

LUMC-PHEG

Data Governance

ELAN-GP datawarehouse

Extramuraal LUMC Academisch Network
17 Mei 2022

Table of Contents

1.	Doelstellingen	2
2.	Afkortingen	2
3.	Infrastructuur	3
3.1.	Overzicht	3
3.2.	Veiligheid	4
3.3.	Toegankelijkheid	4
3.4.	Verantwoordelijkheid	5
3.5.	Mogelijkheid tot bezwaar	5
4.	Data	6
4.1.	Transformatie	6
4.2.	Pseudonimisering & Depseudonimisering	6
4.2.1.	Pseudonimisering	7
4.2.2.	Depseudonimisering	8
4.3.	Aanbod	8
4.3.1.	Metadata	8
4.3.2.	Honorering	9
4.3.3.	Uitgifte	9
4.3.4.	Archivering	10
5.	Onderzoek	10
5.1.	Onderzoek met gepseudonimiseerde data	10
5.2.	Onderzoek met geïdentificeerde patiënten of huisartsen	11
5.3.	Onderzoek met gekoppelde data	12

Versies

Versie	Auteur	Datum
1.3	Margot de Waal	17 mei 2022
1.2	Margot de Waal	17 oktober 2019
1.1	Henk de Jong	16 mei 2019
1.0	Henk de Jong	4 maart 2019
0.6	Henk de Jong	19 februari 2019
0.5	Michiel Meulendijk	20 december 2018
0.4.1	Michiel Meulendijk	18 oktober 2018
0.3	Michiel Meulendijk	20 augustus 2018
0.2	Michiel Meulendijk	27 februari 2018
0.1	Michiel Meulendijk	28 november 2017

1. Doelstellingen

Binnen het Extramuraal LUMC Academisch Netwerk (ELAN) is een infrastructuur opgebouwd voor de uitwisseling van patiëntendata tussen huisartsen en de afdeling PHEG van het LUMC. Deze ELAN-GP datawarehouse voorziening heeft de volgende expliciete doeleinden:

