Haalbaarheidsstudie Nationale TTP-dienst

Inventarisatie rapport

Deze inventarisatie is uitgevoerd door SURFsara in opdracht van NFU Data4LifeSciences, i.s.m. het Mondriaan project.

Mark Hoevers (SURFsara)
Ton Verschuren (SURFsara)
Irene Nooren (SURFsara)
Jan Jurjen Uitterdijk (UMCG/Mondriaan)
Rob Bieringa (UMCG/Mondriaan))

Inhoudsopgave

1. Management samenvatting

Het inventarisatie traject als eerste stap in de haalbaarheid van een Nationale TTP-dienst is uitgevoerd in opdracht van het NFU Data4LifeSciences programma, i.s.m. project Mondriaan en SURFsara en moet inzicht geven in de mogelijkheden voor een nationale onderzoeks-TTP-dienst. Hierin is geïnventariseerd hoe UMC's en RIVM op dit moment te werk gaan bij het pseudonimiseren en koppelen van onderzoeksdata en is onderzocht of er ruimte is om dit op nationaal niveau beter te faciliteren dan wel een nationale TTP dienst in te richten.

Goed georganiseerd research datamanagement biedt kansen voor (hogere kwaliteit van) onderzoek. hergebruik van data en transparantie zijn daarin sleutelwoorden. Voor het hergebruik en analyse van medische data is het uit privacyoverwegingen nodig de persoonsgegevens, die bij die medische data horen, te pseudonimiseren. Het koppelen van medische data ten behoeve van onderzoek kan nieuwe wetenschappelijke inzichten opleveren en zo de zorg verbeteren.

Binnen de Nederlandse universitaire onderzoek-infrastructuur zijn op dit moment slechts beperkte oplossingen voor pseudonimisatie en koppelen aanwezig. Die situatie is suboptimaal: de huidige voorzieningen zijn kwetsbaar omdat ze niet structureel zijn belegd. Daarnaast zijn ze relatief duur wat ten koste gaat van onderzoeksbudgetten en de aangeboden functionaliteiten sluiten onvoldoende aan op de vragen van onderzoekers. De vraag is derhalve aan de orde op welke wijze die infrastructuur geoptimaliseerd kan worden. De haalbaarheidsstudie heeft een nationale focus; in een later stadium wordt de internationale component echter ook relevant.

Uit een interviewronde langs de UMCs blijkt dat UMC onderzoekers het niet aandurven om met andere partijen researchdata te koppelen, omdat dat ingewikkeld en duur is. Er is sprake van een latente behoefte, die niet/moeilijk te kwantificeren is. Wel is duidelijk geworden dat wanneer er op nationaal niveau TTP-functionaliteiten, tegen een betaalbaar tarief, beschikbaar zijn, daarvan zeker gebruik gemaakt zal worden. Het starten van een vervolg traject om tot een nationale TTP dienst te komen is dan ook aan te bevelen.

In het kader van een business model is het belangrijk om bij de vormgeving van een nationale TTP-dienst zorgvuldig te bepalen wat de rollen zijn van de diverse stakeholders, zoals: het programma NFU Data4LifeSciences, het CTMM-TraIT-project, commerciële TTP-dienstverleners, SURF, Mondriaan project, leveranciers van brongegevens, gebruikers/onderzoekers en toezichthouders.

Desgevraagd naar het gewenste business model geven de meeste UMC's aan dat er behoefte is aan zowel een regiefunctie (kennis delen en standaardisatie), als aan operationele dienstverlening (begeleiding bij het daadwerkelijk koppelen). De voorkeur gaat uit naar een governance model waarin één nationale dienst voorziet in deze diensten met borging van standaardisatie in pseudonimisatie voor koppeling.

De wens van de onderzoeker is een nationale TTP-dienst als nutsvoorziening in te richten, zodat duur onderzoeksgeld aan onderzoek en niet aan versleuteling besteed kan worden. De mogelijkheid tot nationale financiering van een dergelijke dienst zal verder onderzocht moeten worden, met de daarbij behorende geldstroom. Andere mogelijkheden zijn een dergelijke dienst te financieren uit 'algemene middelen' van het NFU en/of de UMC's,, of wel via onderzoeksbudgetten als hiervoor door onderzoek financiers gebudgetteerd kan worden.

Op basis van de analyse van de vraagkant (UMC's) komen de onderzoekers tot de volgende aanbevelingen:

a) Faseren: richt een Special Interest Group in, waarin de belangrijkste stakeholders samenwerken aan een ontwikkelmodel voor nationale TTP-dienstverlening;
b) Ambiëren: regel uitsluitend die zaken op nationaal niveau, die niet of niet efficiënt op decentraal niveau uitgevoerd kunnen worden;
c) Prioriteren: zoek met stakeholders naar pilots met meerwaarde voor het onderzoeksveld, zoals bijvoorbeeld het ontwikkelen van standaarden;
d) Kennis delen: is wellicht de eerste prioriteit om partijen te overtuigen van het belang van verantwoord datahergebruik en de noodzaak dat op nationaal niveau te organiseren;

e) Financieren: neem de ontwikkeling van TTP-dienstverlening op nationaal niveau op in het programma NFU Data4LifeSciences; het zal ten goede komen aan de gehele onderzoekspopulatie in de sector hoger onderwijs en onderzoek.

Als eerste stap voorwaarts stellen we de oprichting van een SIG, kennisplatform die zich richt op de volgende thema's:

- veilige uitwisseling en koppeling van persoonsgebonden data (b.v. zorg, biomedisch, sociaal onderzoek),
- standaardisatie in koppeling van data, met oog op internationaal beleid; het ontbreken van standaarden is de hoofdoorzaak van de huidige inefficiënte situatie,
- governance, financiële en juridische aspecten m.b.t. toegang tot gekoppelde data (het uitwerken van de juridische kaders valt buiten de scope van de SIG en hiervoor is afstemming met COREON.

Uitkomsten van dit kennisplatform kunnen via het Data4lifesciences programma (WP6: Good Research Practice) verder worden uitgewerkt in praktische oplossingen die middels outreach bij de UMC's toepasbaar zijn. Dit feitelijke werk bestaat dan uit het faseren, prioriteren, kennisdelen en de begeleiding van de implementatie.

2. Inleiding

Aanleiding

Het inventarisatie traject als eerste stap in de haalbaarheid van een Nationale TTP-dienst is uitgevoerd in opdracht van het NFU Data4LifeSciences programma, i.s.m. project Mondriaan en SURFsara en moet inzicht geven in de mogelijkheden voor een nationale onderzoeks-TTP-dienst. Deze zou op termijn verbreed kunnen worden naar samenwerkende topklinische opleidingsziekenhuizen (STZ) en huisartsenpraktijken en andere bronnen met relevante onderzoeksdata, zoals het CBS en GGD's. Het betreft hier de haalbaarheid van een pseudonimisatie dienst als wel een dienst voor het realiseren van koppelingen tussen research datasets. Dit ter verbetering van de lokale en nationale situatie rond pseudonimiseren en koppelen van patiënt data, waarmee beschikbaarheid en hergebruik van data voor onderzoek wordt bevorderd.

De opdracht is in een nauwe samenwerking tussen het NFU programma Data4Lifesciences en het Mondriaan project uitgevoerd. Dit project heeft als doel de realisatie van een landelijke data infrastructuur waarmee reguliere zorgregistraties en specifieke onderzoeks-cohorten en – populaties op een veilige wijze op persoonsniveau gekoppeld worden zodat verrijkte databases ontstaan. Deze verrijkte databases worden in eerste instantie gebruikt voor wetenschappelijk onderzoek.

Aanpak

In 2015 is, onder begeleiding van SURFsara, een werkgroep opgezet om de problematiek en mogelijke oplossingen in kaart te brengen. Ter verkenning zijn een aantal bijeenkomsten met vertegenwoordigers van UMC's, TTP-dienstverleners en NEN georganiseerd. Hierin zijn diverse functionele specificaties, use-case modellen en governance modellen besproken.

Daarna is gestart met de inventarisatieronde en zijn interviewsessies gehouden met vertegenwoordigers vanuit alle UMC's en het RIVM. De interviews namen gemiddeld anderhalf uur in beslag en werden gehouden met zowel onderzoekers als informatiemanagers, die onderzoekers begeleiden bij hun zoektocht naar de juiste TTP-functionaliteiten.

Op deze wijze heeft de werkgroep proberen te achterhalen wie relevante spelers in het onderzoeksveld zijn, welke normenkaders voor pseudonimisatie en het koppelen van researchdata bestaan en welke mogelijke business modellen voor een nationale TTP-dienst kunnen zijn. Ook is geïnventariseerd op welke wijze UMC's en het RIVM op dit moment te werk gaan bij het uitwisselen en koppelen van researchdata , welke TTP faciliteiten daarvoor gebruikt worden en wat de wensen en verwachtingen zijn m.b.t. de vormgeving van een nationale TTP-dienst voor de gehele sector.

Dit rapport beschrijft de uitkomsten van de inventarisatieronde en een advies over het vervolgstappen in het haalbaarheidsonderzoek. Hoofdstuk 3 gaat in op de probleemstelling en het mogelijke bestaansrecht van de Nationale TTP-dienst. Hoofdstuk 4 geeft een beschrijving van de uitkomsten van de intervies en een overzicht van de randvoorwaarden waaraan een Nationale TTP-dienst zou moeten voldoen. Tot slot geeft hoofdstuk 5 een eerste aanzet van de scope van een toekomstige voorziening met verschillende opties tot vormgeving daarvan,

3. Wat is het probleem

Steeds vaker worden patiënt data uit het zorgdomein gebruikt voor wetenschappelijk onderzoek waarbij gegevens van meerdere bronnen gekoppeld moeten worden. Patiënten die daar mee instemmen (middels een informed consent of geen bezwaar procedure) gaan er van uit dat de instellingen, die hun data hebben verzameld, daar uiterst zorgvuldig mee omgaan en kunnen verantwoorden waar hun data voor wordt gebruikt. er .
Privacybescherming is daarbij van het grootste belang. Bij het koppelen van onderzoeksdata is dan ook coderen, pseudonimiseren[1] en (de-)anonimisatie met een Trusted Third Party (TTP) vereist.

