Factsheet pseudonimisatie ZorgTTP

Stichting ZorgTTP te Houten, maart 2017

Inleiding

In deze factsheet wordt beknopt de door ZorgTTP ontwikkelde systematiek voor het pseudonimiseren van persoonsgegevens beschreven.

Pseudonimisatie

Pseudonimisatie is een maatregel die kan worden ingezet ter bescherming van persoonsgegevens. Pseudonimisatie kan omkeerbaar en onomkeerbaar worden opgezet. In de ZorgTTP dienstverlening wordt onder 'pseudonimiseren' het onomkeerbaar omzetten van een persoonsgegeven naar een niet tot de oorspronkelijke persoon terug te herleiden unieke code verstaan. De omzetting verloopt in een aantal stappen waarbij het cruciaal is dat één van deze stappen bij een zogenaamde Trusted Third Party (TTP) wordt uitgevoerd. De bij de TTP uitgevoerde stap is geheim voor zowel de aanbieder van de gegevens als de ontvangende partij in de pseudonimisatieketen. Op deze wijze kan de relatie tussen pseudoniem en persoonsgegeven worden verbroken en is het niet langer mogelijk om via het aangemaakte pseudoniem terug te gaan naar de direct identificerende gegevens behorende bij de natuurlijke persoon waarop het pseudoniem betrekking heeft.

Voor het pseudonimiseren van bestanden die persoonsgegevens bevatten heeft ZorgTTP een pseudonimisatieplatform ontwikkeld. Dit omvat naast een aantal software modules ook technische en organisatorische voorzieningen.

Pseudonimisatieproces

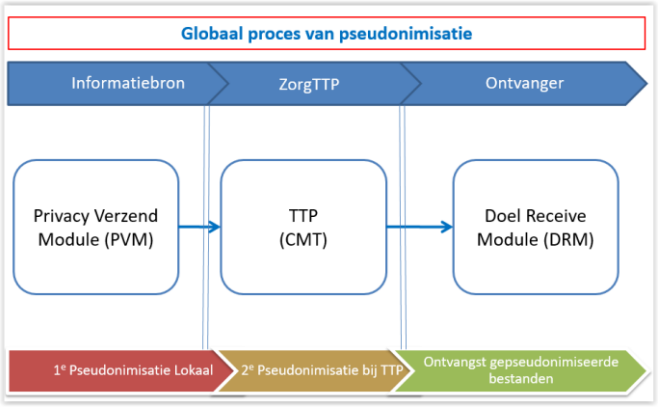
Het pseudonimisatieproces bestaat in het kort uit de volgende stappen:

1. De informatiebron heeft een bestand gegenereerd dat voldoet aan de vooraf gedefinieerde en overeengekomen berichtspecificaties;
2. Het bestand wordt verwerkt met de door ZorgTTP aan de informatiebron beschikbaar gestelde verzendsoftware;
3. Het aangeboden bestand wordt lokaal voor de eerste maal bewerkt met de door ZorgTTP beschikbaar gestelde software;
4. Na de eerste bewerking van het bestand door de informatiebron is een scheiding aangebracht tussen persoonsgegevens en overige inhoudelijke data. Op de persoonsgegevens wordt vervolgens een eerste omzetting uitgevoerd. Hierna volgt transport via een beveiligde internetverbinding naar ZorgTTP.
5. ZorgTTP voert met behulp van centrale pseudonimisatie software een tweede bewerking uit op de ontvangen gegevens waarbij een voor de ontvanger specifiek wachtwoord wordt gebruikt. Na deze bewerking is sprake van definitieve pseudoniemen in de het bestand;
6. Na verwerking wordt het gepseudonimiseerde bestand vrijgegeven en kan het worden opgehaald door de ontvangende partij met een daartoe beschikbaar gestelde ontvangstmodule.

Pseudonimisatieketen

Iedere pseudonimisatieketen bestaat uit drie op elkaar afgestemde componenten:

1. Privacy- en Verzend Module (PVM) wordt gebruikt door de informatiebron;
2. Centrale Module TTP (CMT) wordt gebruikt door ZorgTTP;
3. Doel- en Receive Module (DRM) wordt gebruikt door de ontvangende partij (het doel).