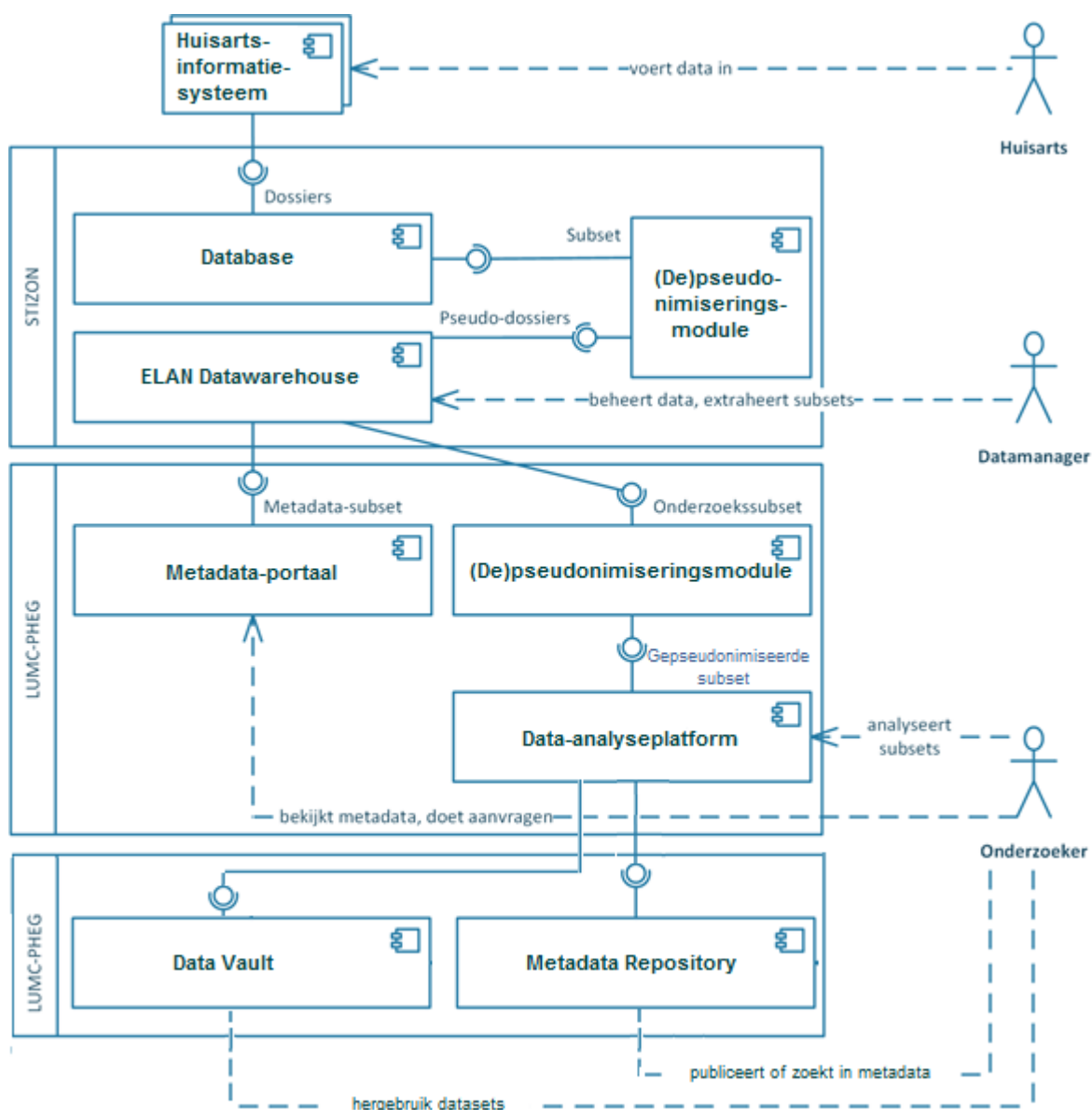
1. Terugkoppeling aan de zorgverlener over de (kwaliteit van de) registratiegegevens op geaggregeerd niveau per praktijk en zorgverlener;
2. Wetenschappelijk onderzoek met betrekking tot gezondheid en gezondheidszorg, niet commercieel, met rapportages op geaggregeerd niveau, per zorggroep of nader te definiëren deelpopulatie;
3. Terugrapportages op het gebied van innovatie van zorg of onderwijs, die voor de zorgverlener in de praktijk via de-pseudonimisatie zo nodig zijn te herleiden tot individuen;
4. Selectie van potentiële deelnemers voor patiëntgebonden onderzoek, die uit naam van de huisarts kunnen worden aangeschreven voor een specifiek project, en na informed consent het verzamelen van follow-up gegevens van deelnemers aan dit specifieke project via het ELAN-GP datawarehouse.

2. Afkortingen

1. Leids Universitair Medisch Centrum (LUMC)
2. LUMC Campus Den Haag (LCD)
3. Public Health en Eerstelijns Geneeskunde (PHEG)
4. Extramuraal LUMC Academisch Netwerk (ELAN)
5. Stichting Informatievoorziening voor Zorg en Onderzoek (STIZON)
6. Trusted Third Party (TTP)
7. Huisartsinformatiesysteem (HIS)
8. Data Transfer Agreement (DTA)

3. Infrastructuur

3.1. Overzicht



Figuur 1: Overzicht van de relaties tussen de verschillende componenten in het ELAN-netwerk.

De infrastructuur is zo ingericht dat data op een veilige manier kan worden uitgewisseld. Huisartsinformatiesystemen (HIS'en) van huisartsen die STIZON gemachtigd hebben hun data te extraheren vormen de bronsystemen van eerstelijnsdata. De dossiers van deelnemende patiënten worden periodiek geëxtraheerd uit de HIS'en en overgebracht naar een centrale database van STIZON.

De patiëntendossiers van aan ELAN deelnemende huisartsen worden als een gepseudonimiseerde subset overgebracht naar het ELAN-GP datawarehouse. De pseudonimisering wordt uitgevoerd door STIZON in een dubbelrol als TTP. Het ELAN-GP datawarehouse bevindt zich in een beveiligde STIZON-omgeving, die alleen toegankelijk is voor een selecte groep medewerkers van PHEG.