Op dit moment zijn er enkele commerciële TTP-dienstverleners actief in Nederland. Deze zijn destijds voor één bepaalde toepassing of project ontwikkeld (bijvoorbeeld een weefselbank t.b.v. pathologisch onderzoek), maar niet structureel belegd; een sustainable TTP-infrastructuur en eenduidig pseudonimisatie proces ontbreekt. Onderzoekers zullen geneigd zijn naar alternatieve pseudonimisatiediensten te zoeken, die wellicht goedkoper zijn (de kosten van codering en pseudonimisatie worden gezien als overhead, die niet bijdraagt aan de onderzoeksresultaten). Bovendien biedt niet elke TTP op dit moment de gewenste functionaliteit, die nodig kan zijn voor een specifieke aanvraag.

Daarnaast is het voor hergebruik van medische data noodzakelijk om deze te koppelen met andere gegevens van die patiënt, zodat verder onderzoek uitgevoerd kan worden. Doordat de verschillende TTP's niet dezelfde standaarden gebruiken is koppelen vaak niet mogelijk en moet eerst het pseudoniem worden verwijderd (decryptie) om vervolgens opnieuw te pseudonimiseren met de standaard van de TTP-dienst die over de koppeldata beschikt. Hierna is het pas mogelijk om de koppeling met succes te realiseren. Deze gang van zaken is ronduit inefficiënt.

Dit betekent dat er behoefte is aan uniforme standaarden en ondersteuning van de onderzoeker bij het bepalen van de juiste en betaalbare TTP-functionaliteit. Redenen genoeg om onderzoek te doen naar de haalbaarheid van een nationale TTP-functionaliteit, die als een nutsvoorziening beschouwd kan worden voor UMC's en een landelijke regiefunctie hierop kan uitoefenen.

[1] De begrippen pseudonimiseren en pseudonimisatie worden doorgaans beide gebruikt.

4. Wat is een TTP-dienst

Begrippen

In het inventarisatietraject werd al snel duidelijk er verwarring is over de begrippen die gebruikt worden in het domein. Die verwarring zit in het gebruik van het begrip TTP en met name welke functionaliteit deze biedt.

Strikt genomen maken we onderscheid tussen TTP's en pseudonimisatie-dienstverleners. Voor de pseudonimisatiedienstverlener wordt vaak ook de term TTP gehanteerd en dat leidt tot verwarring. Een TTP doet niets meer en niets minder dan het sleutelmanagement (sleutel-generatie, -opslag, beheer en –vernietiging), terwijl de pseudonimisatiedienstverlener de versleutelde data beheert. Deze laatste functie wordt ook door een UMC of landelijke organisaties, zoals PALGA, BBMR-NL of IKNL, uitgevoerd. Zie ook de toelichting op pseudonimisatie en versleuteling processen in de volgende paragraaf.

Een aantal UMC's doen zowel sleutelbeheer als opslag van versleutelde data. Dat kan alleen onder strikte scheiding van beide rollen/functionaliteiten.

We hanteren de volgende begrippen:

- **Geanonimiseerde gegevens:** Gegevens van de persoon die niet door de ontvanger van die gegevens tot de persoon kunnen worden herleid
- **Pseudoniem:** Een persoonsidentificatie die verschilt van de gebruikelijk gehanteerde persoonsidentificatie, waardoor de persoon voor derden niet meer te identificeren is (pre-pseudoniem of pseudo-code)
- **Pseudonimisatie:** Bijzondere vorm van anonimiseren, waarbij de verbinding tussen een set identificerende gegevens en het datasubject wordt verwijderd, en een nieuwe verbinding wordt gemaakt tussen een bepaalde set van karakteristieken die verwijzen naar het datasubject en een of meerdere pseudoniemen
- **Versleuteling of encryptie**: een techniek waarbij informatie (softwarematig) onleesbaar voor onbevoegden wordt gemaakt. Voor het coderen en decoderen van de informatie is een speciale encryptie-sleutel (algoritme of vertaaltabellen) nodig.
- **Trusted third party (TTP):** Een vertrouwde onafhankelijke partij die diensten aanbiedt die de betrouwbaarheid van elektronische gegevensuitwisseling en gegevensopslag vergroten

In Bijlage2 is een volledige begrippenlijst volgens ISO 25237 (Health informatics – Pseudonymization) beschreven.

Juridische kader

We onderscheiden op hoofdlijnen twee soorten data met eigen wettelijke regimes:

- Persoonsgebonden data,
- Niet persoonsgebonden data.

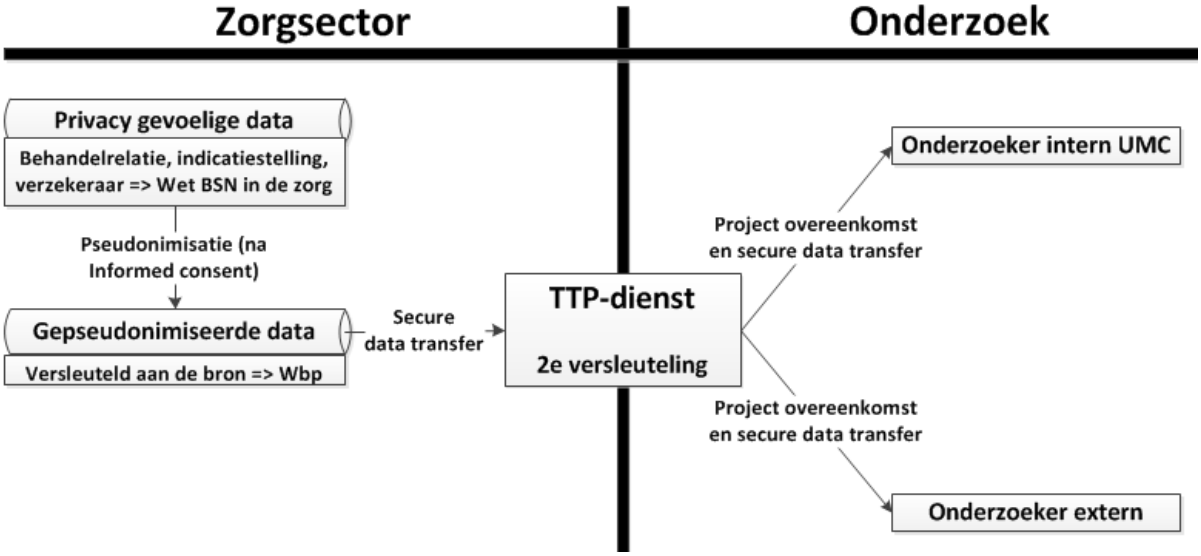
Persoonsgebonden of privacy gevoelige data die in de zorgsector gebruikt worden in het kader van een behandelrelatie vallen onder de Wet 'BSN in de zorg' bij de uitwisseling van deze data tussen behandelaars (waaronder ook het UMC-ziekenhuis), indicatiestellers en verzekeraars.
Zodra privacy gevoelige data buiten het zorgdomein gebruikt gaan worden, bijvoorbeeld voor nader onderzoek (bijv. onderzoeksinstituten buiten hun eigen UMC), dan is de Wet Bescherming Persoonsgegevens (WBP) van toepassing. De WBP eist dat de privacy gevoelige data worden gepseudonimiseerd aan de bron (bij de eerste behandelaar). Een tweede versleutelingmechanisme is optioneel om de relatie tussen de patiënt en zijn medische gegevens te verbreken, maar niet wettelijk verplicht.

Wanneer een onderzoeker geen behandelrelatie heeft met een patiënt, zijn er twee scenario's mogelijk waarmee de onderzoeker volgens wettelijk kader toestemming kan verkrijgen voor het gebruik van data voor onderzoeksdoeleinden:

- Wanneer de patiënt toestemming geeft via een informed consent; de onderzoeksdoeleinden zijn hierin gespecificeerd.

- De gedragscode Gezondheidsonderzoek, ook wel de 'Code Goed Gedrag' genoemd; biedt gezondheidsonderzoekers binnen de grenzen van de wet een duidelijk kader, zodat zorgvuldig wordt afgewogen wat nog wel kan en wat niet.

Onderstaande schets geeft een toelichting op deze situatie.



Figuur 1: versleutelen van persoonsgebonden data voor onderzoeksdoeleinden.

Voor (de uitwisseling van) niet persoonsgebonden data gelden veel minder strengere regels.

Het Autoriteit Persoonsgegevens (voorheen College Bescherming Persoonsgegevens (CBP)) is in Nederland bij wet als toezichthouder aangesteld voor het toezicht op het verwerken van persoonsgegevens. De taken van de AP worden bepaald door de Europese Privacyrichtlijn die voor alle landen van de EU geldt.

Pseudonimisatie

In die gevallen waar het gebruik van persoonsgegevens niet is toegestaan, maar wel behoefte bestaat aan (medische) data van unieke personen, kan het gebruik van *pseudoniemen* uitkomst bieden. Onder pseudonimiseren wordt verstaan het omzetten van persoonsgegevens naar een niet tot de oorspronkelijke persoon herleidbare unieke code. Persoonsgegevens betreffen typisch naam, adres- en woongegevens (NAW) en geboortedatum. Maar ook de combinatie van een zeldzame ziekte en regio, leeftijdscategorie of lidmaatschap van een bepaalde voetbalclub kan de kans op herleidbaarheid van een persoon verhogen.

Het omzetten van persoonsgegevens naar een niet tot de oorspronkelijke persoon herleidbare unieke code (pseudoniem) verloopt, in het meest ideale geval in een aantal stappen, zie ook figuur 2.

