

Data-eigenaarschap

Dr. Jan Rombout, MBA medische data & KNO-arts: 23-10-2021

Hoe hou ik mijn eigen gegevens als zorgprofessional en die van m'n patiënten veilig? Onze zorgdata kunnen hergebruikt worden om de zorg kwalitatief beter en efficiënter te maken: hiermee wordt waarde toegevoegd. Dit document laat zien dat zorgdata breder hergebruikt kunnen worden dan tot nog toe gebruikelijk is.

1. Wie is de eigenaar van data?

Data worden opgeslagen om te worden (her)gebruikt, anders heeft het geen zin ze op te slaan. Bij het opslaan van persoonlijke data gaat de Europese 'General Data Protection Regulation' (GDPR) ¹ ervan uit dat alleen noodzakelijke gegevens bewaard worden. Privacy wordt beter door alleen relevante data te verzamelen en op te slaan.

Inmiddels is ook in China de privacywetgeving op hetzelfde niveau als in Europa ². In Amerika echter zijn de meningen over data-eigenaarschap anders dan in Europa. In Amerika worden data gezien als eigendom van een zorginstelling of bedrijf. De voornaamste reden om ze geheim te houden is een concurrentievoordeel. Deze andere kijk op data heeft zowel voor- als nadelen, waar ik verderop in dit document op in zal gaan.

Data zijn eigendom van de personen tot wie deze data herleidbaar zijn. Volgens de GDPR geven personen vrijwillig toestemming voor opslag en (her)gebruik van hun data. Dit geldt dus zowel voor personen die zowel cliënt, als patiënt, alsook zorgverlener kunnen zijn. Ik voeg daaraantoe dat deze privacyregels ook gelden voor herkenbare groepen van personen.

2. Drie Levels van anonimiteit

Om duidelijkheid te scheppen in de verschillende niveaus van anonimiteit worden in dit document drie levels geformuleerd:

1. Niet-anoniem: persoonlijke data, data met gedeeld eigenaarschap
2. Semi-Anoniem: interne data, die na anonimiseren toch re-identificeerbaar zijn:
 - a. Als de groep waarover gerapporteerd wordt klein is
 - b. De rapportage persoonlijke kenmerken bevat
 - c. Als er genoeg gegevens beschikbaar zijn over een individu uit andere bronnen
3. Volledig-anoniem: zorgdata over grote groepen mensen, welke niet meer herleidbaar zijn en dus openbaar gerapporteerd kunnen worden.

Deze drie levels bepalen wie er toegang heeft tot de data. Niet-anonieme data kunnen alleen door de eigenaar, of met diens expliciete toestemming, gebruikt worden. Dat wil zeggen dat diegenen met toegang geheimhoudingsplicht hebben.



Semi-anonieme data kunnen alleen intern in een organisatie gebruikt worden. Deze gegevens zijn van belang voor het optimaliseren van bedrijfsprocessen en verbeteren van bv. behandelingen.

Anonieme data betreffen altijd een gemiddelde van een flink aantal mensen, zodat ze nooit re-identificeerbaar zijn. Ook zijn anonieme data niet herleidbaar tot groepen mensen. Deze indeling, specifiek voor zorgdata, wordt verder uitgewerkt in de onderstaande tabel:

<i>Naam</i>	<i>Beschrijving</i>
<i>Data-eigenaarschap</i>	Personen zijn eigenaar van hun data, toestemming voor (her)gebruik moet vrijwillig gegeven worden. Anonieme data zijn eigendom van de persoon of organisatie die de data heeft verzameld.
<i>Gedeeld-eigenaarschap</i>	Data die herleidbaar zijn tot duidelijk omschreven groepen van personen zijn eigendom van die groep, een vertegenwoordiger van die groep wordt gevraagd om toestemming voor (her)gebruik, dit geldt dus ook voor groepen groter dan 30 personen (zie volgende punt).
<i>Anonimiteits-levels</i>	1. Niet-anoniem (persoonsgegevens): 1-14 personen. Data van individuen en kleine groepen kunnen alleen met toestemming van de eigenaar en binnen een organisatie door mensen met geheimhoudingsplicht worden behandeld. Dit zijn persoonsgegevens waar de GDPR volledig op van toepassing is. 2. Semi-anoniem : 15-29 personen. Data over groepen van 15-29 personen zijn 'semi-anoniem' en kunnen alleen intern ingezet worden. Dit geldt ook voor zorgdata die geanonimiseerd zijn door alleen persoonsgegevens te verwijderen en dus data per persoon herkenbaar kunnen zijn. 3. Volledig anoniem : 30 of meer personen. Data die statistieken tonen van 30 of meer personen (bv. gemiddelden), waarin individuen niet meer te onderscheiden zijn, kunnen als volledig anoniem worden beschouwd en mogen openbaar gerapporteerd worden.

