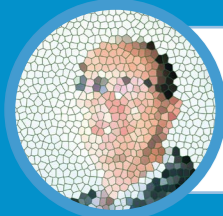


ZorgTTP Team



Hans van Vlaanderen

Directeur / Bestuurder



Marnix Bindels

Software Architect



André Kaldenhoven

DevOps Engineer



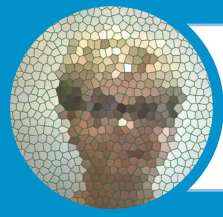
Tom

Software Engineer



Erik van Leer

Operationeel Coördinator



Edith Roerhorst

Operationeel Beheer & Servicedesk



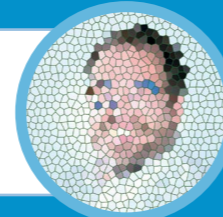
Mark van Kersen

Operationeel Beheer & Servicedesk



Simone van Wijngaarden

Teammanager Advies



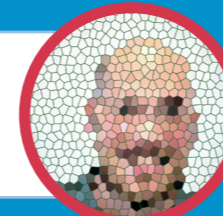
Michiel Vlastuin

Adviseur & Product Owner



Joost Verduijn

Adviseur & Functionaris Gegevensbescherming



Bastiaan van Schijndel

Adviseur & Scrum Master



Martijn Kuijpers

Adviseur



Leon de Jong

Adviseur

Inhoudsopgave

Programma

Het programma van het jubileumcongres op dinsdag 29 november 2022

Sprekers en presentaties

Introductie van de sprekers en de presentaties

Tijdlijn

Belangrijke momenten uit de afgelopen 15 jaar en een blik in de toekomst

Introductie van de PVM Link

Standaard software voor onderzoekers bij eenvoudige en eenmalige koppelingen

De inzet van een Trusted Third Party bij data migraties

Met een praktijkvoorbeeld van Dick Schuurman van de Nederlandse Hart Registratie over de uitgevoerde migratie van de NHR pacemakerregistratie

Overzicht Pseudonimisatieplatform

Kengetallen over het gebruik van het pseudonimisatieplatform

Synthetische data

Een toelichting van Khaled el Emam van Replica Analytics over synthetische data en de mogelijke use cases

Prematch

Prematch factsheet: over het bepalen van de overlap tussen twee of meer datasets ten behoeve van het proportioneel delen van data

Koppelservice

Veilig, eenvoudig en snel kijken naar overlap middels het koppelservice platform

JUBILEUM: 15 JAAR ZORGTTP

Locatie:
Het Spoorwegmuseum
Maliebaanstation 16
3581 XW Utrecht

De pendeltrein
vertrekt één keer per
uur vanaf Utrecht CS.
Vertrek 09:31 uur.

Programma dinsdag 29 november 2022

- 09:45 - 10:15 Inloop en aanmelding
- 10:15 - 10:25 Welkomstwoord - Dagvoorzitter Michiel Eijsbouts
- 10:25 - 11:45 Data delen via een Trusted Third Party

Wat is een TTP en wat kun je er mee?
Positieve blik op data delen voor secundair gebruik
Onderzoekswerkplaats "routine zorgdata voor passende zorg"
Het proces van data linkage en data uitgifte

- 11:45 - 12:30 Lunch
- 12:30 - 13:40 Onderzoekers aan het woord
Gebruik van het gepseudonimiseerd BSN bij gegevenskoppelingen
Analyse van gebruikte naam- en adrespseudoniemen
Onderzoeker aan het woord; de Healthy Pregnancy 4 All-2-studie
Onderzoeker aan het woord; de SELECT-studie
- 13:40 - 14:00 Pauze
- 14:00 - 15:30 Vooruitkijken en mogelijkheden voor de toekomst
De European Health Data Space
E-Health; kansen voor data gestuurde zorg
De toekomst van de kwaliteitsregistratie in het registratielandschap
Multi Party Computation
- 15:30 - 16:30 Borrel

Sprekers en presentaties

Data delen via een Trusted Third Party



Hans van Vlaanderen
Directeur/Bestuurder
Stichting ZorgTTP

Wat is een TTP en wat kun je er mee?

Wat is een TTP en wat kun je er mee?
Welke technische en organisatorische maatregelen maken ZorgTTP een Trusted Third Party voor het beveiligen van data voor beleid en onderzoek? Het ontstaan en de (door)ontwikkeling van de ZorgTTP pseudonimisatieservice aan de hand van een tijdlijn.



Robert Verheij
Programmaleider Zorgdata en het Lerend Zorgsysteem
Nivel Zorgregistraties Eerste Lijn

Positieve blik op data delen voor secundair gebruik

Nivel Zorgregistraties en het Lerend Zorgsysteem. Toelichting op de samenwerking tussen Nivel en ZorgTTP voor Nivel Zorgregistraties. Over de gegevensverzameling en onderzoeken binnen de eerste lijn gezondheidszorg en de vervolgstappen voor een lerend zorgsysteem.



Christel van Dijk
Coördinator Onderzoekswerkplaats
Zorginstituut Nederland

Onderzoekswerkplaats "routine zorgdata voor passende zorg"

Over het samenbrengen van reeds gepseudonimiseerde data met inachtneming van het proportionaliteitsbeginsel. Een samenwerking tussen Nivel en het Zorginstituut waar kennis en expertise over zorg en het zorgstelsel bijeen wordt gebracht voor overstijgende beleidsvragen.



Trynke de Jong
Projectleider & Privacy Officer Lifelines
Medische Biobank Noord-Nederland

Het proces van data linkage en data uitgifte

Lifelines is een grote cohortstudie en biobank met meer dan 167.000 deelnemers, uit drie generaties, die minstens 30 jaar worden gevolgd om inzicht te krijgen in gezond ouder worden. Het doel van Lifelines is om een bron van data en lichaamsmaterialen te zijn voor de nationale en internationale wetenschappelijke gemeenschap met waarborging van de privacy van haar deelnemers. Een presentatie over het proces van gepseudonimiseerd koppelen, prematchen en veilig beschikbaar stellen van gegevens in een beveiligde workspace aan onderzoekers.



Klik voor treininfo



Klik voor parkeerinfo

Onderzoekers aan het woord



Sandra van de Belt-Dusebout

Postdoc Epidemioloog - Antoni van Leeuwenhoek-Nederlands Kanker Instituut



Rinus Voorham

Adviseur Gegevensaanvragen
Stichting Palga

Gebruik van het gepseudonimiseerd BSN bij gegevenskoppelingen

Een door BBMRI gefinancierde validatie koppelingenstudie van het Nederlands Kanker Instituut, samen met Palga en IKNL, met gebruik van gepseudonimiseerde persoonsgegevens vanuit het OMEGA-cohort van VUMC en UMCU.