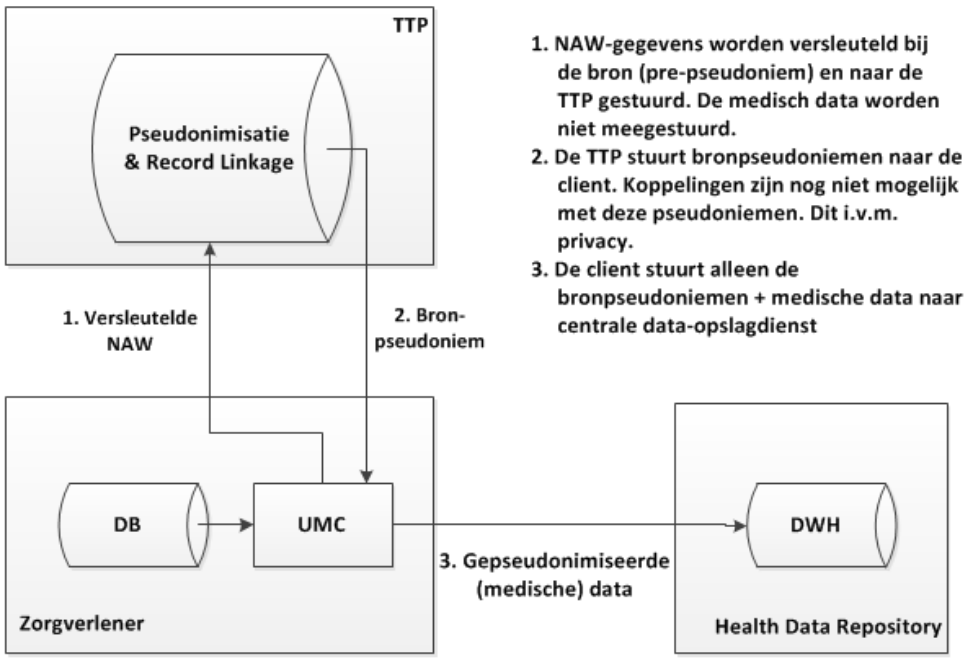
1. Versleuteling van de persoonsgegevens aan de pre-pseudoniem (conform 1e vereiste CBP/AP). Daarbij wordt een scheiding aangebracht tussen de persoonsgegevens en de medische gegevens (die niet naar de TTP worden verstuurd). Het UMC beschikt na deze versleuteling over het bronpseudoniem en de medische data van de persoon. Beiden worden beveiligd opgeslagen bij het UMC en/of een centrale voorziening, zoals bijvoorbeeld een biobank. Overigens is deze stap niet altijd vanzelfsprekend en zal in de meeste gevallen gelijk stap 2 worden uitgevoerd.
2. Door de TTP wordt een tweede versleuteling aangebracht op het pre-pseudoniem, die geheim is voor zowel de aanbieder van de data, als de latere ontvangende partij. Hiermee is de relatie tussen pseudoniem en persoonsgegevens verbroken en is het in principe niet langer mogelijk om van het aangemaakte pseudoniem terug te gaan naar de oorspronkelijke persoon.

3. Pseudoniemen en (medische) data worden in een centrale data repository opgeslagen. Deze (centrale) data repository wordt beheerd door het UMC of een sectorale databank. De TTP heeft hier geen toegang toe.

N.B.: transport van data dient beveiligd plaats te vinden.

Met bovenstaande stappen is het voor zowel de bron als de ontvangende partij niet meer mogelijk om de oorspronkelijke persoonsgegevens uit de bron in verband te brengen met het aangemaakte pseudoniem. Zie ook Figuur 2.

Vereiste is dat de opslag van data en de versleuteling twee gescheiden systemen zijn die onafhankelijk van elkaar opereren en elkaar niet kunnen manipuleren (zie daarvoor bijvoorbeeld ISO 25237, pag. 30).



Figuur 2: Pseudonimisatie schematisch wedergegeven.
Bron: Presentatie Willem de Bruin, Mondriaan project, juli 2014

Record linkage wordt toegepast wanneer er meerdere onderzoeken worden uitgevoerd op dezelfde patiënt en het noodzakelijk is om in het pseudoniem van zijn persoonsgegevens telkens aan te geven dat het om diezelfde patiënt gaat (m.b.v. een codering). Als dat correct gebeurt kunnen de gepseudonimiseerde (medische) data steeds tot diezelfde patiënt herleid worden (zonder te weten wie het is). Op die manier is het mogelijk om het medische traject dat die patiënt in de tijd doorlopen heeft zichtbaar te maken.

Deze stappen komen overeen met de in ISO 25237 gehanteerde entiteiten (zie bijlage 1):

1. Data source: bron
2. Person identification service: 1e versleuteling tot pre-pseudoniem
3. Pseudonymization service: definitieve pseudonimisatie, 2de versleuteling
4. Data target: centrale opslag.

Koppelen

De databases waarin de pseudoniemen en bijbehorende medische data zijn opgeslagen kunnen zich bevinden bij de individuele UMC's, of bij centrale sectorale of nationale partijen, bijvoorbeeld de beheerders van catalogi met categorieën van medische gegevens (Pathologisch-Anatomisch Landelijk

Geautomatiseerd Archief (PALGA), Biobanking and Biomolecular Research Infrastructure for The Netherlands (BBMRI-NL)), IKNL, CBS, NIVEL, e.d..
Onderzoekers kunnen uit zo'n catalogus een keuze maken van datasets, die ze zouden willen koppelen aan bijvoorbeeld een eigen verzamelde dataset.

Er bestaan ruwweg twee methoden voor het koppelen van gepseudonimiseerde registratiebestanden: een deterministische koppeling en een probabilistische koppeling. In beide methoden worden records uit de te koppelen bestanden vergeleken op een set van koppelvariabelen en wordt een beslissing genomen welke paren bij elkaar horen.

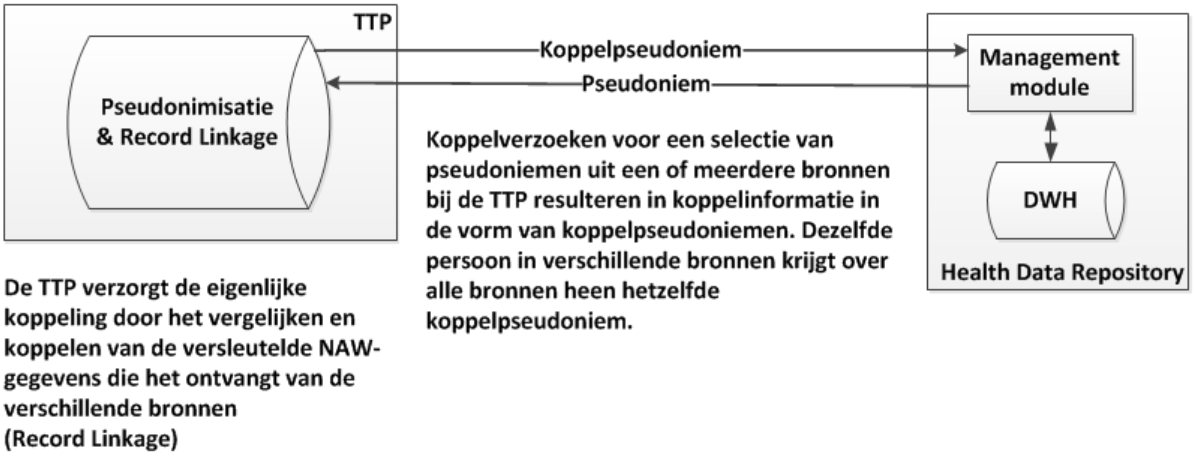
Deterministische koppeling

Bij een deterministische (of 'exacte') koppeling wordt gekeken naar overeenkomst op de koppelvariabelen en op basis van het patroon van overeenkomst wordt besloten welke paren worden meegenomen als koppeling. Bij overeenkomst op de volledige koppelsleutel (derde versleuteling) spreekt men van een 'full' koppeling. Wanneer een verschil op 1 of meer koppelvariabelen wordt geaccepteerd spreekt men van een N-1 of N-2 koppeling.

Probabilistische koppeling

Bij probabilistisch koppelen wordt gebruik gemaakt van een kansmodel (Fellegi en Sunter) op basis waarvan een koppeling van twee records al dan niet wordt gemaakt. Een koppelgewicht wordt toegekend op basis van de waarschijnlijkheid van overeenkomst van de variabele, onder recordparen die in werkelijkheid bij elkaar horen en onder recordparen die in werkelijkheid tot verschillende personen behoren. Deze kansen (waarschijnlijkheden) worden geschat met behulp van een statistische methode op basis van de waargenomen patronen van overeenkomst en verschil op de koppelvariabelen. Een overeenkomst op geboortedatum geeft bijvoorbeeld meer bewijs dat twee records bij elkaar horen dan een overeenkomst op geslacht (dit is op basis van toeval al 50%). De gewichten van de afzonderlijke koppelvariabelen worden bij elkaar opgeteld en dit gewicht is dus een maat voor de mate van waarschijnlijkheid dat twee records bij elkaar horen. Recordparen met een heel laag (negatief) koppelgewicht horen niet bij elkaar en recordparen met een heel hoog koppelgewicht horen bij elkaar.

Parelsnoer koppelt deterministisch (m.b.v. ZorgTTP). Mondriaan heeft gekozen voor een hybride oplossing. Er wordt deterministisch gekoppeld bij aanwezigheid van een uniek identificerend veld, zoals een BSN-nummer. In alle andere gevallen wordt er probabilistisch gekoppeld (m.b.v. Custodix e.a.).



Figuur 3: koppelen op persoonsniveau via een TTP.
Bron: Presentatie Willem de Bruin, Mondriaan project, juli 2014

Een verzoek tot koppelen van datasets kan worden ingediend bij de beheerder van de betreffende data. De eigenlijke koppeling wordt verzorgd door een TTP-dienst, door het vergelijken en koppelen van de versleutelde kenmerken, zoals bijvoorbeeld NAW-gegevens, achternaam, postcode of verzekeraar.

Trusted Third Party (TTP)

De TTP-dienst is de enige partij die weet op welke wijze het pseudoniem is aangemaakt. We zien dat trusted third parties (TTP diensten) een belangrijke rol spelen in het faciliteren van het pseudonimisatie- en koppelproces. De TTP dienst is in het bezit van de vertaaltabellen tussen pre-pseudoniem en pseudoniem, en heeft koppelpseudoniemen ter beschikking van specifiek gekoppelde data.

Maar daaraan zijn wel degelijk ook beperkingen gesteld. De TTP speelt strikt genomen uitsluitend een rol in het sleutelmanagement (sleutelgeneratie, -opslag, -beheer en –vernietiging) en heeft geen bemoeienis met de data (persoonsgegevens) die versleuteld worden. De TTP en pseudonimisatie-dienstverlener zijn twee gescheiden entiteiten.

