# Waarom bestuurders bij databases vaak onterecht het woord “anonimiseren” gebruiken

**In een recent interview in het dagblad Het Parool op 16 augustus 2019 figureerde Jeroen Muller. Hij is de bestuursvoorzitter van de grootste GGZ-instelling van Amsterdam. Het interview vond plaats naar aanleiding van de vernietiging door Akwa GGZ van de ROM-database die afkomstig was van de opgeheven Stichting Benchmark GGZ. Het ging daarbij om gepseudonimiseerde zorgdata uit de geestelijke gezondheidszorg(GGZ). Jeroen Muller sprak in het interview over het “dubbel anonimiseren” van de data.** Het gebruik van de term “anonimiseren” van data in plaats van “pseudonimiseren” is niet zo maar een verspreking. In de loop der tijd is het op gaan vallen hoe voorstanders, vaak afkomstig uit de bestuurslaag van de zorg, de pil van het op centraal(landelijk) niveau data verzamelen proberen te vergulden. Dat doen ze door te spreken over “anoniem”, “bijna anoniem” of “vrijwel anoniem” als het om gepseudonimiseerde data gaat. Ik zal in deze bijdrage ingaan op de vraag waarom ze dit doen.

## Pseudonimiseren

Dit is een procedure waarmee identificerende gegevens met een bepaald algoritme worden vervangen door versleutelde gegevens (het pseudoniem). Het algoritme kan voor een persoon altijd hetzelfde pseudoniem berekenen, waardoor informatie over de persoon, ook uit verschillende bronnen, kan worden gecombineerd. Daarin onderscheidt pseudonimiseren zich van

anonimiseren, waarbij het koppelen op persoon van informatie uit verschillende bronnen niet mogelijk is. Pseudonimiseren is een techniek van informatiebeveiliging, meer specifiek: een 'privacy enhancing technique'.(Deze alinea is afkomstig uit Wikipedia: lemma pseudonimiseren)

## Ultrakorte uitleg

PSEUDONIMISEREN: persoon nog herleidbaar, bestand en sleutel gescheiden bewaren
ANONIMISEREN: persoon niet meer herleidbaar, sleutel weggegooid

Men vervangt  bij pseudonimisatie van zorgdata het "wie" deel(direct identificerende data) door een versleuteld pseudoniem, terwijl het "wat"-deel onversleuteld blijft. Helaas zitten daar maar al te vaak toch identificerende kenmerken bij. Je hoeft daarbij maar te denken aan zeldzame ziekten of familiekenmerken.

## Kanttekening bij anonimiseren

Bij bovenstaande uitleg over wat anonimiseren is, moet  echter door het voortschrijden van techniek en programmatuur thans gezegd worden dat ondanks anonimiseren het zeer wel mogelijk is geanonimiseerde data tot individuen te herleiden. Een voorbeeld is een publicatie rond de recente jaarwisseling in de Journal of the American Medical Association. Op 23 juli 2019 stond in The Guardian ook een artikel onder de kop 'Anonymised' data can never be totally anonymous, says study'

## Nederland

Matthijs R. Koot besteedde er in 2012 in zijn proefschrift uitgebreid aandacht aan. Hij vroeg daarin zich af hoe anoniem geanonimiseerde gegevens zijn. Hij toonde aan dat  67% van zijn onderzoekspopulatie uniek identificeerbaar binnen Nederland was op basis van de vier cijfers van de postcode in

combinatie met geboortedatum (gegevens die niet zelden afleidbaar zijn uit profielen op sociale media).

## Waarom?

Met de gedachte dat het grote publiek het verschil tussen pseudonimiseren en anonimiseren alleen een academische kwestie zal vinden probeert men tegelijkertijd het publiek gerust te stellen door de term “anoniem” of “bijna anoniem” te gebruiken. Men gaat er dan vanuit dat die term gelijk staat met het volledig geheim blijven van de identiteit van individuen, terwijl zulks niet het geval is. “Anoniem” lijkt ook de connotatie “vertrouwd” te hebben.

## Niet overdreven

Uit het bovenstaande moge blijken dat het geenszins overdreven is zeer nauwkeurig te letten over wat voorstanders van grote (zorg)data-verzamelingen in de mond nemen. Het kritisch blijven over dit soort zaken blijft een vereiste om niet in de valkuil van mooi-praat en juich-taal te vallen.

W.J. Jongejan, 20 augustus 2019

Afbeelding van Arek Socha via Pixabay