Henrike Galenkamp

Wetenschappelijk Coördinator HELIUS-studie
Amsterdam UMC, locatie AMC

Analyse van veelgebruikte naam- en adrespseudoniemen

Voor de studie "Ethnicity and COVID-19" zijn test- en vaccinatiedata van de GGD Amsterdam en data van het HELIUS-cohort van het Amsterdam UMC met de grootste zorgvuldigheid samengebracht. Onderdeel van het traject was een analyse van de betrouwbaarheid van de gebruikte naam- en adrespseudoniemen.



Vooruitkijken en mogelijkheden voor de toekomst



Roger Lim

Coördinerend Beleidsmedewerker
Ministerie van Volksgezondheid, Welzijn & Sport

De European Health Data Space

Op weg naar een Europese ruimte voor gezondheidsgegevens. Een toelichting op de EDHS met een focus op secundair gebruik van data, de juridische grondslagen en de positionering ten opzichte van andere EU-wetgeving, de maatregelen en wat dat van ons vraagt. Daarnaast wordt een toelichting gegeven op de Nederlandse aanpak in de voorbereidingen op de onderhandelingen binnen de EU.



Lisette van Gemert-Pijnen

Hoogleraar Gezondheidstechnologie
Universiteit Twente & Lid RvT Stichting ZorgTTP

E-Health; kansen voor data gestuurde zorg

Een toelichting op de uitkomsten van het rapport "Digitalisering in de gezondheidszorg nader beschouwd" (opdrachtgever VWS), het Integraal Zorgakkoord en de implicaties ervan over de zorgdomeinen heen. Hoe wordt de zorg straks georganiseerd en wat is de rol van E-Health? Welke informatie is daarbij van belang en wat zijn de kansen voor het delen en benutten van data voor de toekomst?



Minke van Minde

Promovenda, Erasmus MC Rotterdam
Huisarts in Opleiding, UMC Utrecht

Onderzoeker aan het woord; de Healthy Pregnancy 4 All-2-studie

Een presentatie over het onderzoek waar met behulp van gepseudonimiseerde data vanuit consultatiebureaus onderzoek is gedaan met als uiteindelijk doel het terugdringen van perinatale gezondheidsverschillen en verschillen in kindergezondheid in Nederland.



Max Ragusi

AIOS Radiologie
UMC Utrecht

Onderzoeker aan het woord; de SELECT-studie

De resultaten van het onderzoek voor het identificeren van borstkankerpatiënten die wel of geen baat zullen hebben bij anti-hormonale behandeling door middel van de beoordeling van beeldmateriaal (MRI's) en onderzoeksgegevens vanuit Palga, IKNL en negen betrokken ziekenhuizen.



Gert-Jan van Boven

Directeur
Stichting Dutch Hospital Data

De toekomst van de kwaliteitsregistratie in het registratielandschap

Naast veel verbetermogelijkheden brengen kwaliteitsregistraties de noodzaak van registratie met zich mee. Er wordt hard gewerkt om die goed te kunnen inpassen in de dagelijkse praktijk. Zelfs de wetgever doet een duit in het zakje. Hoe zorgen we ervoor dat alles straks gaat werken?



Pieter Verhagen

Co-founder en COO
Linkisight



Hans van Vlaanderen

Directeur/Bestuurder
Stichting ZorgTTP

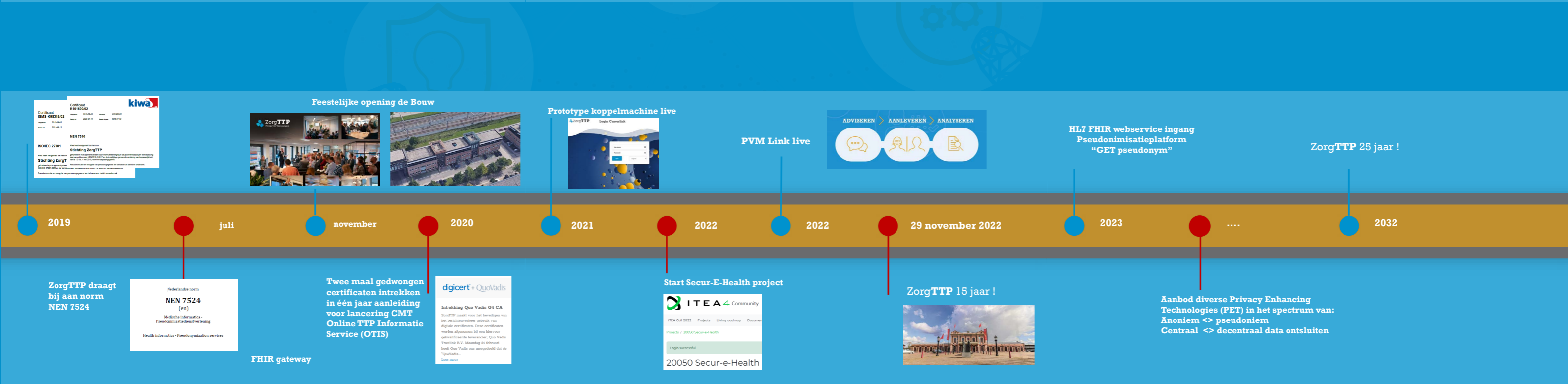
Multi Party Computation

Een duo-presentatie over de innovatieve samenwerking tussen ZorgTTP en Linkisight op het gebied van Multi Party Computation binnen het Secur-E-Health project. Het project richt zich op "privacy preserving" technieken bij organisatieoverschrijdende gegevensanalyse.



Tijdslijn

Belangrijke momenten uit de afgelopen 15 jaar en een blik in de toekomst



ZorgTTP introduceert de PVM Link 1.0 voor eenmalige en eenvoudige koppelingen

Generieke software

Stichting ZorgTTP innoveert en ontwikkelt voortdurend, om zo onderzoekers en organisaties die onderzoek doen efficiënter te laten werken. We zien van dichtbij hoe het afstemmingsproces in de voorbereidende fase van onderzoek vaak een remmende factor is. Daarom heeft ZorgTTP intensief geïnvesteerd in de ontwikkeling van een generieke Privacy- en Verzendmodule (PVM). Met de PVM Link 1.0, zoals we deze software noemen, ondersteunen we onderzoekers bij het efficiënt realiseren van eenmalige en relatief eenvoudige koppelingen.