Onderzoekers kunnen een data-aanvraag doen in het kader van een specifiek onderzoek. Die aanvraag zal, mits gehonoreerd door de ELAN-huisartsen stuurgroep, door een datamanager geëxtraheerd worden, op subsetniveau nogmaals gepseudonimiseerd worden, en met de onderzoeker worden gedeeld. Gehonoreerde subsets worden beschikbaar gesteld aan voor het onderzoek geautoriseerde onderzoekers via een data-analyseplatform.

3.1.1. Toekomst visie

In de toekomst zullen metadata over de patiëntendossiers inzichtelijk worden gemaakt aan LUMC-onderzoekers via een intern data-portaal. Ook de data-aanvraag kan dan via dit portaal worden gedaan.

3.2. Veiligheid

In het ELAN-project wordt gestreefd naar een zo groot mogelijke veiligheid van de data. Met dat doel worden alle opslagplaatsen en overdrachtsroutes adequaat beveiligd.

1. STIZON, waar de STIZON-database is ondergebracht, is ISO 27001 en NEN 7510 gecertificeerd.
2. Het ELAN-GP datawarehouse wordt op dezelfde manier beveiligd als de STIZON-database. De toegang voor de selectie groep LUMC-medewerkers tot de data van het ELAN-GP datawarehouse is beveiligd d.m.v. twee-factor-authenticatie.
3. Bij de infrastructuur van het LUMC worden studiedatasets afkomstig uit het ELAN-GP datawarehouse gearchiveerd. Het LUMC hanteert NEN 7510 als richtsnoer voor de inrichting van de informatiebeveiliging en heeft o.a. de volgende maatregelen geïmplementeerd:
 - a. Er zijn adequate procedures over communicatie, support en beheer van de infrastructuur (in het bijzonder m.b.t. informatieveiligheid);
 - b. Wijzigingen in gegevens of in informatieverwerking worden uitsluitend uitgevoerd onder een procedure voor wijzigingsbeheer.
 - c. Er wordt gebruik gemaakt van logische toegangsbeveiliging waardoor enkel geautoriseerde personen toegang hebben tot informatie;
 - d. De apparatuur is te allen tijde opgesteld in een beveiligde ruimte;
 - e. Er worden reguliere back-ups gemaakt van gegevens;
 - f. Software wordt regelmatig geüpdatet en er wordt frequent gecontroleerd op virussen.
4. Aan dataplatforms waarop ELAN-data geanalyseerd wordt, worden dezelfde eisen gesteld als aan andere partijen in de samenwerking; d.w.z. ISO 27001 en NEN 7150-certificering of geïmplementeerde richtlijnen die op basis hiervan opgesteld zijn.

3.3. Toegankelijkheid

De data in het ELAN-GP datawarehouse is alleen toegankelijk voor een selecte groep PHEG-medewerkers bestaande uit datamanagers en dataspecialisten. In de toekomst zal via een dataportaal metadata beschikbaar gesteld worden aan onderzoekers. Deze metadata omvat informatie over de data, maar niet de inhoud van de data zelf. Via de metadata zijn o.a. inzichtelijk:

1. De structuur van de data;
2. Beschrijvingen van de structuur met als doel variabelen inzichtelijk te maken;
3. Het aantal rijen dat beschikbaar is per dataset.

Alleen LUMC-onderzoekers hebben toegang tot deze metadata. Na aanvraag (en goedkeuring) van een subset van de data voor onderzoek, wordt deze beschikbaar gesteld aan de aanvrager d.m.v. een dataplatform.

3.4. Verantwoordelijkheid

De volgende verantwoordelijkheden zijn te onderscheiden m.b.t. ELAN-data:

1. De zeggenschap over de in ELAN opgenomen data blijft bij de registrerende huisartsen liggen.
2. De verantwoordelijkheid voor het al dan niet honoreren van data-aanvragen voor onderzoeksprojecten ligt bij de ELAN-huisartsen stuurgroep.
3. De juridische verantwoordelijkheid voor de data-subsets binnen de infrastructuur van het LUMC, ligt bij het LUMC.