3. Ook zorgprofessionals hebben privacy

De rapportage-regels om privacy te waarborgen uit bovenstaande tabel zijn ook van toepassing op zorgprofessionals. Volgens de Nederlandse wet moeten zorgprofessionals hun BIG-registratienummer onder alle correspondentie en data zetten. Het is niet toegestaan gegevens op basis van BIG-nummer te verzamelen en over individuele artsen te rapporteren.

Ook als data van artsen herleidbaar zijn tot een omschreven groep, zoals 'de vakgroep KNO-artsen van een ziekenhuis', vallen ze onder *gedeeld eigenaarschap* en zal in dit specifieke geval de betreffende KNO-vakgroep toestemming moeten geven voor (her)gebruik.



4. Wij kunnen veel leren van verzamelde zorgdata

Uit de tabel blijkt dat er binnen een organisatie veel mogelijk is met zorgdata. Er mag over groepen groter dan 14 personen vrijelijk intern semi-anoniem gerapporteerd worden. Rapportages over kleinere groepen dan 15 personen zijn mogelijk, als daar toestemming voor is gegeven, dit laatste geldt bv. voor wetenschappelijk onderzoek.

Rapportages over groepen van personen groter dan 30, waarin geen individuen meer te herkennen zijn (dus bv. geen puntenwolk) en die niet over een groep gaan, kunnen zelfs openbaar gemaakt worden. De zorg kan een stuk transparanter worden als vergelijkbare rapportages gemaakt worden, over bv. uitkomsten van zorg, zodat verschillende zorginstellingen zich onderling kunnen vergelijken.

In Amerika is het makkelijker om gegevens te verzamelen in een groot 'data lake'. De meeste datasets met trainingsdata voor Artificiële Intelligentie worden dan ook daar samengesteld. Dit lijkt een voordeel, maar is ook een nadeel omdat er maar een paar grote datasets zijn die telkens hergebruikt worden en waarvan de relevantie onduidelijk is. Het blijkt bv. dat CT-herkenningsalgoritmes geen significante bijdrage hebben gehad aan de herkenning en behandeling van Covid³. Deze meer dan 300 algoritmes waren getraind op een *data lake* en goedgekeurd door de USA Food and Drug Administration. Daarnaast weten wij niet of die Amerikaanse data wel van toepassing zijn op de Europese situatie.

'Federated learning'⁴ is een decentrale en privacy-vriendelijke vorm van 'machine learning'. Dat wil zeggen dat er geen centrale database nodig is waar alle gevoelige data in moet worden gestopt (*data lake*). In plaats van dat de data naar het *machine learning* model wordt gebracht, brengt *Federated Learning* het *machine learning* model naar de data. Hiermee blijven de data binnen de digitale muren van de zorginstelling en komt het algoritme langs om getraind te worden. Op deze manier kunnen wij in Nederland en binnen Europa nieuwe kennis ontwikkelen over dagdagelijkse zorg, die direct toepasbaar is op Nederlandse/ Europese patiënten en niet geïmporteerd is uit een onvergelijkbare populatie.

Een voorwaarde voor rapportages over dagdagelijkse zorg, betrouwbare statistiek en *federated learning* is een gedeeld dataformat (zie eerdere RomRom blogpost⁵). Het MPS®-dataformat is in ontwikkeling om gestandaardiseerde rapportages te kunnen maken. Uiteindelijk komt het MPS®-dataformat in het publieke domein als opslag-standaard.

5. Zorgdata kunnen goed hergebruikt worden

In dit document worden rapportageregels voorgesteld, volledig binnen de GDPR. Het blijkt dat het goed mogelijk is om zorgdata semi-anoniem binnen een zorgorganisatie te bewerken, zodat er volledig anoniem naar buiten gerapporteerd kan worden.

Door zorgdata te hergebruiken, bijvoorbeeld met *federated learning* kan er een fantastische nieuwe bron van kennis aangeboord worden. Wij kunnen met elkaar waarde creëren en toevoegen aan de zorg door onze zorgdata optimaal (en anoniem) in te zetten.

¹ In Nederland vertaald naar de 'Algemene Verordening Gegevensbescherming' (AVG) Wikipedia: https://nl.wikipedia.org/wiki/Algemene_verordening_gegevensbescherming

² Dutch IT channel: <https://dutchitchannel.nl/679699/china-scherpt-privacywetgeving-aan.html>

³ Eric Topol op twitter: <https://twitter.com/erictopol/status/1371496436764975104>

⁴ TNO: <https://www.tno.nl/nl/aandachtsgebieden/informatie-communicatie-technologie/roadmaps/data-sharing/federated-learning>

⁵ RomRom: <https://www.romrom.nl/zorg-datamodel-voor-een-level-playing-field>