Proces versnellen met een standaard werkwijze

Voor het gebruik van de PVM Link hanteert ZorgTTP een standaard werkwijze. Hiermee kunnen we het proces aanzienlijk versnellen. Wij beoordelen in het eerste adviesgesprek of de PVM Link 1.0 bij jouw vraagstuk past. Is dit het geval, dan stellen we een offerte op met vaste prijzen voor het gebruik van de software. Daarnaast sluiten we per verwerking een verwerkersovereenkomst af met de verantwoordelijke partij (Verwerkersverantwoordelijke). Voor deze overeenkomst hebben wij een template beschikbaar. Door de generieke opzet van de software is er geen configuratie van de software nodig. Daardoor is, na het afstemmen van een aantal projectspecifieke parameters, de software direct beschikbaar voor eenmalige en eenvoudige koppelingen.

Eenvoudig gebruik

Standaard software

Research IT en datamanagement

Volgen over tijd en plaats

Software van ZorgTTP vervangt persoons-identificerende kenmerken, zoals naam of BSN, door een specifiek pseudoniem. Dat houdt in dat voor iedere persoon steeds hetzelfde pseudoniem wordt gebruikt. Individuen zijn hierdoor koppel- en volgbaar in tijd en plaats, zonder de oorspronkelijke persoonsgegevens te gebruiken. Zo is mogelijk informatie op individueel niveau uit te wisselen en onderzoek uit te voeren conform de wettelijke vereisten.

Direct beschikbaar voor gebruik

ZorgTTP levert de PVM Link aan, of je kunt deze downloaden van onze website. Om de PVM Link te kunnen gebruiken, moet de te versturen data aan bepaalde berichtspecificaties voldoen. De PVM genereert een aantal veelgebruikte pseudoniemen waarmee je eenvoudige koppelingen kunt uitvoeren. Daarnaast kunnen de meeste organisaties eenvoudig overweg met het te gebruiken bestandsformat.

Borgen op een centrale plek in de organisatie

Een ander voordeel van deze generieke software is dat je de tool binnen organisaties beschikbaar kunt stellen aan specifieke (research)afdelingen of datastewards. Zo kun je effectief inspelen op de toenemende ondersteuningsvraag van onderzoekers voor het veilig gebruik van data. Bovendien kun je op deze manier de software van ZorgTTP op een centrale plek binnen een organisatie borgen.

Heb je als onderzoeker of als medewerkers van een research - IT afdeling vragen over onze dienstverlening in het algemeen, of over de PVM Link in het bijzonder, neem dan contact op met één van onze adviseurs via advies@zorgtpp.nl of 030-6360649.

De inzet van een Trusted Third Party bij datamigraties

In opdracht van de Nederlandse Hart Registratie (NHR) heeft ZorgTTP als Trusted Third Party opgetreden bij de migratie van registratie gegevenssets, met daarin opgenomen persoonsgegevens, naar MijnNHR. Een Trusted Third Party (TTP) is een vertrouwde derde partij en speelt een onafhankelijke rol bij de uitwisseling van data tussen twee of meer partijen. De TTP draagt zorg voor het beschermen van de privacy van betrokkenen met inachtneming van geldende wet- en regelgeving en met gebruik van state-of-the-art technologie. Een TTP wordt veelal ingezet door onderzoekers en organisaties waar enkel met gepseudonimiseerde gegevens gewerkt mag worden. De TTP speelt hier een belangrijke rol bij het pseudonimiseren van de gegevens en de beveiligde levering van data, maar heeft hier nadrukkelijk zelf geen toegang tot de gepseudonimiseerde data. Een TTP kan echter ook ingezet worden bij projecten waar partijen liever zien dat data-uitwisseling door een TTP wordt gefaciliteerd dan door één van de betrokken partijen of in situaties waar gewerkt wordt met persoonsgegevens, bijvoorbeeld bij migraties waar overgestapt wordt van versleutelingsmethode en/of leverancier.



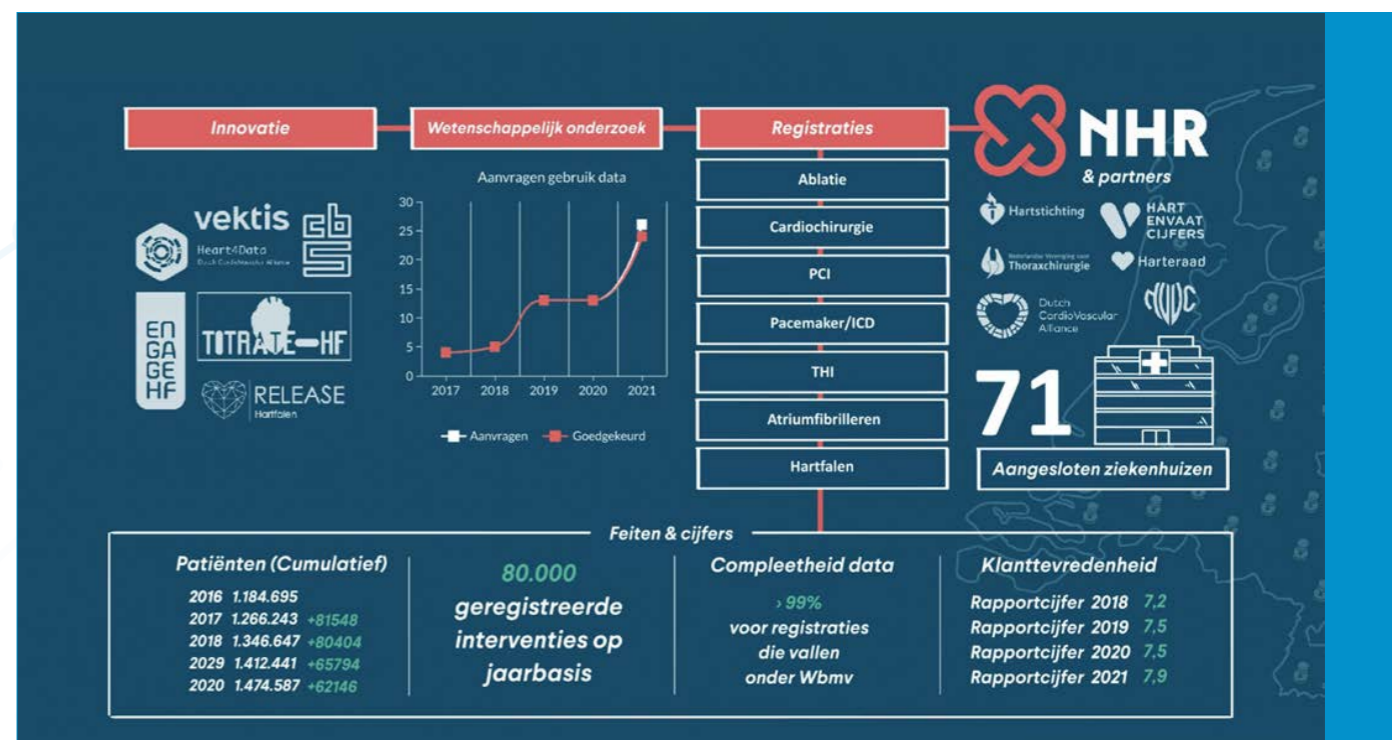
“ Een TTP is een mooie oplossing als partijen gevoelige persoonsgegevens willen uitwisselen. Dit omdat een TTP speciaal is ingericht om te zorgen dat de veiligheid en vertrouwelijkheid gewaarborgd is. ”

