



Advies FG inzake remote acces projecten op lager regionaal niveau

5.1.2.e

FG van het CBS

projectnummer Klik en typ projectnummer
 Klik en typ sector
 18 januari 2018

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

1. Inleiding

Bij urban data centers, rural data centers, provinciale data centers, city deals en eventueel andere laag regionale onderzoekscentra speelt mee dat de betreffende lokale onderzoekers veel kennis hebben van hun omgeving. Dat is enerzijds mooi, want dan zijn de projecten direct relevant voor het beleid. Maar het brengt anderzijds ook een aanvullend risico met zich mee op zichtherkenning bij het zien van de microdata gedurende het onderzoek. Bij zichtherkenning wordt het object van onderzoek, de persoon of het bedrijf, door de onderzoeker herkend op basis van de eigen kennis ondanks dat de bestanden ontdaan zijn van (direct) identificerende kernmerken.

Het CBS moet zich bij laag regionale onderzoekscentra bewust zijn van de verhoogde kans op zichtherkenning en de extra risico's op oneigenlijk gebruik ingeval van zichtherkenning en daarop mogelijk passende aanvullende maatregelen nemen.

2. Aandachtspunten

De CBS wet (artikel 41) zegt het volgende:

- 1. In afwijking van artikel 37 kan de directeur-generaal op verzoek, ten behoeve van statistisch of wetenschappelijk onderzoek, een verzameling van gegevens met betrekking tot het gebruik waarvan passende maatregelen zijn genomen om herkenning van afzonderlijke personen, huishoudens, ondernemingen of instellingen te voorkomen, verstrekken aan een dienst, organisatie of instelling als bedoeld in het tweede lid, dan wel daartoe toegang verlenen.*

2. Een verzameling van gegevens als bedoeld in het eerste lid kan worden verstrekt, dan wel daartoe kan toegang worden verleend aan:
- a. een universiteit in de zin van de Wet op het hoger onderwijs en wetenschappelijk onderzoek;
 - b. een bij wet ingestelde organisatie of instelling voor wetenschappelijk onderzoek;
 - c. bij of krachtens de wet ingestelde planbureaus;
 - d. de communautaire en nationale instanties voor de statistiek van de lidstaten van de Europese Unie;
 - e. onderzoeksafdelingen van ministeries en andere diensten, organisaties en instellingen, voor zover daartoe instemming van de CCS is verkregen.

De huidige remote access omgeving kent een aantal maatregelen om **(zicht-)herkenning** te voorkomen. Net zo als bij verwerking binnen het CBS voor de reguliere statistiekproductie worden direct identificerende variabelen verwijderd: de bestanden worden verrint of bij bedrijven anderszins aangepast. Daarnaast wordt de uitput gecontroleerd op onthulling.

De tweede beperking in de CBS-wet zegt ons dat het doel van onderzoek **statistiek en wetenschap** moet zijn. Zichtherkenning bij het uitvoeren van onderzoek op microdata helemaal uitsluiten is met de huidige werkwijze niet mogelijk. Vandaar dat naast een geheimhoudingsverklaring en een awareness toets wordt geëist dat de onderzoeker deel uit maakt van een onderzoeksafdeling en van daaruit ook aangestuurd wordt. De onderzoekers worden door het CBS hierop gecontroleerd en krijgen op basis hiervan een volledige onderzoekslicentie of een meekijklicentie. Een tweede belangrijke pijler in het kader van deze nota is naast het verkleinen van de kans op herkenning, het beperken van het risico op oneigenlijk gebruik ingeval van herkenning.

De kans op herkenning en de risico's bij herkenning worden tot op heden mede aanvaardbaar geacht omdat onderzoekers bij bijvoorbeeld een universiteit of ministerie vaak geen detailkennis hebben van de objecten en daarnaast verder weg zitten bij uitvoering van beleid, handhaving of opsporing in het geval van zichtherkenning. Een dergelijke onderzoeker zal en kan in geval van zichtherkenning veelal niets doen met deze kennis.

Bij laag regionale onderzoekscentra kan dit essentieel anders zijn. Deze lokale onderzoekers kunnen beschikken over detailkennis van de objecten in onderzoek en zitten soms ook vrij dicht bij uitvoering, handhaving en opsporing. Het CBS ziet bijvoorbeeld dat sommige onderzoekers tijdelijk vanuit de lokale organisatie bij een **onderzoeksafdeling** geplaatst worden. Dit maakt zowel de kans op zichtherkenning alsmede het risico op verkeerd gebruik bij zichtherkenning groter.

3. Advies FG

3.1 Algemeen

Het CBS moet zich bewust zijn van de extra risico's op zichtherkenning bij laag regionale onderzoekscentra en daarop mogelijk passende aanvullende maatregelen nemen op basis van een risico analyse:

De kennis of de veronderstelde kennis van de onderzoeker bepaalt welke passende maatregelen genomen moeten worden om herkenning te voorkomen.

3.2 Risico bij zichtherkenning

Naast het vermijden van zichtherkenning moet het CBS zich ook bewust zijn van het risico op **onbedoeld gebruik** in geval van zichtherkenning.

Mogelijke maatregelen¹ om dit risico te verkleinen:

- 1 vermijdt onderzoekers welke (heel kort geleden) in uitvoerende onderdelen werkzaam waren (zijn) bij de opdrachtgever.
- 2 Beperk het aantal meekijklicenties tot één.

3.3 Kans op zichtherkenning

Mogelijke maatregelen om de kans op zichtherkenning te verkleinen zijn:

- 1 Remote execution: door microdata niet daadwerkelijk aan te bieden maar alleen via scripts benaderbaar te maken verkleint men de kans op zichtherkenning;
- 2 Pas datameer- en datavisualisatietechnieken toe om het zicht op microdata te beperken;
- 3 Bepaal welke informatie echt nodig is gezien de onderzoeksopdracht (dataminimalisatie):
 - a. lever niet hele SSB-componenten maar alleen selecties van variabelen en objecten.
 - b. Rond variabelen maximaal af of deel deze in in klassen
- 4 Bied in bepaalde gevallen alleen maatwerk als optie aan.

¹ De maatregelen in dit hoofdstuk zijn indicatief en alleen ter verduidelijking om een idee te geven waar aan gedacht kan worden.