Wanneer een data leverancier (UMC) zelf een tool gebruikt voor de pseudonimisatie aan de bron, dan moet de afdeling waar dat gebeurt technisch en organisatorisch gescheiden zijn van het zorgdomein, zodat er geen toegang is tot de patiënt data.
De versleutelde medische data worden dus nooit bij een TTP opgeslagen. Dit onderscheid wordt (te) vaak niet helder gemaakt, waardoor sprake kan zijn van begrips- en spraakverwarring. In bijlage 2 zijn de belangrijkste begrippen uit ISO 25237 opgenomen.

Nadere uitleg over soorten privacy-risico's in het pseudonimisatieproces wordt gegeven in bijlage 3. De vereisten aan de beveiliging van processen en systemen zijn uit ISO 25237 overgenomen in bijlage 4.

5. Uitgangspunten TTP-dienst

Dit hoofdstuk beschrijft de uitkomsten van de gevoerde interviews bij de UMC's en RIVM. Vanuit elk instituut zijn gemiddeld zo'n 2-3 personen bij de sessies betrokken geweest. Hierbij is, voor zover mogelijk, rekening gehouden met een afvaardiging vanuit meer dan één afdeling/programma per instituut.
Over het algemeen is de diversiteit groot (afdelingen in een UMC en UMC overstijgende) inzake de huidige invulling van pseudonimisatie en koppelvoorzieningen. Sommige afdelingen pseudonimiseren handmatig (vervanging van gegevens in Excel), terwijl andere – vaak grotere – afdelingen of programma's dit via een commerciële TTP laten verzorgen.

Processen

Naar mate van professionaliteit en reikwijdte onderscheiden we drie categorieën voor de uitvoering van pseudonimisatie en koppelingen:

De ***eerste categorie*** betreft UMC's, die bij het pseudonimiseren gebruik maken van professionele *externe* partijen en ook daadwerkelijk datasets koppelen. De 1ste versleuteling (aan de bron) wordt in sommige gevallen intern gedaan. De 2de versleuteling wordt altijd door een TTP dienst gedaan.
Dit betreft:

- het LUMC (ZorgTTP, alle parels in Parelsnoer);
- het MUMC (MEMIC en Custodix, koppelen met Movare, GGD, CBS en RNH);
- het RIVM, voor m.n. het koppelen van het datawarehouse met IKNL, PALGA en DICA (ZorgTTP). Waar RIVM optreedt als dataleverancier hebben ze meestal te maken met een TTP die optreedt namens de partij die van het RIVM bepaalde datasets afneemt. Bij deelname aan bevolkingsonderzoek bij RIVM mag, omdat sprake is van een behandelrelatie, gekoppeld worden via het BSN.

Een ***tweede categorie*** wordt gevormd door de UMC's die zelf *interne* voorzieningen hebben getroffen voor het pseudonimiseren en het koppelen met andere onderzoekers, maar in sommige situaties ook gebruik maken van *externe* partijen.
Dit zijn:

- het UMCG, Universitair Centrum Psychiatrie (eigen software + handmatige koppelingen) en Oncologie (via Trial Coordination Center (TCC) en ZorgTTP). Met GGZ Noord Nederland en DICA wordt gekoppeld;
- het UMCU (eigen software voor pseudonimisatie + zelf intern koppelen). In geval van polymerase chain reaction (PCR), gebruikt in DNA onderzoek, wordt extern gekoppeld. Gebruik interne bronnen door derden wordt afgehouden, vaak door gebrek aan juridische kennis m.b.t. wet-en regelgeving medische data;
- het AMC (Parelsnoer, dus ZorgTTP), voor de Intensive Care koppelingen met Vektis, MARS en Achmea; verder wordt niet gekoppeld vanwege privacy.

De ***laatste categorie*** bestaat uit UMC's die zelf niet koppelen, maar wel gepseudonimiseerde datasets uitwisselen (zowel intern als extern). Zij hebben daarvoor vaak zelf een tool ontwikkeld of pseudonimiseren handmatig (postcode 4, e.d.).
Het gaat om:

- het ErasmusMC (1e versleuteling/prepseudonimisatie handmatig, 2e versleuteling/pseudonimisatie met industriële software (Aquarius net). Bij MammoXL koppelen afnemende partijen met ErasmusMC. Verder wordt niet gekoppeld, maar uitgewisseld;
- het VUmc (DICOM voor imaging). Externe uitwisseling (géén koppeling) voor Alzheimer en MS in heel Europa, alsmede enkele parels in Parelsnoer;
- het RadboudMC (gebruiken een mini 1-weg hash om te anonimiseren). Er wordt niet gekoppeld (2-weg hash, te ingewikkeld en duur).

Overigens is de laatste categorie, vanuit het perspectief van de onderzoeker, een meest voor de hand liggende oplossing en . zowel van toepassing bij singlecenter als multicenter studies. Een externe TTP-dienst zal pas worden ingezet bij het koppelen van een research dataset aan grote registraties zoals Lifelines, IKNL, CBS etc. Vanuit 'Good Research Practice' verdient natuurlijk de eerste of tweede categorie de voorkeur om de onafhankelijkheid op een netjes manier te kunnen waarborgen.

Kennis

Een grote diversiteit werd ook waargenomen in de kennis van onderzoekers over pseudonimisatie en koppeling van data. In veel gevallen waren onderzoekers niet bekend met de mogelijkheden van het koppelen van data. Dit kan verschillende redenen hebben:

- kennis is (nog) niet nodig voor het onderzoek waar de onderzoeker bij betrokken is
- disseminatie van kennis over processen is onvoldoende bij de onderzoeker bekend. Het proces van de verspreiding van kennis is onvoldoende, of de onderzoeker weet niet waar hij kennis kan halen.

De behoefte aan pseudonimisatie is binnen de UMC's vaak een latente behoefte. Deze behoefte is met name gerelateerd aan multidisciplinair en longitudinaal onderzoek, bijv. waarbij data vanuit verschillende ziektebeelden aan elkaar gekoppeld wordt.

Kosten

De kosten voor pseudonimisatie en TTP dienstverlening worden door afzonderlijke onderzoekers over het algemeen hoog bevonden. Dit is één van de redenen waardoor geen gebruik wordt gemaakt van de diensten en daarmee verrijking van onderzoek wordt ingeperkt. Met een centrale faciliteit zou de TTP dienstverlenging wellicht goedkoper georganiseerd kunnen worden.

De minst gevorderde UMC's maken gebruik van een gratis tool van een bevriende leverancier, dan wel van interne diensten waarvan de kosten intern niet doorbelast worden.

Bij ZorgTTP kost een jaarabonnement voor pseudonimiseren + koppelen ongeveer € 30.000, waarbij niet bekend is hoeveel koppelverzoeken hierbij per jaar toegestaan zijn. De kosten bij een commerciële TTP dienst worden door onderzoekers als zeer hoog ervaren. Tussen deze twee uitersten zien we dat er vaak met interne uurtarieven wordt gewerkt variërend tussen de € 25 en € 75.

Overigens wordt bij de grotere projecten die gesubsidieerd worden door de EU of ZonMW steeds meer gepleit voor het oormerken van een deel van de projectbegroting t.b.v. data stewardship. Eventuele TTP kosten zouden hier onderdeel van kunnen zijn.

Partner stakeholders

Bij een nationale TTP zijn de volgende stakeholders als partners momenteel aan de orde:

- Leveranciers brongegevens:
    - Ziekenhuizen: UMCs, STZ ziekenhuizen
    - Huisartsen
    - NIVEL, CBS
    - Palga, Promise
-
- NEN: Toezicht / regulaties informatievoorziening en privacy NL: Autoriteit Persoonsgegevens, NEN 7510.
- Mondriaan project, Lygature
- SURF
- NFU
- Commerciële TTP diensten zoals Custodix en ZorgTTP; wellicht in het adopteren van standaarden die bovenstaande partijen ontwikkelen.

Daarnaast de volgende projecten die helpen om de use cases in kaart te brengen, coördinatie te voeren en infrastructuur op te zetten

- Programma NFU Data4Lifesciences;
- CTMM-TraIT project: ontwikkeling nationale ICT diensten;
- BBMRI
- Dutch TechCentre for Life Sciences (DTL)

Op dit moment wordt in een NEN-werkgroep gesproken over standaardisatie. De voortgang ervan is momenteel onduidelijk.
Het veld bestaat momenteel uit verschillende commerciële TTP-aanbieders waarin het ontwikkelen van standaarden een nadelige invloed kan hebben op hun marktpositie.

Wie uiteindelijk betrokken moet zijn in de realisatie van een sustainable business model is afhankelijk van de gekozen propositie en governance structuur. Gezien de propositie is een nationale partij zoals SURF, NFU, DTL, Mondriaan project / Lygature of NEN, een aangewezen partij om de lead te nemen in het business model.

Markt

Onder de gebruikers van een nationale TTP dienst vallen de onderzoekers UMCs en wellicht andere onderzoeksinstituten. De coördinatie hiervan ligt bij nationale projecten zoals NFU Data4LifeSciences, BBMRI en TraIT, en/of contacten via SURF. Er is niet 1 partij die representatief is voor de groep van onderzoekers.

De behoefte aan een nationale TTP dienst bestaat bij UMCs met name uit de klinische onderzoeksgebieden waarbij klinische informatie uit verschillende bronnen geïntegreerd dient te worden :

- verschillende disciplines
- verschillende zorgverleners, bijv. bij longitudinaal onderzoek over een bepaalde tijd na verhuizing van een patiënt

Deze behoefte is, naast de UMCs, ook bij andere onderzoeksprogramma's en instellingen (STZ ziekenhuizen, huisartsen netwerken, NKI) aanwezig.

Propositie

Een standaardisering van pseudonimisatiediensten maakt het koppelen van verschillende onderzoeksdatasets eenvoudiger. Met name de kosteneffectiviteit om datasets onderling te koppelen en niet opnieuw dezelfde gegevens te moeten verzamelen kan een flinke besparing opleveren.

Het ontbreken van de standaardisering biedt een achterstand voor onderzoek in Nederland t.o.v. landen als Denemarken en UK.