Dick Schuurman, Adviseur bij de NHR

Migratie van de Pacemaker/ICD registratie van de NHR

De NHR beheert diverse registraties waaronder Ablatie, Atriumfibrilleren, Cardiochirurgie, Hartfalen, PCI, Pacemaker/ICD (PM/ICD) en THI. Alle registraties, met uitzondering van de Pacemaker registratie, waren al opgenomen in MijnNHR. MijnNHR is de afgeschermd online omgeving waar gegevenssets over behandelingen en ziektebeelden veilig worden aangeleverd en opgeslagen. Deelnemende ziekenhuizen krijgen in deze omgeving vervolgens directe feedback.

De opdracht aan ZorgTTP betrof enerzijds het decrypteren van persoonsgegevens in de gegevenssets die werden aangeleverd door de partij die op dat moment de registratie beheerde en anderzijds het uploaden van deze gegevenssets naar MijnNHR. Om de vertrouwelijkheid van de vertrouwelijke gegevens te borgen, heeft de NHR aan ZorgTTP gevraagd deze migratie-opdracht uit te voeren.



De Nederlandse Hart Registratie; drie primaire processen

De NHR kent een drietal primaire processen: registratie, wetenschap en innovatie, met als doel kwaliteitsbewaking en -verbetering te bevorderen in de hartzorg in Nederland. De NHR werkt hiervoor intensief samen met verschillende partijen, waaronder de centra, de Nederlandse Vereniging voor Cardiologie (NVVC), de Wetenschappelijke Vereniging voor cardio-thoracaal chirurgen (NVT), de Dutch CardioVascular Alliance (DCVA) en de Hartstichting.

Binnen het primaire proces "Registratie" worden registraties die een verplicht karakter hebben, of dusdanig ver ontwikkeld zijn dat er sprake is van een stabiele situatie die onderhouden moet worden, uitgevoerd.

Projectaanpak en samenwerking

Aansturing van de migratie werd geleid door Dick Schuurmans, wiens werkzaamheden binnen de NHR zijn gericht op ICT en informatiebeveiliging. Voor het project is een hoofdlijnenplan gemaakt, aangevuld met gedetailleerde stroomschema's. Overlegmomenten hebben met name plaatsgevonden bij aanvang van het project met een groot aantal betrokkenen. Op die manier is bij aanvang zoveel mogelijk input verzameld en een gemeenschappelijk kader gecreëerd. Op basis van die gesprekken is structuur gegeven aan de uit te voeren werkzaamheden.

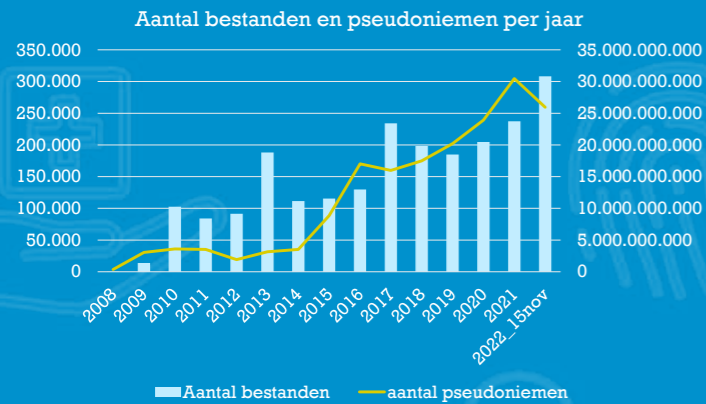
Omdat veel partijen betrokken waren en deze geen toegang mochten en wilden hebben tot de vertrouwelijke gegevens gedurende de migratie wilde de NHR bij voorkeur een TTP inschakelen. Dit omdat een TTP speciaal is ingericht om te zorgen dat de vertrouwelijkheid gewaarborgd is. Binnen het migratietraject had ZorgTTP de verantwoordelijkheid voor het decrypteren van de vertrouwelijke gegevens en het uploaden van gegevenssets in MijnNHR. Het traject is succesvol, veilig en volgens afspraak verlopen en alle registraties van de NHR zijn nu opgenomen in MijnNHR. Grote voordelen voor de NHR zijn nu dat alle registraties door dezelfde leverancier worden beheerd, overal dezelfde versleuteling wordt gehanteerd en koppeling tussen registraties mogelijk is.

Volgend project; migratie aangeboren en verworven hartaandoeningenregistraties

Een volgend project rondom de overgang van twee registraties van kinderen en volwassenen met aangeboren en verworven hartafwijkingen staat al op stapel. De opzet hiervan is complex en vraagt de nodige voorbereiding om de migratie straks soepel te laten verlopen. Het borgen van de vertrouwelijkheid van de gegevens is ook hier een belangrijk item, om deze reden heeft de NHR aan ZorgTTP gevraagd wederom als TTP in deze migratie op te treden. ZorgTTP zal hierbij worden ingezet voor het decrypteren van versleutelde gegevens, het verrijken van de gegevens en het uploaden van gegevenssets naar MijnNHR. De reden voor de NHR om voor dit project weer voor ZorgTTP te kiezen is de prettige en professionele samenwerking bij de migratie van de PM/ICD registratie.