A. Werking Privacy en Verzend Module (PVM)

Deze module wordt gebruikt door de aanbiedende partij, dat is de informatiebron. De module kent een aantal functies. Allereerst wordt een aantal controles uitgevoerd op het aangeboden bestand. Daarna worden de persoonsgegevens omgezet in zogenaamde pre-pseudoniemen. Vervolgens wordt een scheiding aangebracht tussen de pseudoniemen (het sleuteldeel) en de bijbehorende data (het datadeel). Beide delen worden vervolgens beveiligd met behulp van encryptie op zodanige wijze dat het sleuteldeel enkel kan worden geopend door ZorgTTP en het datadeel enkel kan worden geopend door de ontvangende partij, het doel.

Bewerkingen in de PVM

De PVM voert drie typen handelingen uit op het aangeboden bestand:

1. De PVM controleert gegevens die na pseudonimisatie worden verwijderd;
2. De PVM zet aangeboden persoonsgegevens om in een zogenaamd pre-pseudoniem;
3. De PVM voert een aggregatie slag uit op gegevens die niet mogen worden doorgegeven.

Controle op aangeboden persoonsgegevens

Op de aangeboden persoonsgegevens worden logische controles uitgevoerd zoals:

'een datum moet voldoen aan het voorgeschreven formaat (ddmmeejj)'

Pre-pseudonimisatie

De Autoriteit Persoonsgegevens (AP) eist dat een eerste versleuteling bij de *informatiebron* (de partij die beschikt over de te verzenden persoonsgegevens) wordt uitgevoerd. Deze eerste versleuteling wordt ook wel pre-pseudonimisatie genoemd. Een voorbeeld van een pre-pseudoniem is de tekenreeks:

OS1B0039iaf4etutr0su85qv9gfsipex

In het voorbeeld vormen de eerste vier tekens (OS1B) de zogenaamde handtekening van het pseudoniem. Aan deze handtekening kan herkend worden dat het gaat om een *'Onbewerkte Sleutelwaarde' (OS)* van het *1e niveau (1)* voor het type pseudoniem *'B'.* Daarbij slaat het 1e niveau op de eerste bewerking bij de informatiebron en de 'B' op het gebruikte persoonsgegeven; het burgerservicenummer (BSN). Het feitelijke pseudoniem wordt gevormd door de reeks van 28 tekens volgend op de handtekening.

Aggregatie

Voorbeelden van aggregatie op de aangeboden gegevens zijn:

a. De postcode wordt omgezet van 6-karakters (NNNNAA) naar 4-cijferig (NNNN);
b. De geboortedatum (ddmmeejj) wordt bewerkt naar geboortejaar en geboortemaand.

De uitkomst van bovenstaand proces, in combinatie met de hieronder beschreven vervolgstappen, wordt verder in dit document in beeld gebracht onder de kop: 'voorbeeld werking pseudonimisatie'.

B. Centrale Module TTP (CMT) - Centrale pseudonimisatiesoftware:

De Centrale Module TTP ontvangt het in de PVM versleutelde bestand. Dit bestand bestaat uit twee onderdelen: een datadeel en een sleuteldeel. Het sleuteldeel bevat de pre-pseudoniemen, deze worden door de centrale applicatie omgezet in de definitieve pseudoniemen. De centrale applicatie heeft geen toegang tot het datadeel. Alleen in de ontvangstapplicatie kunnen deze gegevens zinvol verder verwerkt worden. Na verwerking verstuurt CMT het bericht naar de ontvangende partij.

C. Doel- en Retour Module (DRM) - Lokale pseudonimisatiesoftware:

De ontvangstmodule wordt gebruikt door de ontvangende partij. De DRM ontvangt van de centrale TTP applicatie de berichten. Na ontvangst worden beide delen samengevoegd. De gegevens zijn dan niet meer te herleiden tot de oorspronkelijk aangeboden persoonsgegevens.

Het definitieve pseudoniem voor het eerder genoemde voorbeeld wordt:

DS2B008rtd2wkt2rnm7wcdj5hyasbv8u

Aan de handtekening kan nu worden herkend dat het een pseudoniem bestemd voor DBC-Informatiesysteem (DIS) betreft (DS) dat 2 maal bewerkt is en gebaseerd is op het BSN. Dit voorbeeld is gebaseerd op pseudonimisatie ten behoeve van DBC-Informatiesysteem.

Voorbeeld werking pseudonimisatie
In onderstaande figuur wordt het in dit document beschreven proces in beeld gebracht aan de hand van de oorspronkelijk aangeboden data (input) en de aan het einde van het proces resterende data (output).