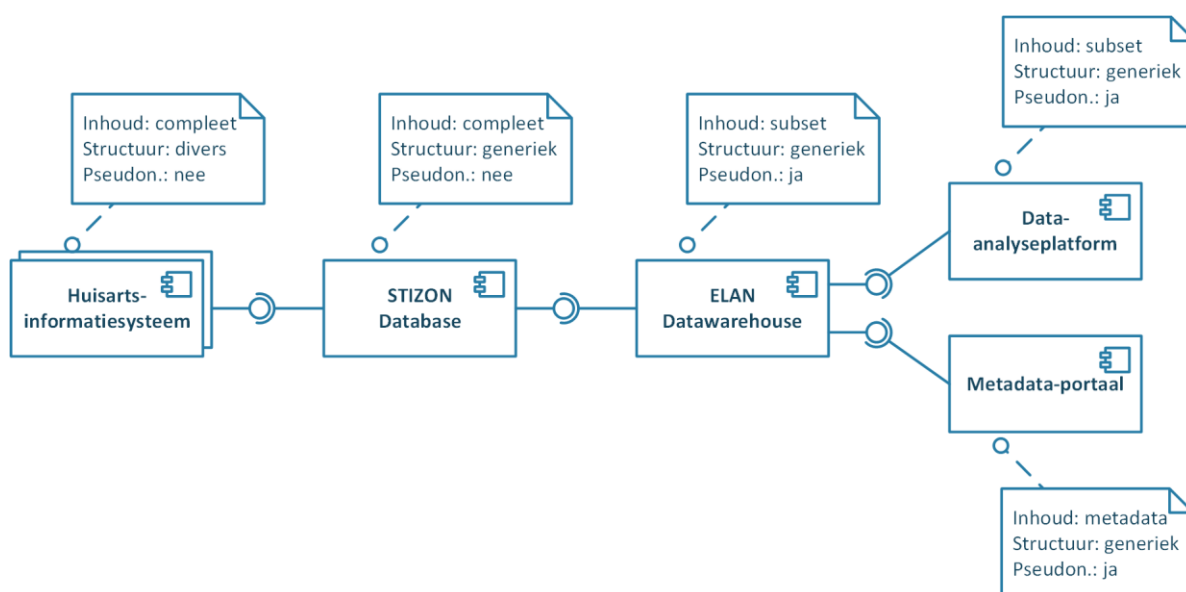
3.5. Mogelijkheid tot bezwaar

Middels voorlichtingsmateriaal bij de huisartspraktijken zelf (posters, folders, informatieschermen, website) worden de patiënten geïnformeerd dat hun praktijk gecodeerde gegevens levert ten behoeve van wetenschappelijk onderzoek. Hierin wordt uitdrukkelijk vermeld dat tot personen herleidbare gegevens niet meegeleverd worden. Mocht een patiënt echter besluiten dat hij/zij liever geen medische data beschikbaar wil stellen voor dit doel, bestaat er de mogelijkheid tot opt-out: in het HIS van de huisarts wordt een aantekening gemaakt dat betreffende patiënt bezwaar maakt. Deze aantekening van bezwaar wordt bij STIZON verwerkt en zorgt er voor dat geen data van desbetreffende patiënt wordt opgenomen in het datawarehouse.

Deze opt-out werkwijze wordt gedekt door nationale wetgeving die stelt dat verstrekking van patiëntdata voor onderzoeksdoeleinden toegestaan is, mits kan worden aangetoond dat het vragen van toestemming aan iedere patiënt afzonderlijk in redelijkheid niet kan worden verlangd én dat herleiding tot individuele personen redelijkerwijs wordt voorkomen (Burgerlijk Wetboek 7:458). Ook artikel 9.2.j van de Algemene verordening gegevensbescherming (AVG) ondersteunt deze werkwijze: verwerking van [...] gegevens over gezondheid zijn verboden, tenzij deze verwerking ten behoeve van wetenschappelijk onderzoek geschiedt, waarbij de evenredigheid met het nagestreefde doel wordt gewaarborgd en het recht op bescherming van persoonsgegevens wordt geëerbiedigd.

4. Data

4.1. Transformatie



Figuur 2: Overzicht van de datatransformaties tijdens de ELAN-extracties.

Tijdens de extracties wordt de data verschillende keren getransformeerd om uiteenlopende redenen (o.a. generalisatie, pseudonimisering, toegankelijkheid).