Pseudonimisatieplatform

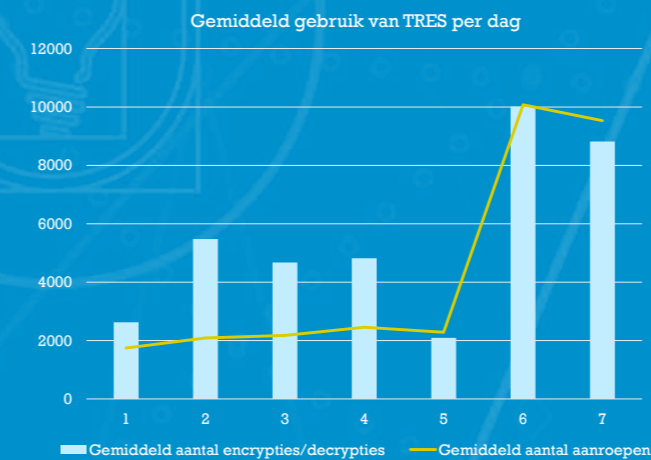
ONOMKEERBAAR



178.956.430.544

Totaal aantal verwerkte pseudoniemen

OMKEERBAAR



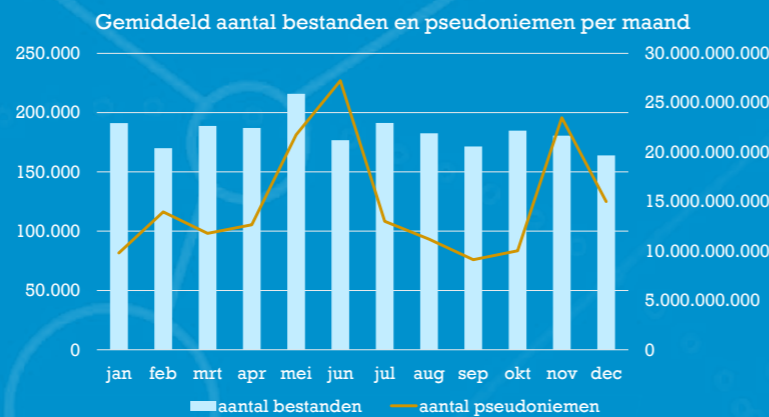
750.798.657

Totaal aantal TRES verwerkingen



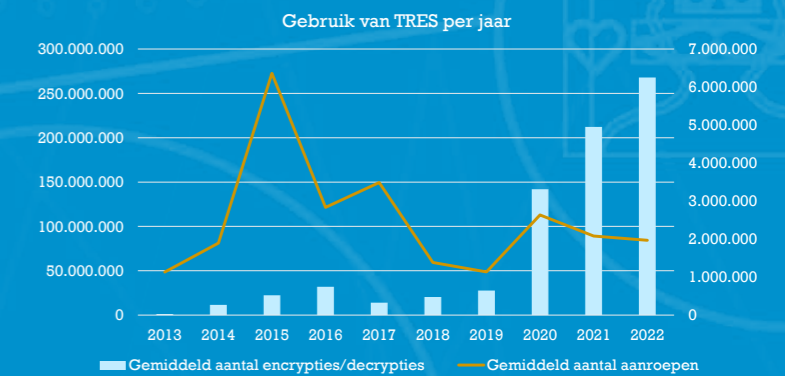
1.887.019.801

Het grootste aantal persoonsgegevens op één dag verwerkt



90.616.687

Het grootste aantal verwerkingen op één dag



2.203.994

Totaal aantal bestanden verwerkt



8.932

Totaal aantal bestanden voor domeinconversie aangeboden



150

Actieve domeinen



2007

In gebruik sinds



675

Actieve gebruikers in 2022



3.671

TRES gebruikers 2013-2022



122

Totaal ingerichte projecten



2013

In gebruik sinds

UNDERSTANDING AND LEVERAGING THE REAL PROMISE OF SYNTHETIC DATA

Regulation, cost and other factors can at times impede the great many societal and organizational benefits of sharing data. Data synthesis is rapidly emerging as a solution and the future of data sharing within and across borders- a practical privacy enhancing technology (PET) for sharing data for secondary purposes. The rate of adoption of data synthesis has been growing steadily in recent years.

What is synthetic data?

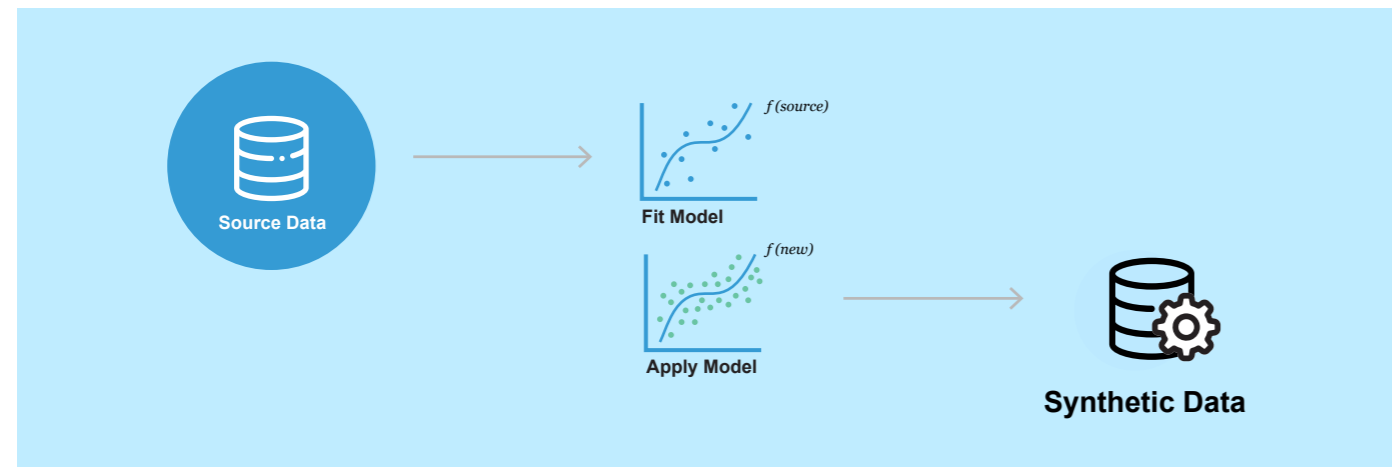
Synthetic data is generated from real data. It is created with a machine learning model that captures the patterns in the real data, and then generates new data from that model. The values in the synthetic data are generated from the model and closely capture the statistical properties of the original dataset.

If we are modelling the structure of the original data very accurately then the end result would effectively be replicating the original data, which would have negative

privacy implications. Therefore, there is a balancing act between how accurate the model needs to be and how close the synthetic data is to the original data. It is the classic trade-off between data utility and data privacy.

When done well the synthetic data retains sufficient statistical properties of the original data and has a very low risk of identifying individuals. This allows synthetic data to be shared more freely with minimal administrative and technical controls, and to be treated as non-identifiable data.

*Synthetic Data =
Non-identifiable Data*



How good is synthetic data?

Evidence is accumulating that synthetic data has high utility, in that it models the patterns in real data well. This means that the analysis done on synthetic data gives the same conclusions as the real data (reproducibility), and the population inferences from the analysis of synthetic data are statistically valid.

