ projectnummer  
                          Klik en typ sector  
                          7 juni 2021

samenvatting    Klik en typ de samenvatting  
trefwoorden    Klik en typ de trefwoorden

## 1. Aanleiding en context

Eind mei 2021 is er een kennismakingsgesprek geweest tussen de Autoriteit persoonsgegevens (AP) en de DG van het CBS. Namens het CBS waren daarbij ook aanwezig Erik Bruinsma en 5.1.2.e

5.1.2.e

Tijdens het gesprek kwam de zorg van de AP naar voren hoe het CBS in de (nabije) toekomst denkt om te gaan met open data, waaronder CBS-data, en steeds meer geavanceerde tools als algoritmes, Machine Learning, Artificial Intelligence welke samen een potentieel onthullingsgevaar vormen voor de gepubliceerde statistische informatie van het CBS. Voor de kortere termijn beschikt het CBS over veel ervaring, awareness en tools als T-Argus om statistische informatie te beveiligen. T-Argus is een tool om een (CBS-)tabel of een geselecteerde set van (CBS-)tabellen statistisch te beveiligen tegen een bepaald aanvalsscenario. Dat scenario houdt geen rekening met breed beschikbare openbare informatie. Ook houdt T-Argus geen rekening met het scenario waar bijvoorbeeld internet robots ingezet worden die heel breed zoeken naar beschikbare (hulp)informatie. Dat laatste scenario is tegenwoordig zeker niet meer denkbeeldig.

Dit kan op termijn de wettelijk uitgangspositie van het CBS (geen onthulling van informatie over herkenbare individuele entiteiten) in gevaar brengen.

## 2. Advies FG

Het risico op onthulling van individuen op basis het gepubliceerde statistische informatie wordt bepaald door een samenspel tussen security en statistische geheimhouding. Security bepaalt hoe de statistische informatie beschikbaar gesteld wordt. Laat het CBS bijvoorbeeld toe dat internetrobots bij de cijfers kunnen of kan StatLine bijvoorbeeld afgeschermd worden voor internet robots? Statistische geheimhouding bepaalt welke statistische informatie beschikbaar gesteld wordt zodat de gebruiker zinvolle informatie krijgt en het individu tegen onthulling beschermd wordt. Overigens speelt dit niet alleen bij StatLine tabellen, maar bij alle CBS publicaties inclusief visualisaties, modellen, ASD publicaties, etc..

Het advies van de FG is om beleid te formuleren op basis van een risicoanalyse hoe de beveiliging tegen komende bedreigingen vorm te geven. Het beleid heeft duidelijk een breder karakter dan alleen van statistische geheimhouding. Ook security speelt hier zoals aangegeven een rol. Nog breder speelt ook over welke entiteiten en concepten er gepubliceerd wordt. Denk hierbij aan een toenemend onthullingsgevaar bij deels overlappende classificaties en concepten en open data over gerelateerde entiteiten (scholen en leerlingen bijvoorbeeld). Standaardisatie van classificaties en variabelen bij ASD en RA helpt.

Verder zijn er vast nog wel meer aspecten te benomen welke het risico positief of negatief beïnvloeden. Ik heb zeker niet de bedoeling gehad een volledig overzicht te geven.