Factsheet pseudonimisatie ZorgTTP

Stichting ZorgTTP, mei 2021

Inleiding

Deze factsheet is gericht op lezers zonder uitgebreide voorkennis over pseudonimisering. We beschrijven de uitgangspunten voor en procestappen van de door ZorgTTP ontwikkelde systematiek voor het pseudonimiseren van persoonsgegevens.

Uitgangspunten

Pseudonimisatie is een maatregel die wordt ingezet ter bescherming van persoonsgegevens in grootschalige gegevensverwerkingen. Door de herleidbaarheid van de verwerkte gegevens te beperken wordt de privacy van de betrokkenen beschermd. Pseudonimiseren is niet hetzelfde als anonimiseren. Bij anonimiseren is het niet langer mogelijk om te herleiden. Dit is in de praktijk zeer ingewikkeld als je nog steeds zicht wilt hebben op individuen binnen een dataset. Bij pseudonimisatie wordt ingezet in combinatie met andere beschermende maatregelen.

De kern van pseudonimiseren is het omzetten van direct herleidbare gegevens zoals de naam of unieke identificerende nummers naar één of meerdere codes; de pseudoniemen. Voor de overige gegevens wordt de afweging gemaakt of deze worden verwijderd, gecodeerd of geaggregeerd. Deze afweging is gebaseerd op het uitgangspunt van dataminimalisatie uit de Algemene Verordening Gegevensbescherming (AVG)

Varianten

Pseudonimisatie kan omkeerbaar en onomkeerbaar worden opgezet. ZorgTTP biedt beide varianten aan. In deze factsheet wordt het proces van onomkeerbaar pseudonimiseren beschreven. Hierbij wordt pseudonimiseren gedefinieerd als *het onomkeerbaar omzetten van een persoonsgegeven naar een niet tot de oorspronkelijke persoon terug te herleiden unieke code*.

Essentie

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 als de afnemer van de gegevens in de pseudonimisatieketen. Op deze wijze kan de relatie tussen pseudoniem en persoonsgegeven zowel technisch als organisatorisch worden verbroken. Na pseudonimisatie 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. Het is wel mogelijk om met dezelfde input tot hetzelfde pseudoniem te komen. Op deze wijze kunnen individuen gevolgd worden ten behoeve van beleid en onderzoek zonder dat hiervoor direct identificerende gegevens beschikbaar komen buiten de omgeving waarbinnen ze zijn vastgelegd.

Procesverloop

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.

Het pseudonimisatieproces bestaat uit de volgende stappen:

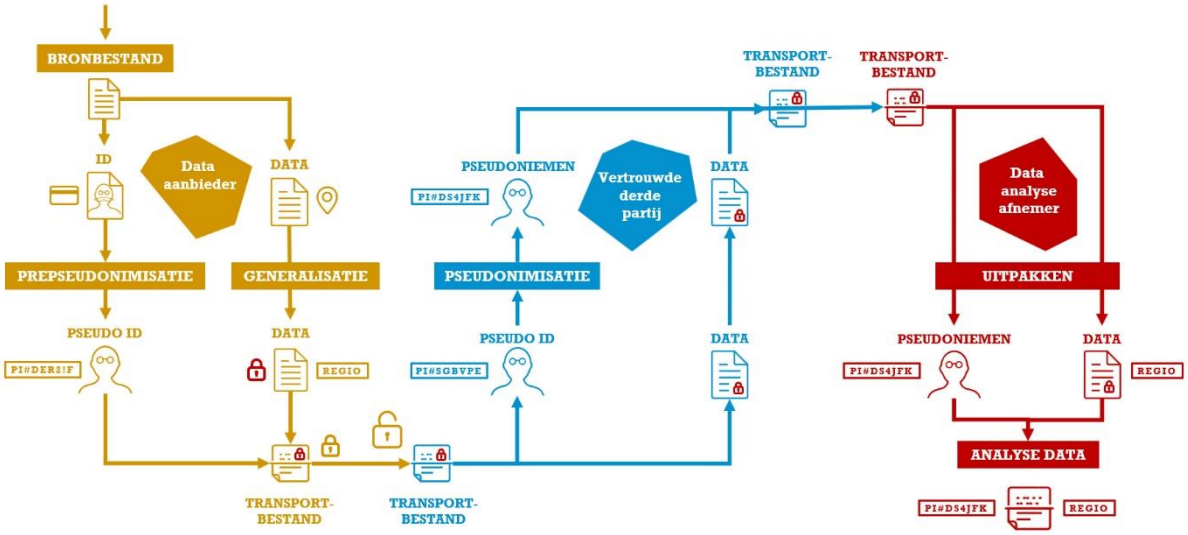
1. De informatiebron biedt een bestand aan dat voldoet aan de vooraf gedefinieerde en overeengekomen berichtspecificaties. Het bestand bevat persoonsgegevens en (medisch) inhoudelijke data;
2. Het bestand wordt bij en door de aanbieder verwerkt met de door ZorgTTP beschikbaar gestelde verzendsoftware;
3. De verwerking bestaat uit het:
    a. Scheiden van de persoonsgegevens en overige (medisch) inhoudelijke data.
    b. Omzetten van de persoonsgegevens tot een pre-pseudoniem.
    c. Beveiligen van de te verzenden gegevens.
4. Hierna volgt transport via een beveiligde internetverbinding naar ZorgTTP;
5. ZorgTTP voert op de centrale verwerkingsomgeving een tweede bewerking uit op de ontvangen pre-pseudoniemen. Voor iedere afnemer wordt een specifieke encryptiesleutel gebruikt. Uitkomst

van deze bewerking is een definitief pseudoniem dat niet kan langer algoritmisch kan worden omgezet naar de oorspronkelijke input;
6. Vervolgens worden de gepseudonimiseerde gegevens vrijgegeven om te worden opgehaald door de beoogde afnemer met een daartoe beschikbaar gestelde ontvangstmodule.
7. ZorgTTP heeft tijdens het pseudonimisatieproces geen toegang tot het (medisch) inhoudelijke datadeel. Dit is beveiligd en kan enkel door de ontvangende partij worden ontsleuteld middels de ontvangstmodule.

Tijdens het pseudonimisatieproces wordt een gelaagd model van beveiliging gehanteerd:
- Pseudonimisatie op recordniveau;
- Versleuteling op bestandsniveau;
- Transportbeveiliging;
- Controle van afzender van verzonden bestand middels certificaat

Voordeel van deze aanpak is dat als één laag doorbroken wordt, dan resteren er nog andere om de gegevens te beschermen.

Pseudonimisatieketen
Een pseudonimisatieketen is een klantspecificatie configuratie van drie op elkaar afgestemde (software)componenten. Dit zijn de:
1. Privacy- en Verzend Module (PVM) die wordt gebruikt door de aanbieder (gele deel in onderstaand schema);
2. Centrale Module TTP (CMT) die wordt gebruikt door ZorgTTP (blauwe deel);
3. Doel- en Receive Module (DRM) die wordt gebruikt door de afnemer (rode deel).



A. Werking Privacy en Verzend Module (PVM)
Deze module wordt gebruikt door de aanbiedende partij; het gele deel in bovenstaand schema. 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.

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 eerste versleuteling die plaatsvindt bij de partij die beschikt over de te verzenden persoonsgegevens 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 afnemer. De DRM ontvangt berichten van de centrale TTP applicatie. Na ontvangst wordt de transportbeveiliging verwijderd en worden beide berichtdelen samengevoegd. Het samengevoegde bestand bevat gegevens die niet zondermeer te herleiden zijn 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).