Synthetic data also has lower privacy risks because, when done well, there is no one-to-one mapping between the synthetic records and

real people. Two types of disclosure risks are relevant for synthetic data: attribution and membership disclosures. Empirical evaluations of these disclosure risks of synthetic data have demonstrated this in multiple studies.

Both the utility and privacy risks can be evaluated objectively for any specific dataset quite efficiently.

How can synthetic data be used?

Synthetic data is generated from real world data or clinical trial data. The process is largely automated which reduces the amount of training needed to use it effectively. It is not necessary to know how the generated synthetic data will be used (e.g., the kinds of statistical or machine learning analyses that will be performed) a priori to generate useful synthetic data.

Who is using synthetic data?

Private and public organizations that collect and analyze health data for secondary purposes are adopting synthetic data. These

Replica Synthesis Software

The Replica Synthesis software provides a highly automated and user-friendly capability to create synthetic data and to evaluate the data utility and privacy. In addition, users can create data simulators and share them with other users through sophisticated data access policies. Our objective was to make data synthesis and the sharing of synthetic data as simple as possible for end-users.

include the pharmaceutical industry to synthesize clinical trial datasets, insurers, health IT companies, health analytics companies, academic medical centers, community and specialty pharmacies, and health systems. Furthermore, organizations that are not primarily in the healthcare business, such as retail, finance, and urban development are linking with health data and applying synthetic data generation technologies to manage their risks and enhance their datasets.

Gartner has predicted that:

- By 2024, 60% of the data used for the development of AI and analytics solutions will be synthetically generated
- By 2024, synthetic data will halve the volume of real data needed for machine learning
- By 2025, synthetic data will reduce personal customer data collection, avoiding 70% of privacy violation sanctions

HOW CAN SYNTHETIC DATA BE USED?

Synthetic data solutions can help mitigate various data privacy and quality challenges, and ultimately represent tremendous opportunities for a variety of sectors and society at-large. Uses for synthetic data include the following.

Data sharing and data access

- **AI and data science projects.**

One of the biggest challenges when developing AI and machine learning algorithms is getting sufficient realistic datasets to train the models and test on. Data synthesis makes it easier to get rapid data access and to augment data.

- **Data Simulators.** Instead of sharing datasets, data custodians can share data simulators, which are models of the data. The privacy risks from data simulators are very small. Plus, it is much easier to accept the sharing of models rather than the sharing of data.

- **Proof of concept and technology evaluations.**

Oftentimes technology developers or acquirers need to quickly evaluate whether a new technology works well in practice and they need realistic data to work with as well as have minimal constraints on access to that data.

- **Data exploration.**

Organizations that want to maximize the use of their data can make synthetic versions available for exploration and assessment by potential users. If the exploration yields positive results, the users can validate the results on the real data by sending the modeling code developed on the synthetic data to a verification server.

- **Software testing.** Testing data-driven applications require realistic data for functional and performance testing. Random data cannot replicate what will happen when a system goes into production.

- **Open data / open science.** Sharing complex datasets publicly is challenging because of privacy concerns. This can now be achieved by sharing synthetic data instead.

- **Hackathons and data competitions/challenges.** These require datasets that can be distributed widely with minimal burden on the entrants. The datasets have to be realistic to enable meaningful innovation, which is what synthetic data provides.

Data enhancement

- **Amplifying small datasets.**

When datasets are small (e.g., rare diseases, pediatric clinical studies) then synthetic data generation methods can amplify the datasets by adding virtual patients with the same characteristics and patterns. This can improve machine learning model accuracy and stability.

- **Correct bias.** If a dataset is biased because, for example, of poor representation of certain individuals or groups, data synthesis can re-balance these datasets while maintaining the key relationships. This way inherent biases in the data can be mitigated.

- **Rescue studies with virtual patients.**

When recruitment is poor in clinical studies or participant attrition is high, virtual patients can be simulated using the data that has already been collected to correct for these challenges.

- **Virtual controls.** For single-arm studies, instead of using external controls in clinical trials, internal controls using already collected data can be simulated as virtual patients. This allows the design of studies with a smaller target recruitment.

Where can I learn more?

At Replica Analytics, we have world-renowned expertise and experience in data management, data analytics and privacy enhancing technologies. You can get more information about our work related to synthetic data from the **knowledgebase** on our website, which has some general information as well as links to published articles with deep dives into the methods and algorithms. You can always reach out to us if you have specific questions.



ZorgTTP en Replica Analytics

Stichting ZorgTTP innoveert en ontwikkelt voortdurend, om zo onderzoekers en organisaties die onderzoek zo goed mogelijk te ondersteunen. In dat kader zoeken we ook altijd de samenwerking met andere partijen.

Heb je als onderzoeker of als medewerker van bijvoorbeeld een research of IT-afdeling vragen over onze dienstverlening in het algemeen, of over synthetische data in het bijzonder, neem dan contact op met één van onze adviseurs via advies@zorgttp.nl of 030-6360649.

Prematchen: proportioneel data delen

Inleiding

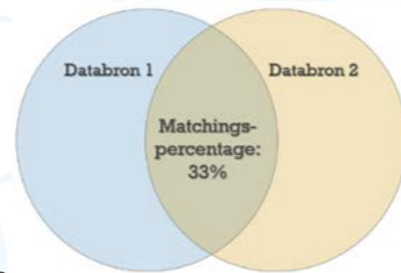
In gevallen waarin het nuttig is om enkel gegevens van individuen te verzamelen die bekend zijn bij meerdere databronnen, maar op voorhand niet bekend is om welke individuen het gaat, adviseren we om een pré-match uit te voeren. In deze factsheet wordt beknopt de door ZorgTTP ontwikkelde systematiek voor het uitvoeren van een pré-match bij de inzet van de pseudonimisatieketen beschreven.

Pseudonimisatie

De pseudonimisatieketen van ZorgTTP wordt ingezet als beveiligingsmaatregel bij gevoelige gegevensverzamelingen. Persoons-identificerende kenmerken zoals naam, geboortedatum, postcode en/of geslacht, worden bij pseudonimisatie vervangen door een pseudoniem, zodanig dat voor ieder persoonsgegeven steeds hetzelfde pseudoniem wordt gegenereerd. Individuen worden op deze wijze koppelbaar in tijd en over verschillende bronnen heen zonder dat daartoe de oorspronkelijke persoonsgegevens verstrekt hoeven te worden.