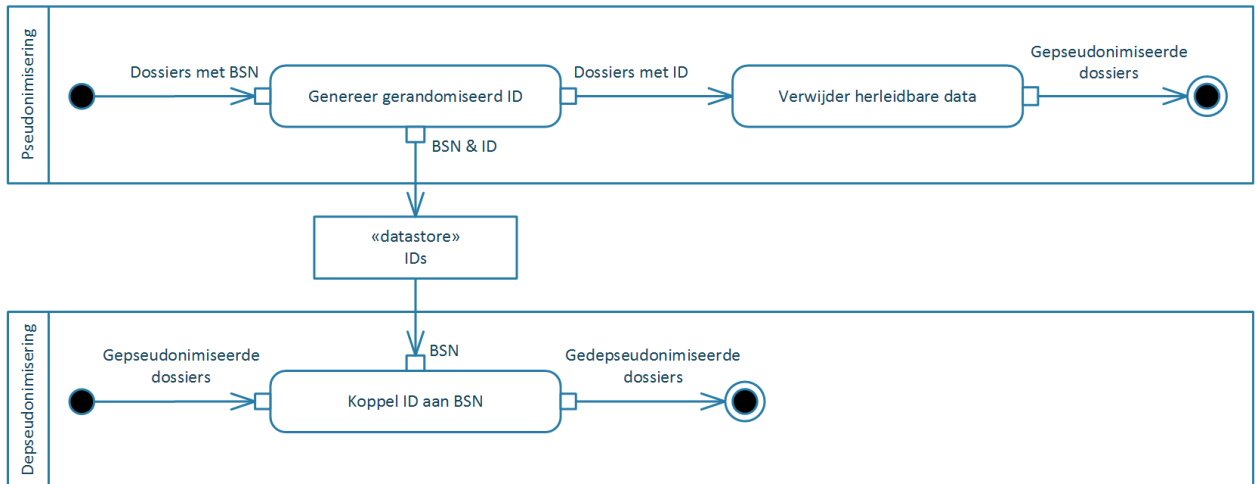
1. Gedurende de eerste extractie, vanuit de HIS'en naar de STIZON-database, worden de diverse uitleverstructuren van de HIS-leveranciers gegeneraliseerd tot één structuur. Hierbij wordt dataverlies tot een minimum beperkt. Het doel is de data zo volledig mogelijk te transformeren naar een structuur die verdere bewerking mogelijk maakt. In deze vorm bevat de data zowel gecodeerde als ongestructureerde journaal-, episode-, lab-, en medicatiedata.
2. Tijdens het pseudonimiseringsproces wordt een subset samengesteld die alle tot patiënt- en huisartsniveau herleidbare data verwijdert. Zowel huisartsen als patiënten zijn na pseudonimisering niet meer direct herleidbaar door onderzoekers. De procedure wordt uitvoerig beschreven in de sectie Pseudonimisering. De route om contact te zoeken met huisartsen of patiënten loopt via STIZON welke fungeert als TTP. Deze subset vormt de inhoud van het voor PHEG toegankelijke ELAN-GP datawarehouse.
3. Via het dataportaal wordt metadata beschikbaar gesteld aan onderzoekers.
4. Tijdens het prepareren van een subset voor onderzoek wordt een selectie van de data ontsloten voor onderzoekers via een dataplatform dat aansluit op de eisen van het onderzoek.

4.2. Pseudonimisering & Depseudonimisering

Gegevens worden via tweewegpseudonimisering versleuteld. Dat wil zeggen dat alle herleidbare gegevens worden verwijderd, maar dat een sleutel bewaard wordt om gepseudonimiseerde data te kunnen depseudonimiseren.

Data wordt tweemaal gepseudonimiseerd: de eerste keer als STIZON de data gereedmaakt voor inclusie in het ELAN-GP datawarehouse, en de tweede keer als een subset uit het ELAN-warehouse wordt uitgegeven voor onderzoek. Deze methode garandeert dat voor verschillende onderzoeken uitgegeven subsets niet aan elkaar gekoppeld kunnen worden d.m.v. overlappende pseudoniemen.

4.2.1. Pseudonimisering



Figuur 3: Illustratie van de pseudonimiserings- en depseudonimiseringsprocessen.