Een nationale TTP zou moeten voorzien in een dienst waarmee door standaardisering hergebruik van data wordt bevorderd. Dit biedt een grote toegevoegde waarde voor onderzoek in Nederland, om daarmee een concurrerende positie ten opzichte van andere landen (Europees, wereldwijd) te kunnen handhaven. Nederland loopt voorop in biomedisch onderzoek. Het is helaas lastig te kwantificeren in hoeverre Nederland een verbeterde positie zou kunnen pakken door betere faciliteren van koppeling van data, omdat niet bekend is welke data exact beschikbaar is, is welke koppelingen mogelijk vernieuwende onderzoeksvragen kunnen beantwoorden.

In andere landen, zoals Denemarken, wordt patiënt data gekoppeld middels een nationaal identificatie nummer voor burgers, analoog aan het Nederlandse Burger Service Nummer (BSN). Organisaties zijn door de overheid bevoegd om gebruik te maken van het nationale burger identificatie nummer. Daarnaast zijn aanvragen mogelijk die door een ethische commissie beoordeeld worden. De cultuur in Denemarken is dat er meer vertrouwen in overheidsorganisaties bestaat, en daarmee koppeling middels een BSN mogelijk is.

De propositie voor een nationale TTP dienst is afhankelijk van de nationale keuze geen BSN te gebruiken voor het uitwisselen van patiënt data buiten het zorg domein. Een andere mogelijke route is tot politiek-nationale afspraken te komen voor het gebruik van BSN voor onderzoek, t.b.v. hergebruik van klinische en onderzoeksdata. De tijd tot implementatie zal voor deze route naar verwachting langer zijn.

Wensen, suggesties en aandachtspunten

In het traject voorafgaande aan de interview sessie zijn door de werkgroep een aantal requirements opgesteld die van toepassing zijn bij de inrichting en/of vormgeving van een nationale TTP(-infrastructuur). Deze zijn tijdens de verschillende interviewsessies geverifieerd.

- Onafhankelijkheid van aanbieders van pseudonimisatiediensten, dan wel scheiding tussen de TTP en de pseudonimisatiedienst;
- Centraal aanspreekpunt met kennis van zaken;
- Support naar onderzoekers in keuze TTP-functionaliteit / oplossing;
- Uniformiteit in de dienstverlening met mogelijkheid tot differentiatie naar behoefte;
- Aanbod van diverse functionaliteiten waaruit gekozen kan worden op basis van de *use case*;
- Kostenreductie, gebruik maken van schaalvoordelen en gezamenlijke inkoop;
- Gegarandeerde continuïteit / beschikbaarheid van de dienstverlening;
- Centraal volgen van, en waar mogelijk participeren in, de ontwikkelingen op gebied van standaarden, techniek en wet- & regelgeving, onder andere op het gebied van privacy, WBP, Algemene Verordening Gegevensbescherming van de EU, en het uniform vertalen van deze nieuwe ontwikkelingen naar de dienstverlening;
- Behouden van huidige TTP-services die in gebruik zijn bij de UMC's door samenwerking en/of overdracht;
- Regel alleen centraal wat niet (of niet goed(koop)) decentraal kan;
- Voldoe aan kwaliteitsnormen, o.a. ISO 25237 (w.o. para's 9.3 en 9.4), NEN-standaarden, ISO 27001 en wellicht later ook ETSI-normen voor grensoverschrijdende vertrouwensdiensten.

Tijdens de interviewfase konden UMC's hun wensen en verlangens m.b.t. een nationale TTP-dienst ventileren. Hieronder een overzicht van aanvullende requirements:

- Bied kennis en expertise aan om onderzoekers te kunnen adviseren en faciliteren: meest voorkomende soorten onderzoek, welke functionaliteiten passen daar het beste en goedkoopste bij. Op welke wijze kan (afhankelijk van type onderzoek) optimaal gepseudonimiseerd en gekoppeld worden (bijv. door gebruik standaarden);
- proefperiode voor gebruik TTP diensten door onderzoekers, zodat onderzoekers kunnen uitzoeken wat ze nodig hebben, voordat ze in een duur contract stappen;
- Uitsluitend centraal uitvoeren wat niet of niet goed en goedkoop decentraal kan;
- Data laten waar ze nu zijn: geen centrale database. Wel centraal register/catalogus: welke datasoorten (kenmerken) zijn waar verkrijgbaar;
- Groeimodel: start met 1 of 2 pilots voor zaken waaraan grote behoefte bestaat. Bouw zo de expertise geleidelijk op;
- Governance: je kunt wel beginnen onder de vlag van SURFsara of NFU, maar als er bij succes uitgebreid gaat worden naar andere ziekenhuizen, huisartsen, GGD's, CBS en andere data-eigenaren, kan dat dan wel onder vlag van SURFSara of NFU;
- Regiemodel op technische randvoorwaarden (standaarden), kennisdeling, uitvoeringsvraagstukken (interoperabiliteit, e.d.), catalogus, e.d.;
- Streef naar partnerships met bestaande TTP-aanbieders: dienstenmodel? Dit o.a. omdat marktverstoring voorkomen dient te worden;
- Streven naar gratis nutsmodel, waarin UMC's en andere partners een vaste jaarbijdrage leveren. Dit omdat er bij veel onderzoek geen budget is voor dit soort zaken, dus liever centraal binnen een UMC financieren;
- De nationale TTP-dienst plaatsen binnen het D4LS-programma, met eigen budget?

Naast het leveren van een nationale TTP-dienst voor pseudonimisatie en verdere versleuteling t.b.v. het koppelen van gepseudonimiseerde data, is er ook een sterke behoefte aan kennisuitwisseling, mede ten behoeve van bewustwording en het leren van anderen.

De propositie, kernactiviteiten en prijsmodel hangen verder af van het gekozen governance model voor een nationale TTP dienst.

Inkomsten

Financiering van dergelijke ondersteunende diensten voor onderzoek is een uitdaging. Onderzoeksbudgetten zijn momenteel veelal niet ingericht om te voorzien in dergelijke diensten en/of is een onderzoeker niet bewust van feit dat dergelijke faciliteiten nodig zijn om mee te begroten.

De kennisuitwisseling en ontwikkeling van een nationale gestandaardiseerde dienst zal bij voorkeur op nationaal niveau financieel ondersteund moeten worden middels onderzoeksbudgetten of algemene middelen bij de UMCs. Mogelijk kan ook financiering worden gezocht bij de onderzoeksinstellingen ter ondersteuning van onderzoek binnen de instellingen.

Mogelijke financieringsbronnen, onderzoekfinanciers zijn ZonMW, NWO. Te onderzoeken zijn NEN en NFU. Mogelijk kunnen ook Europese financieringsbronnen worden aangeboord als de propositie Europees wordt ingezet.

De geldstroom voor dergelijke diensten zal bepaald moeten worden:
- financiering middels onderzoeksbudgetten, overeen te komen met onderzoekfinanciers
- financiering middels interne middelen (overhead ICT ondersteuning)

De belangen liggen grotendeels bij onderzoek Nederland waardoor logischerwijze de financiering via het ministerie OCW (middels ZonMW, NWO) zou verlopen.

Verdienmodel

De kosten voor operationele activiteiten zullen doorberekend moeten worden naar klanten. Dit kan per record (per patiënt ID) of koppeling gedaan worden.

De wens van de onderzoeker is middels een freemium model gebruik te kunnen maken van de diensten ter verkenning van de toegevoegde waarde. Kosten hiervoor zullen doorberekend moeten worden in afname in staffels vanaf bijv. 10 records of koppelingen, afhankelijk van de use case.

Kernactiviteiten

Een nationale TTP-dienst zou naast het sleutel-management ook pseudonimisatie- en koppeldiensten kunnen aanbieden, waarbij er tussen de beide functies door technische en organisatorische maatregelen een strikte scheiding is aangebracht conform de eisen van de Autoriteit Persoonsgegevens. TTP diensten zullen geaudit worden op o.a. functiescheiding.

De commerciële TTP-dienstverleners gebruiken voor de 1ste versleuteling verschillende hashes en algoritmen, waardoor het koppelen van researchdata niet optimaal kan verlopen (vendor lock-in). Door de 1ste versleuting op een gestandaardiseerde manier uit te voeren, wordt koppeling van gepseudonimiseerde data vereenvoudigd.

De kernactiviteiten van een nationale TTP-dienst zouden kunnen zijn:
- Pseudonimisatiedienst en/of ter beschikking stellen van gestandaardiseerde software voor het creëren van een pre-pseudoniem
- sleutel-management
- koppeldiensten

Tussen het sleutel-management is door technische en organisatorische maatregelen een strikte scheiding nodig conform de eisen van de Autoriteit Persoonsgegevens. Dit is goed realiseerbaar, omdat TTP's geaudit worden en daarbij een oordeel over de functiescheiding ontstaat en eventueel bijgestuurd kan worden.

Onderstaande paragraaf beschrijft de kaders waarmee een nationale TTP-dienst te maken zal krijgen m.b.t. de kernactiviteiten. Er wordt onderscheid gemaakt naar nationale wet- en regelgeving en internationale normen.

Wet- en regelgeving

Er zijn verschillende wetten van belang voor het verantwoord gebruik van zorgdata in wetenschappelijk onderzoek; dat is dus *buiten* de wet- en regelgeving m.b.t. het zorgdomein zelf.

De belangrijkste is de Wet bescherming persoonsgegevens. Daarnaast bestaan er internationale normen die voorschrijven op welke wijze pseudonimisering en het koppelen van researchdata zo veilig mogelijk

gedaan kan worden. Het is van belang dat TTP-dienstverleners kunnen aantonen dat ze voldoen aan genoemde normen; dat geldt dus ook voor een nationale TTP-dienst.

Privacy wetgeving (nationaal)
De Wet bescherming persoonsgegevens (Wbp) stelt eisen aan een behoorlijke en zorgvuldige verwerking van persoonsgegevens (onder meer: uitdrukkelijke toestemming[2], doelbinding, bewaartermijnen, passende beveiliging, recht op inzage en correctie van je dossier en melding van datalekken).
Art. 13: "De verantwoordelijke legt passende technische en organisatorische maatregelen ten uitvoer om persoonsgegevens te beveiligen tegen verlies of tegen enige vorm van onrechtmatige verwerking".