Pré-match & Algemene Verordening Gegevensbescherming (AVG)

Door middel van een pré-match kan ZorgTTP het matchingspercentage in twee of meerdere populaties bepalen, en daarbij aan de databronnen terugkoppelen welke individuen bekend zijn bij meerdere databronnen. De databronnen hebben nu de mogelijkheid om alleen de gegevens van de individuen die bij meerdere databronnen bekend zijn te verwerken via de pseudonimisatieketen naar het datadoel. Door bij een pré-match te bepalen van welke individuen het nuttig is om onderzoeksgegevens te verwerken, worden alleen de minimaal noodzakelijke gegevens verwerkt en alleen die gegevens die effectief zijn om het gestelde doel te bereiken. Zo wordt voldaan aan het beginsel van minimale gegevensverwerking (data-minimalisatie) en het proportionaliteitsvereiste uit de AVG.



Beveiliging

Bij het uitvoeren van een pré-match wordt gebruik gemaakt van de voor de opdrachtgever ingerichte pseudonimisatieketen. ZorgTTP ontvangt alleen gepseudonimiseerde persoonsgegevens en krijgt nimmer inzage in de oorspronkelijke persoonsgegevens. De betrokken databronnen zullen ook geen inzage krijgen in de persoonsgegevens van de andere databronnen. Na het uitvoeren van de pré-match kan de verrijkingfase van start gaan.

Specificaties

Voor de totstandkoming van een pré-match is het noodzakelijk om te weten welke persoonsgegevens bekend zijn bij alle betrokken databronnen en op welke wijze de matchingspercentages en matchende ID's moeten worden gerapporteerd. ZorgTTP adviseert haar opdrachtgevers graag om een optimale pré-match uit te voeren en de rapportage te verzorgen. Voor elke pré-match opdracht zal daarom uitgebreid worden overlegd met de opdrachtgever om tot de definitieve afstemming te komen.

Pré-match proces

Het Pré-matchproces bestaat in het kort uit de volgende stappen:

1. Pré-matchfase: Aanlevering aan ZorgTTP

Uitgangspunt is dat de verzendende partijen een bestand genereren dat voldoet aan vooraf vastgestelde specificaties met daarin alleen de identificerende persoonsgegevens, en door de verzendende partijen zelf bepaalde unieke identifiers;

2. Pré-matchfase: Uitvoering

ZorgTTP haalt de gepseudonimiseerde bestanden met de ontvangstmodule op en voert de pré-match uit om te bepalen welke pseudoniemen voorkomen in beide bestanden;

3. Pré-matchfase: Rapportage

ZorgTTP rapporteert het matchingspercentage van de betreffende pseudoniemen en koppelt de unieke identifiers van individuen die matchen terug aan de databronnen;

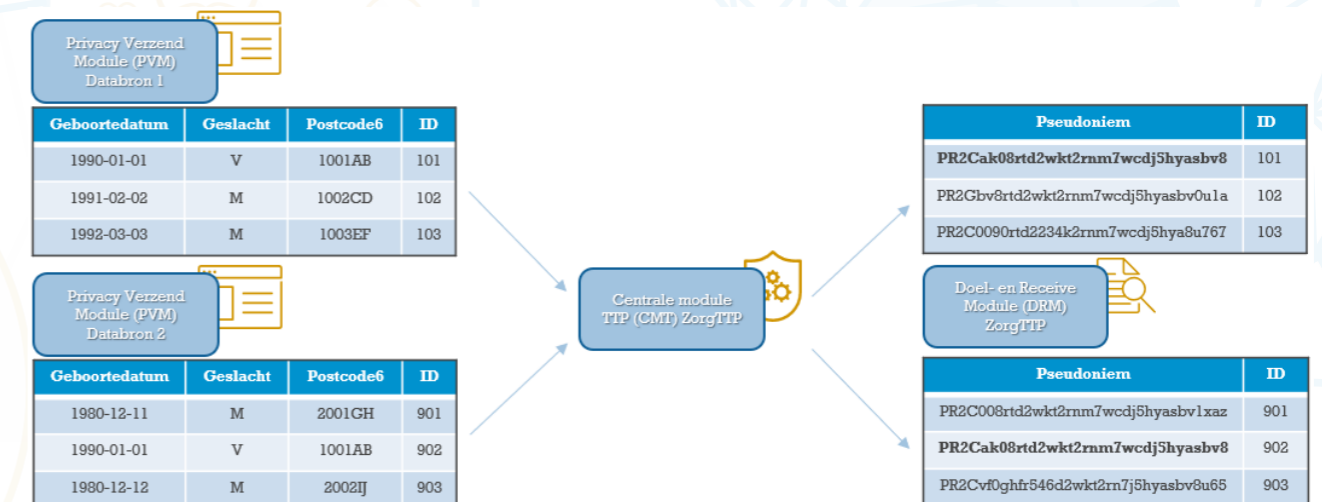
4. Verrijkingfase: Aanlevering aan onderzoekende partij

De databronnen kunnen van de matchende individuen de persoonsgegevens en onderzoeksgegevens naar de onderzoekende partij te sturen.

Fase 1: Pré-matchfase

Aanlevering aan ZorgTTP

In onderstaand versimpeld voorbeeld versturen twee databronnen elk van drie bij hen bekende individuen de persoonsgegevens en unieke identifiers aan ZorgTTP. Eén individu is bekend bij beide databronnen. Voor dit individu wordt in beide bestanden hetzelfde pseudoniem gegenereerd.



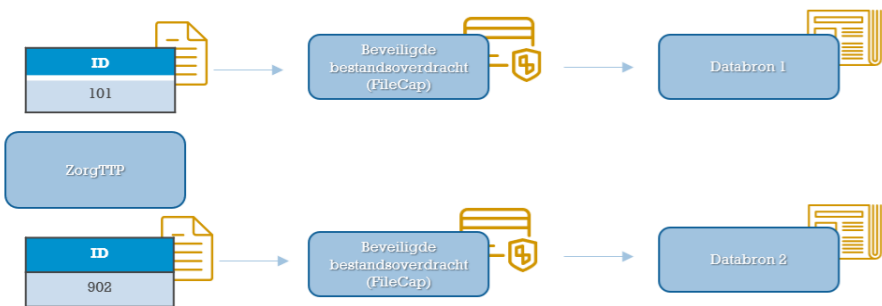
Uitvoering

ZorgTTP voert de pré-match uit op de gepseudonimiseerde gegevens. Afhankelijk van de wens van de opdrachtgever kunnen de matchingspercentages en ID's op verschillende wijze worden gerapporteerd. Het is onder andere mogelijk om:

- een pré-match uit te voeren op meerdere pseudoniemen of een combinatie daarvan;
- te rapporteren over hits (is een individu bekend bij meerdere databronnen);
- te rapporteren over matches (bij welke databronnen is een hit bekend);
- te rapporteren op basis van een beslisboom (om te bepalen met welk pseudoniem of welke pseudoniemen de hoogste matchingsgraad wordt bereikt).

