

Aanleverspecificaties bestand pseudonimisatie CBS

Inleiding

Dit document betreft de aanleverspecificaties voor het aanbieden van persoonsgegevens aan de PVM CBS voor pseudonimisatie.

De Privacy Verzend Module (PVM) leest de te pseudonimiseren gegevens uit een bestand in CSV formaat met de volgende kenmerken:

- Het scheidingsteken is de puntkomma (;)
- Het quote-teken is het dubbele aanhalingsteken (")
- De tekenset van het bestand is ISO-8859-1
- De bestandsnaam dient te eindigen op de extensie .csv

Bestandsnaam

De bestandsnaam is vrij te kiezen, zodanig dat betekenis aan de zending omtrent inhoud, afzender, datum en gebruik zijn toe te kennen. ZorgTTP raadt aan een conventie te hanteren (zonder dat de PVM deze afdwingt of zich hiertoe beperkt): `cbs_test` of `productie_afzender_datum_volgnummer.csv`, bijvoorbeeld:

Naam	Betekenis
<code>cbs_prod_\$afzender\$_20180801_1.csv</code>	Productie aanlevering van \$afzender\$ op 1 augustus 2018, 1 ^e bericht
<code>cbs_test_\$afzender\$_20180201_2.csv</code>	Test bericht van \$afzender\$ op 1 februari 2018, 2 ^e poging

Opbouw bestand

Voor de verwerking van bestanden maakt de PVM gebruik van kolomlabels. De volgende kolomlabels worden als input gebruikt voor pseudonimisatie: **BSN**, **GEBDAT**, **GESLACHT** en **PC6**. Alle andere variabelen worden ongewijzigd aangeboden aan het CBS.

Kolom	Criteria
BSN	een geldig burger service nummer met een lengte van 8 of 9 karakters, alléén cijfers.
GEBDAT	een datum in 'EEJJMMDD' formaat van 8 karakters ná 1 januari 1850 en niet in de toekomst, 27 april 2011 wordt dan 20110427.
GESLACHT	1 karakter waarbij 1, M of m voor mannelijk en 2, V of v voor vrouwelijk staat, 0 voor onbekend en 9 voor niet gespecificeerd
PC6	6 karakters voor een Nederlandse postcode van 4 cijfers en 2 letters A-Z

Afhankelijk van de aanwezigheid van één of meerdere kolomlabels voor pseudonimisatie kunnen één of meerdere van de volgende pseudoniemen worden aangemaakt:

Type pseudoniem	Kenmerk	Brongegeven(s) pseudoniem
BSN	OS1B	BSN
CBS	OS1C	Postcode (NNNNAA), geboortedatum, geslacht
NIVEL	OS1T	Postcode (NNNN), geboortedatum, geslacht

De informatie uit het inputbestand in andere dan de genoemde, voor pseudonimisatie relevante kolom wordt versleuteld door de keten getransporteerd en *onveranderd* bij het CBS afgeleverd.

Voorbeeld input pseudonimisatie CBS:

VOLGNR;GESLACHT;GEBDAT;PC6

1;2;19750301;3448KX

2;1;19671020;2215PR

3;2;19620624;2321HG

4;1;19550914;2223AA

5;2;19360404;2242SC

6;1;19531027;2211MV

7;2;19770617;2316CV

Waarbij de PVM een bestand alleen verwerkt wanneer de eerste regel kolomnamen bevat en de kolommen **BSN**, **GEBDAT**, **GESLACHT** en **PC6** hierin voorkomen.