Een persoonsgegeven is elk gegeven over een geïdentificeerde of identificeerbare natuurlijke persoon. Dit betekent dat informatie ofwel direct over iemand gaat, ofwel naar deze persoon te herleiden is. Dat het om een natuurlijke persoon moet gaan, houdt in dat gegevens van overleden personen of van organisaties geen persoonsgegevens zijn (zie https://www.cbpweb.nl/nl/over-privacy/persoonsgegevens/wat-zijn-persoonsgegevens?qa=persoonsgegevens).

Er zijn veel soorten persoonsgegevens. Voor de hand liggende gegevens zijn iemands naam, adres en woonplaats. Maar ook telefoonnummers en postcodes met huisnummers zijn persoonsgegevens. Gevoelige gegevens als iemands ras, godsdienst of gezondheid worden ook wel bijzondere persoonsgegevens genoemd. Deze zijn door de wetgever extra beschermd.

Patiëntgegevens zijn zogenaamde "bijzondere gegevens", waarvoor het strengste regime geldt.
Art. 21 Wbp geeft regels voor de verwerking van patiëntgegevens (o.a. geheimhouding).

De Autoriteit Persoonsgegevens, voorheen het College Bescherming Persoonsgegevens (CBP), functioneert als toezichthouder. In de Wet Geneeskundige Behandelingsovereenkomst (WGBO) is het gebruik van persoonsgegevens in het kader van een behandelrelatie geregeld.

De Autoriteit Persoonsgegevens heeft een aantal criteria opgesteld waaraan voldaan moet worden bij de toepassing van pseudonimisatie:

1. Er wordt vakkundig gebruik gemaakt van pseudonimisering, waarbij de eerste van de twee uitgevoerde versleutelingen van gegevens plaatsvindt bij de aanbieder van de gegevens (dus aan de bron bij de medisch behandelaar);
2. Er zijn technische en organisatorische maatregelen getroffen om herleidbaarheid van de versleuteling te voorkomen (2e versleuteling);
3. De verwerkte gegevens zijn niet indirect identificerend;
4. Deze drie voorwaarden worden onderworpen aan periodiek te houden audits;
5. Daarnaast dient de pseudonimiseringsoplossing op heldere en volledige wijze te worden beschreven in een actief openbaar gemaakt document, zodat iedere betrokkene kan nagaan welke garanties de gekozen oplossing biedt.

Deze voorwaarden zijn gesteld bij de vormgeving van pseudonimisering van de risicoverevening en in dat kader richting ministerie van VWS gecommuniceerd. Aangenomen wordt dat dezelfde criteria van toepassing zijn op het gebruik van zorg data voor wetenschappelijk onderzoek.

Internationale normen
Sinds 2008 bestaat er een ISO Technical Standard for Pseudonymization (ISO/TS 25237:2008,IDT), die door NEN is uitgegeven als Nederlandse Praktijkrichtlijn (NPR-ISO/TS 25237:2008,IDT). In deze richtlijn worden de-identificatie-eisen gesteld aan de omgang met persoonsgegevens in de zorg, aan het daarbij te gebruiken proces van pseudonimisatie en eisen aan de daarbij te hanteren technische voorzieningen, de zogenaamde trustworthy practices for re-identification (zie bijlage 4).

[2] In de zorg: informed consent

6. Aanbevelingen TTP-dienst

Uit de inventarisatiefase blijkt dat de behoefte aan pseudonimiseren en koppelen van onderzoeksdata over het algemeen latent is. Er is allereerst behoefte aan begeleiding bij het pseudonimiseren en koppelen van onderzoeksdata. Het kennisniveau bij diverse UMC's laat te wensen over en ook financieel is er vanuit de individuele onderzoeker vaak weinig financiële ruimte om onderzoeksgeld aan dit soort zaken te besteden. Om op nationaal niveau in deze behoefte te kunnen voorzien is het nodig dat er gewerkt wordt aan bewustwording, kennisdeling en coaching van onderzoekers bij de keuze in te gebruiken TTP-functionaliteiten.

Daarnaast kan parallel gewerkt worden aan de verdere uitwerking van de haalbaarheidsstudie. Dit zal vormgegeven worden via een in te richten Special Interest Group 'veilig koppelen van datasets'. Het primaire doel van deze interest group is o.a.:

- Realisatie van een netwerk waarin onderzoekers, informatiebeveiligers, beleidsmedewerkers, ICT managers van UMC's en instellingen, en andere stakeholders samenwerken met een gezamenlijk Nationaal belang,
- Strategieontwikkeling en lobby bij het Ministerie t.b.v. efficiënte uitwisseling van onderzoeksdata, inclusief de discussie om hiervoor het BSN te gaan gebruiken.
- Uitwerken van voor- en nadelen van architectuur modellen voor veilige data koppeling. Daarnaast een roadmap architectuur voor een nationale research infrastructuur die veilige data koppelingen faciliteert.

Governance modellen

Voor de inrichting van een nationale TTP-dienstverlener zijn in het voortraject reeds enkele governance modellen uitgedacht met een rol voor een onafhankelijke partij. De modellen zijn:

Model 1: coördinerend

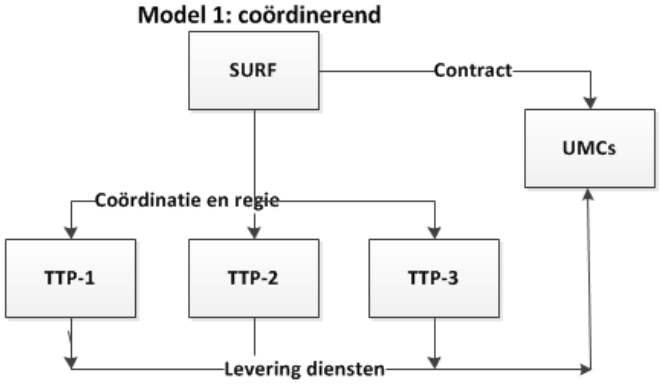
- Contract UMC's met SURF;
- SURF kent de innovatieve klantwensen en –eisen en coördineert hierin, m.n. bij integratie-vraagstukken;
- SURF bemiddelt naar TTP-leveranciers en onderhandelt schaalvoordeel
- SURF voert regie.

Voordelen:

- Verwachte prijsdaling vanwege schaalvoordelen (hogere volumes);
- Centrale coördinatie en
- Bevordering van integratie van diensten van verschillende TTP's.

Nadelen:

- Hoe dwing je integratie en/of standaardisatie af? Vraag-gestuurd: anders geen afname.



In het contract met UMC's verplicht SURF zich om de coördinatie tussen TTP's en de centrale regie in de infrastructuur te verzorgen. We gaan er van uit dat er daarvoor ook iets geregeld moet worden tussen TTP's en SURF. TTP's leveren rechtstreeks aan UMC's.

Model 2: meta TTP

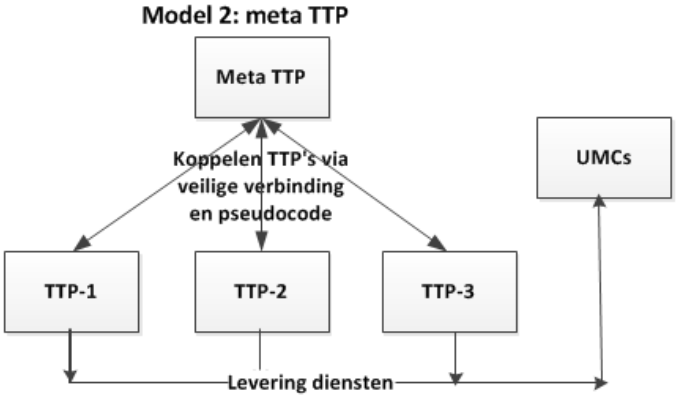
- Een meta TTP-functie zorgt er voor dat data van verschillende TTP's gekoppeld kunnen worden;
- Er is een veilige verbinding tussen de TTP's en de meta TTP mogelijk, waar o.b.v. een pseudocode (3de versleuteling) de TTP's gekoppeld kunnen worden door de meta TTP.

Voordelen:

- 2e en 3e versleuteling is minder kwetsbaar, omdat de identificerende sleutel daarbij niet nodig is
- Houdt het initiatief bij specifieke functionaliteit gevraagd door de aangesloten instellingen;
- Nodigt leveranciers uit dit te realiseren;
- Bevordert de verrijking van researchdata t.b.v. wetenschappelijk onderzoek.

Nadelen:

- Hoe dwing je samenwerking tussen de meta TTP en de bestaande commerciële TTP's af?
- Een extra versleuteling is nodig, waarmee het pseudonimisatie proces nodig verder wordt gecompliceerd.



SURF verzorgt een meta-TTP-functionaliteit, die er voor zorgt dat data van verschillende TTP's gekoppeld kan worden. Communiccatie tussen SURF en TTP's is beveiligd. TTP's leveren aan UMC's.

Model 3: SURF TTP

SURF levert:

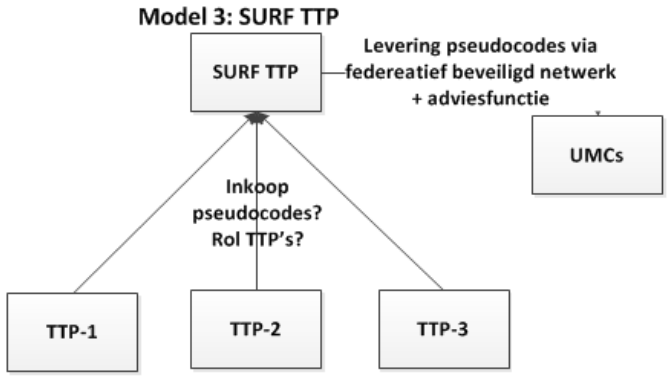
- De kern TTP-functionaliteiten: de pseudocode (3de versleuteling);
- Advies voor gebruik van de juiste functionaliteit;
- Afname diensten via veilig SURF-netwerk (zorgdomein op aparte lichtpaden met firewall er omheen);
- Autorisatie en authenticatie via federatieve model, als bij SURFconext, eduroam;
- Gebruikers niet meer in dienst verdwijnt dan direct bij de aangesloten servies.