Rapportage

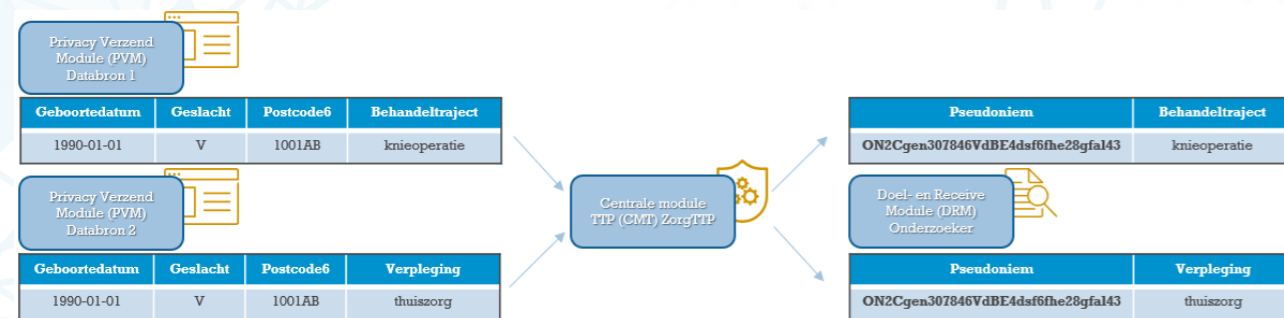
De matchingspercentages en matchende ID's worden via beveiligde bestandsoverdracht teruggekoppeld aan de databronnen. Op basis van het voorbeeld wordt een matchingspercentage van 33% gerapporteerd.



Fase 2: Verrijkingfase

Aanlevering aan onderzoekende partij

Op basis van de teruggekoppelde matchingspercentages en ID's uit de pré-matchfase kunnen de databronnen nu het aanleverbestand voor de onderzoeker opbouwen, ditmaal enkel voor de matchende individuen en inclusief de onderzoeksdata. De onderzoeker kan nu op basis van de pseudoniemen in de twee aangeleverde bestanden de onderzoeksdata aan elkaar koppelen.



ZorgTTP koppelservice

Veilig, eenvoudig en snel kijken naar overlap

In 2021 heeft ZorgTTP de opdracht gekregen om een koppelservice te ontwikkelen. Met de koppelservice kan een onderzoeker veilig, eenvoudig en snel nagaan of de patiënten uit een cohort bekend zijn bij een aantal grote Nederlandse registraties.

Wat is de koppelservice?

De koppelservice is een platform waarop opdrachtgevers en onderzoekers gepseudonimiseerde persoonsgegevens uitwisselen. Met de koppelservice zie je nu veilig, eenvoudig en volledig geautomatiseerd in welke aangesloten registraties de patiënten vanuit een specifiek cohort bekend zijn, maar ook welke individuen in zowel de ene als een andere registratie voorkomen. Voorheen moest de onderzoeker de data waarop hij onderzoek wilde doen bij elke registratie opnieuw laten matchen. Omdat de koppelservice werkt met pseudoniemen, zijn gegevens niet herleidbaar naar personen.

Wat is het voordeel voor de onderzoeker?

Dat is het geautomatiseerd nagegaan in welke aangesloten registraties patiënten bekend zijn, gecombineerd met het bepalen welke individuen in meerdere registraties voorkomen. Hiermee weet een onderzoeker snel of er informatie over deze mensen bekend is. Hij kan zo makkelijker bepalen of het waarde heeft überhaupt dat onderzoek uit te voeren. Wellicht is er onvoldoende data om de onderzoeksvraag te beantwoorden en moet de onderzoeksvraag worden aangepast. Is er voldoende basis om een onderzoek te doen, dan ondersteunen wij als ZorgTTP ook de daadwerkelijke uitwisseling van de informatie.

Wat is het voordeel voor de opdrachtgever?

Onze opdrachtgevers zijn de partijen die de data verzamelen. Ze stoppen veel tijd en middelen in het verzamelen van gegevens en willen de data op een veilige manier ter beschikking stellen voor onderzoek. ZorgTTP ondersteunt opdrachtgevers door het koppelproces te automatiseren, de pseudoniemen veilig op te slaan en de gegevens goed te beschermen. Zo hoeft de registratie niet zoveel moeite te doen om data-uitwisseling te faciliteren. Ze komen pas in beeld als er medisch inhoudelijke data geleverd moet worden.

Beveiliging

Veiligheid vinden we heel belangrijk. Ook bij de koppelservice. Daarom slaan wij geen medisch inhoudelijke data op en ook geen direct herleidbare persoonsgegevens. Wij werken alleen met onomkeerbare pseudoniemen. Dat betekent dat we de persoonsgegevens waarop dat pseudoniem gebaseerd is niet kunnen achterhalen. Daarnaast slaan we de data van opdrachtgevers niet op in dezelfde database. Gegevens van verschillende opdrachtgevers worden los van elkaar opgeslagen in aparte pseudoniemen kluisen: secure pseudonym stores (SPS).

Contracten

Onze werkzaamheden bestaan nu uit het standaardiseren en automatiseren van het proces. We werken toe naar een situatie waarbij een onderzoeker eenvoudig van alle registraties toestemming krijgt om gebruik te maken van de data uit de stores. In de toekomst kunnen we opdrachtgevers en onderzoekers ook helpen bij de administratieve afhandeling van het proces. Zodat een onderzoeker te maken krijgt met dezelfde set afspraken. Waarbij een onderzoek niet bij elke registratie afzonderlijk moet worden getoetst. Voor de onderzoeker en opdrachtgever is het prettig als alle administratie rondom zo'n onderzoek bij hetzelfde 'loket' afgehandeld kan worden.



De privacybeschermer bij datasamenwerking

ZorgTTP dankt haar opdrachtgevers voor het vertrouwen



Sinds 2007

ADVIES@ZORGTTP.NL

WWW.ZORGTTP.NL



ZorgTTP
Privacy & vertrouwen

Locatie:

Het Spoorwegmuseum
Maliebaanstation 16
3581 XW Utrecht

Meer informatie?

Heeft u vragen over onze dienstverlening neem dan contact op met een van onze adviseurs.

e-mail

advies@zorgttp.nl

internet

<https://www.zorgttp.nl>

bel

030-63 606 49