Voordelen:

- Uniformiteit en flexibiliteit in authenticatie/autorisatie en toegang tot data;
- Netwerkbeveiliging eenduidig en transparant te regelen;
- Minder vervuiling igv bevoegdheden bij personeelsmutaties;
- SURFnet is vertrouwd met federatieve beveiligings services;
- Custodix heeft op dit idee positief gereageerd

Nadelen:

- Rol overige commerciële TTP's onduidelijk;
- Toeleverancier, inkooprelatie.



SURF levert pseudocodes via federatief model aan UMC's, adviseert onderzoekers over de te gebruiken functionaliteit.
Rol TTP's onvoldoende duidelijk.

Deze drie governance modellen zijn in het voortraject van deze haalbaarheidsstudie opgesteld, maar niet verder uitgewerkt. Op zich is het wel nodig om de verantwoordelijkheden tussen SURF en de commerciële TTP's duidelijk te omschrijven per model. Zo is nu nog onduidelijk welke rol de TTP's krijgen in model 3 en/of er sprake zou zijn van inkoop. Dit zal nog verder onderzocht moeten worden.

Desgevraagd naar het voorkeursmodel van de geïnterviewde(n) gaven veel partijen aan dat als er iets op nationale schaal ingericht gaat worden, dat deze dan zowel de regiefunctie als operationele diensten zou kunnen aanbieden, d.w.z. model 3 nationale TTP dienst.

Faseren

Om voor de hand liggende redenen is het niet verstandig om in één keer een volledige nationale TTP-dienst voor sleutelmanagement én pseudo- en koppeldienstverlening in te richten:

- Het is nog niet altijd even duidelijk welke functionaliteiten er op nationaal niveau behoefte is;
- Er is nog geen budget om zo'n organisatie, zowel qua techniek, als qua kennis en bemensing in te richten;
- Er zijn reeds bestaande TTP-dienstverleners, met wie nog geen afstemming heeft plaatsgevonden;
- Het business model dient nog verder uitgewerkt te worden, met name financieel.

Het advies is derhalve: ga een traject in van samenwerking, kennisvergaring en –deling en doe ervaring op in een aantal bescheiden pilots. Richt hiervoor een Special Interest Group TTP in, waarin de belangrijkste stakeholders samenwerken aan een ontwikkelmodel voor nationale dienstverlening.

Ambiëren

Start met een bescheiden ambitie, waarbij rekening gehouden wordt met wat reeds bereikt is op decentraal niveau. Breng uitsluitend die zaken onder een nationale TTP-dienst die niet of onvoldoende efficiënt decentraal uitgevoerd kunnen worden, zoals bijvoorbeeld:

- Het vergaren en delen van kennis;
- Het participeren in standaardisatie-overleg;
- Het verzorgen van grensoverschrijdende koppelingen.

Versterken business case

Zoek samen met belangrijke stakeholders naar pilots met een duidelijke meerwaarde voor het veld. Bijvoorbeeld het standaardiseren van de 1e (en wellicht ook 2e) versleuteling in het pseudonimisatie-proces, waardoor hergebruik van data bevorderd wordt indien die standaarden gebruikt worden. Onderzoek daarvoor eerst of er veel voorkomende onderzoeks *use cases* zijn, waarbij de data op eenzelfde gestandaardiseerde wijze kunnen worden aangeleverd ten behoeve van pseudonimisatie.
Er zijn in het voortraject drie *use cases* benoemd, waarvan onderzocht moet worden hoe representatief ze zijn, alsmede of er nog meer veel voorkomende *use cases* zijn. Die zouden als eerste op nationaal niveau gefaciliteerd kunnen worden.

Een onderzoeker kan dan kiezen om met bijvoorbeeld één externe partij decentraal te koppelen of via de centrale gestandaardiseerde voorziening wanneer externe beschikbaarheid een issue is. Dit laatste is ook internationaal van belang.

Ook lobby kan hieronder vallen. Bijvoorbeeld over het gebruik van het BSN bij koppelen (is deterministisch) is nog lang niet in alle omstandigheden toegestaan; lobby kost ook tijd en inspanning.

Probeer een of twee pilots met wat grotere projecten, waarin data stewardship al aan de orde is en de kosten voor pseudonimisatie begroot zijn of kunnen worden.

Zorg voor een landelijk overzicht van welke onderzoekdata (kenmerken) waar beschikbaar zijn. Zo'n centraal register of catalogus zou het hergebruik van onderzoekdata kunnen bevorderen.

De tweede prioriteit moet zijn het opzetten van de nationale TTP-dienst zelf. Daarbij moet goed gekeken worden naar behoeften van onderzoekers, voor de hand liggende *use cases*, gewenste governance, duurzaamheid en belangen van andere stakeholders.

Kennis delen

Wellicht moet de eerste prioriteit zijn om via voorlichting en bewustwording onderzoekers te informeren over de voordelen van goede pseudonimisatie en koppelingen. Het beeld uit de interviewfase is dat veel onderzoekers 'er' niet aan beginnen, omdat 'het' zo moeilijk en duur is. Vertrouwen ontbreekt en dat kan alleen met goede informatie doorbroken worden. Een *centraal kennisplatform* moet deze impasse doorbreken: organiseer masterclasses, voorlichting, praktijkoefeningen, *best practices*, seminars, etc. onder de vlag van D4LS; dat hoeft niet duur te zijn, maar is wel een noodzakelijke voorwaarde voor acceptatie en groei.

Financieren

De inrichting van een SIG TTP zou gefaciliteerd kunnen worden door SURF.

De kennisuitwisseling en ontwikkeling van een nationale gestandaardiseerde dienst zal op nationaal niveau financieel ondersteund moeten worden. Mogelijke financieringsbronnen zijn Zon-MW, en NWO.

Het begint daarmee onder de vlag van SURF (SIG TTP) en kan later, indien nodig, als onafhankelijke stichting neergezet worden met externe financiering.

Vervolgstappen

Met de haalbaarheidsstudie is de noodzaak om op nationaal niveau bepaalde activiteiten te ontplooien op het gebied van *trusted third parties* aangetoond: de vraagzijde heeft duidelijk behoefte aan sturing en begeleiding bij het opzetten van het pseudonimisatie- en koppelproces voor researchdata. Er hoeft verder geen tijd verspild te worden.

Deze paragraaf schetst op welke wijze op projectbasis gewerkt kan worden aan het ontwikkelen van een Nationale TTP-dienst.

Om voor de hand liggende redenen is het verstandig om te werken met een groeimodel. Daarbij onderscheiden we de volgende fasen:

1) **Project**
    a. Richt onder de vlag van SURF een SIG TTP op: deelname belangrijkste stakeholders (op eigen kosten), zowel behoeftestellers (onderzoekers en datamanagers) als bestaande (commerciële) aanbieders (nadat eerst nog apart met ze gesproken is);
    b. Zorg voor budget voor uitvoering van het projectplan via D4LS of anderszins;
    c. Benoem een kwartiermaker Nationale TTP, die als trekker en wegbereider kan functioneren. Zijn eerste taak is om dit projectplan verder vorm te geven in samenspraak met de SIG;
    d. Respecteer datgene wat bereikt is op decentraal niveau en breng uitsluitend zaken onder een nationale TTP-dienst die niet of onvoldoende efficiënt decentraal uitgevoerd kunnen worden.

2) **Selectie pilots**
    a. Kies gezamenlijk enkele pilots, waarin kan worden samengewerkt; bij voorkeur de wat grotere projecten waarbij reeds sprake is van data stewardship;

b. Kies voor die projecten die leiden tot het oplossen van praktische problemen en op basis waarvan aan kennisdeling en awareness gedaan kan worden; zet hierbij bijvoorbeeld enkele veel voorkomende *use cases* centraal;
c. Bekijk welke rol standaardisatie hierin kan spelen.

3) Kennisplatform

a. Richt een centraal kennisplatform op, waarin leertrajecten ontwikkeld worden voor onderzoekers en data managers: hoe doe je dit en dat… Leer bijv. van de problemen in de TraIT-pilot met PALGA;
b. Verzamel *best practices,* geef *masterclasses*, neem onderzoekers bij de hand, e.d.
c. Kijk ook naar de gang van zaken in het buitenland: wat kunnen we daar van leren?

4) Standaardisatie

a. Onderzoek welke standaarden nodig zijn om tussen instellingen te kunnen koppelen, niet bilateraal, maar multilateraal ook in de toekomst.

5) Centrale catalogus

a. Ga na hoe een centrale catalogus georganiseerd kan worden, waarin bekeken kan worden welke bestaande datasets op welke kenmerken gekoppeld kunnen worden;
b. Maak zo'n catalogus, waarbij de data blijft waar ze nu is en uitsluitend de informatie over datasets gecentraliseerd openbaar gemaakt wordt. Dit in samenspraak met D4LS WP2 (catalogi).

6) Evalueer periodiek

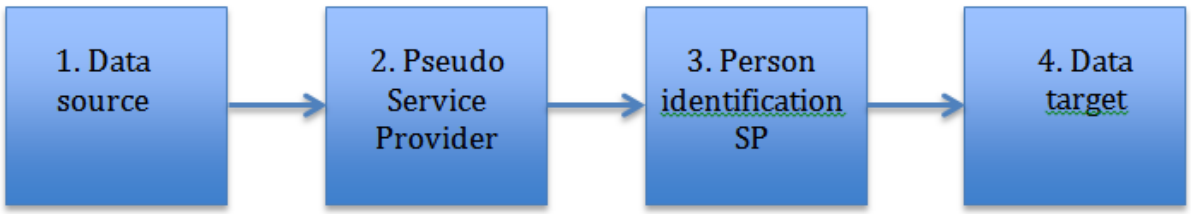
a. Evalueer periodiek of er nog steeds in een behoefte wordt voorzien en of datgene wat centraal ontwikkeld wordt voor iedereen bruikbaar is.

7. Referenties

[1] NPR-ISO/TS 25237:2008,IDT: Medische informatica – Pseudonimisatie;
[2] E. Flikkenschild, De TTP-functionaliteit voor research central via SURF beschikbaar? Een eerste notitie voor verdere discussie, 4-8-2014;
[3] Innovatievoorstel, verkenning van een nationale Trusted Third Party dienst, auteur en datum niet bekend;
[4] Architectuur Centrale infrastructuur, Parelsnoer Initiatief, 11-02-2011;
[5] B. Franken, Een standaard voor pseudonimisatiediensten? 2013;
[6] W. de Bruijn, Mondriaan project Health Fesearch Data, Mondriaan pseudonimiseren, koppelen en ….. voor onderzoek, 7-7-2014;

Bijlage 1: Het pseudonimisatieproces

Bron: ISO TS 25237



Het pseudonimisatieproces bestaat uit vier delen:

1. De **bron**, die de volgende functies vervult:
   a. Voorbereiden en structureren van de gegevens voor overdracht aan de 2e en 3e functie (2. **person identification service** en de 3. **pseudonimisatie-dienst**): de pseudo-dienst moet weten wat er verwacht wordt dat er gedaan wordt met elk data element, bijvoorbeeld door het plaatsen van 'tags' aan elk element of door de elementen in een voorgedefinieerde locatie te plaatsen. Een vorm van standaardisatie dus: zo moet het worden aangeleverd, opdat het correct verwerkt kan worden;
   b. De feitelijke overdracht naar de person identificatie service en de pseudo-dienst. Dat kan bijvoorbeeld door ze eerst te bellen;
   c. Toezicht op de resulterende code van de pseudo-dienst: voldoet de code aan de vereisten en kan deze verstuurd worden.
4. Het **doel** is de eenheid die de gepseudonimiseerde data ontvangt en zorgt voor de verdere procesgang van de gegevens. Die procesgang is mede afhankelijk van de lokale wet- en regelgeving en het privacy-risiconiveau. Gepseudonimiseerde data kunnen ook onder de privacywetgeving vallen, zoals:
   d. Ontcijfering van de gepseudonimiseerde data;
   e. Regels voor de opslag van deze data;
   f. Statistische analyse van de opgeslagen datasets.

Bijlage 2: de belangrijkste begrippen uit ISO 25237

Onderstaand worden de meest gebruikte begrippen gedefinieerd, gebaseerd op ISO 25237.

Data-subject
De persoon naar wiens gegevens verwezen wordt

Direct identificerende gegevens
Gegevens die direct herleidbaar zijn tot een natuurlijke persoon

Indirect identificerende gegevens
Gegevens die uitsluitend tot een natuurlijke persoon herleidbaar zijn wanneer ze in combinatie met andere indirect identificeerbare gegevens worden gebruikt

De-identificatie
Algemene term voor elk proces waarmee de verbinding tussen een set identificerende gegevens en het data subject wordt verwijderd (bijvoorbeeld anonimisatie en pseudonimisatie)

Anonimiseren
Het proces dat de verbinding tussen de identificerende gegevens en het data subject verwijdert

Geanonimiseerde gegevens
Gegevens van de persoon die niet door de ontvanger van die gegevens tot de persoon kunnen worden herleid

Persoonsidentificatie
Elke informatie die als doel heeft een persoon in een bepaalde context uniek te identificeren

Pseudoniem
Een persoonsidentificatie die verschilt van de gebruikelijk gehanteerde persoonsidentificatie, waardoor de persoon voor derden niet meer te identificeren is (pre-pseudoniem of pseudo-code)

Pseudonimisatie
Bijzondere vorm van anonimiseren, waarbij zowel de verbinding tussen een set identificerende gegevens en het data subject wordt verwijderd, als een nieuwe verbinding wordt gemaakt tussen een bepaalde set van karakteristieken die verwijzen naar het data subject en een of meerdere pseudoniemen

Trusted third party (TTP)
Een vertrouwde onafhankelijke partij die diensten aanbiedt die de betrouwbaarheid van elektronische gegevensuitwisseling en gegevensopslag vergroten

Persoonsgegevens

Identifying data
Dat deel van de data dat een set karakteristieken bevat die tot unieke identificatie van het data subject leiden
Payload data
Dat deel van de data dat een set karakteristieken bevat die het niet mogelijk maken het data subject uniek te identificeren, bijvoorbeeld medische gegevens

Bijlage 3: Soorten privacy-risico's

Bron: ISO TS 25237

Er bestaat altijd een risico dat ongeautoriseerde her-identificatie van gepseudonimiseerde data bereikt kan worden, bijvoorbeeld met behulp van krachtige computers en decryptie software. Daarom wordt onderscheid gemaakt naar drie niveaus van privacy-risico's:

1. Risico's verbonden aan de persoonlijk identificerende data;
2. Risico's verbonden met geaggregeerde data;
3. Risico's verbonden met bijzondere data in de database, bijvoorbeeld gegevens over een bijzondere, zeldzame ziekte, die wel tot een natuurlijk persoon te herleiden is.

Om deze redenen moet aantoonbaar gemaakt worden dat op elk van deze niveaus afdoende maatregelen genomen zijn.
Niveau 1 van privacybescherming: verwijdering van duidelijk identificeerbare gegevens en eenvoudig te verkrijgen indirect identificerende gegevens;
Niveau 2 van privacybescherming: maatregelen tegen aanvallers die externe data gebruiken in combinatie met de gepseudonimiseerde data;
Niveau 3 van privacybescherming: maatregelen die identificatie van bijzondere data in databases onmogelijk maken.

Bijlage 4: Technische vereisten uit ISO 25237

Pagina 17/18:
The service entrusted to protect the patient identities shall conform to minimum trustworthy practices requirements:

- there is a need to assure the health consumer's confidence in the ability of the health system to manage the confidentiality of their information;
- there is a need for the service to provide physical security protection;
- there is a need for the service to provide operational security protection;
- re-identification keys, transformation tables and protection need to be subject to multi-person controls and/or multi-organization controls consistent with the assurances claimed by the service;
- the service shall be under the control of (e.g. contractually or operationally) the custodian of the source identifiers;
- legal and environmental constraints surrounding release of re-identification keys and protections need to be disclosed in support of the privacy protection levels claimed by the service;
- quality and availability of service needs to be specified and provided in accordance with the information provision and access needs;
- some identifiers may simply be blanked as they are unnecessary for the use;
- some identifiers may be blurred in a way consistent with the intended use.

Zie ook **pag. 30 e.v.**, 9.3 Trustworthy practices for operations en 9.4 re-identification, waar e.e.a. nader wordt gespecificeerd:
Provision of pseudonymization services in healthcare shall meet the following objectives in order to be effective in securing the privacy protection of personal health information:

- the reliable and secure binding of unique pseudonyms to individuals or organizations that are the subject of pseudonymized personal health information;
- the protection of the pseudonyms from unauthorized re-identification;
- the provision of authorized re-identification of the subject's source identifier(s) in accordance with reidentification policy parameters as agreed between the service provider and the service subscriber.

The above objectives shall be accomplished in a manner that maintains the trust of all who rely upon the confidentiality of the personal health information that is protected through the pseudonymization service. As pseudonymization is particularly suited to primary and secondary research and analysis purposes, information resources that rely upon these services to protect personal health information may implicitly require the trust of the patients whose information is being examined, as well as that of the general public. It is unlikely that either healthcare providers or patients will cooperate in the contribution of personal health information for analytical purposes if such identity protection services are believed to be insecure.

In order to satisfy these requirements, a pseudonymization service:

- should be strictly independent of the organizations supplying source data;
- shall be able to guarantee security and trustworthiness of its methods by publishing to its subscribers its operating practices;
- shall be able to guarantee security and trustworthiness of its software modules:
    - shall provide assurances as to the source, processes and integrity of its software modules,
    - code integrity shall be asserted through code signing;
- shall be able to guarantee security and trustworthiness of its operating environment, platforms and infrastructure:
    - shall restrict network traffic to restrict all unnecessary traffic,
    - shall disable all unnecessary operating system services,
    - shall provide technical, physical, procedural, and personnel controls in accordance with ISO 27799;
- shall implement monitoring and quality assurance services and programs:
    - to assure quality of service,
    - to monitor against network penetration and malicious attacks;
- cryptographic key management:
    - shall be under multi-person control,
    - identifiers shall be encrypted by two keys, one under control of the data source, and one under the control of the pseudonymization service;
- instantiation of the pseudonymization service:
    - shall be documented,

    - shall be recorded and audited to be able to demonstrate service integrity;
- business continuity of the pseudonymization service:
    - shall be assured through backup,
    - shall be assured through a disaster recovery plan;
- internal audit procedures:
    - shall be documented,
    - shall be executed on no less than a monthly basis;
- external audit procedures:
    - shall be used to establish to the satisfaction of subscribers and any relying party that it fully complies
    - with published operating procedures,
    - the auditor shall be completely independent of the audited party by belonging to a separate organization from the pseudonymization service provider,
    - the auditor shall have no financial interest in the audited party,
    - the auditor shall be a qualified information systems auditor to the extent necessary for admission to the relevant professional body,
    - service subscribers shall immediately be notified of any pseudonymization services that are found by an auditor to be deficient;
- participants:
    - shall maintain integrity of the organization's key(s) associated with pseudonymization,
    - shall maintain physical, network, personnel, and technical controls of the associated systems in accordance with ISO 27799,
    - are responsible for the anonymization of payload data and privacy protection of any pseudonymized information resources maintained by the organization.
- risk assessment shall be conducted regarding access by the data source to the resulting pseudonyms and specification of such restrictions shall be expressed in operational policies.

9.4 Implementation of trustworthy practices for re-identification

The pseudonymization service should provide support for controlled re-identification. A pseudonymization service providing such support shall make available to subscribers and data subjects a re-identification policy that specifies the criteria required for an authorized re-identification event.

- Re-identification shall be subject to multi-person controls, and multi-organization control.
- Time-sensitive re-identification shall be accommodated within the defined policy and process communicated to those needing time-sensitive processes (e.g. public health authorities).
- Audits shall:
    - be provided for all re-identification events in accordance with RFC 3881;
    - minimally include:
        - the party to whom the identity was disclosed;
        - the time/date of the re-identification;
        - the reason for re-identification as defined in 5.5.
- Re-identification from the pseudonymization service shall re-identify only the local pseudonym from the source organization.
- The controller of the data is responsible for re-identification of the patient, and may, as permitted by local jurisdiction, validate further the re-identification request.