Tijdens het pseudonimiseringsproces wordt een subset samengesteld die alle tot patiëntniveau herleidbare data verwijdert. Dit gebeurt d.m.v. de volgende deelprocessen:

- a. Verwijdering van burgerservicenummers (BSN);
- b. Verwijdering van naam-, adres-, en woonplaats- (NAW)-gegevens;
- c. Normalisering van geboortedatum: geboortedata worden gereduceerd tot jaren;
- d. Transformatie en / of verwijdering van vrije tekst, volgens een van de volgende deelprocessen:
 - i. D.m.v. automatische tekstanalyse wordt relevante niet-herleidbare data uit vrije tekst geëxtraheerd. Deze data maakt deel uit van de subset. Vrije tekst wordt verwijderd.
 - ii. D.m.v. automatische tekstanalyse worden herleidbare persoonsgegevens uit vrije tekst verwijderd. De niet-herleidbare vrije tekst maakt deel uit van de subset.
 - iii. Vrije tekst wordt verwijderd.

Een soortgelijk proces wordt uitgevoerd om huisartsen te pseudonimiseren:

- a. Verwijdering van naam-, adres-, en woonplaats- (NAW)-gegevens;
- b. Verwijdering van UZI- en andere direct herleidbare codes;
- c. Verwijdering van de naam- en adresgegevens van de huisartspraktijk;
- d. Verwijdering van de naam van de zorggroep.

De TTP kent zowel op huisarts- als op patiëntniveau een uniek gerandomiseerd nummer toe, en gebruikt dit als sleutel voor eventuele identificering na uitlevering van de data. Zowel huisartsen als patiënten zijn na pseudonimisering niet meer direct herleidbaar door onderzoekers. De route om contact te zoeken met

huisartsen of patiënten loopt via de TTP, met enkele uitzonderingen voor het benaderen van huisartsen zoals hieronder beschreven in 4.2.3.

Wanneer een subset die specifiek is samengesteld voor een onderzoek wordt uitgegeven door de ELAN datamanager, wordt deze opnieuw gepseudonimiseerd. Deze stap voorkomt dat subsets die voor verschillende projecten zijn uitgegeven alsnog gekoppeld kunnen worden. Deze tweede pseudonimisering houdt in dat er een nieuw gerandomiseerd nummer wordt toegekend aan patiënten en huisartsen in een onderzoekssubset.

4.2.2. Depseudonimisering

Via het depseudonimiseringsproces kan een patiënt of huisarts identificeerbaar gemaakt worden, bijv. voor het terugkoppelen van individuele informatie of voor inclusie in een cohortstudie. Via depseudonimisatie kunnen individuen weliswaar geïdentificeerd worden, maar identificerende gegevens worden niet gedeeld met de onderzoeker. Depseudonimisatie gebeurt via de TTP en bestaat uit de volgende stappen:

- a. Koppeling van het unieke onderzoeks-ID aan het gerelateerde pseudo-ID (door ELAN datamanager).
- b. Koppeling van het unieke pseudo-ID aan het gerelateerde BSN (door TTP).
- c. (Optionele) koppeling van identificerende gegevens (bijv. NAW-gegevens) d.m.v. het BSN.
- d. (Optionele) benadering van geïdentificeerde patiënt via huisarts.

4.2.3. Depseudonimisering huisarts door datamanager ELAN

De ELAN datamanagers kunnen via een apart sleutelbestand, aangeleverd door STIZON, de huisartscode depseudonimiseren. Dit mag alleen voor goedgekeurde projecten, ofwel na toestemming van de ELAN-huisartsen stuurgroep. Het betreft benaderen van huisartsen voor toestemming tot deelname aan patiëntgebonden onderzoek (zie procesbeschrijving in 5.2.1.), of wanneer huisartsen zelf een expliciete vraag hebben -en dus toestemming geven- tot feedback op de gegevens van de eigen praktijk.

4.3. Aanbod

De ELAN-data wordt voor onderzoekers inzichtelijk gemaakt d.m.v. een data portal. In dit portal wordt alleen metadata weergegeven, niet de data zelf. Dit houdt in dat informatie over de data direct voor onderzoekers toegankelijk is, maar dat een studie-aanvraag goedgekeurd moet worden voordat de data zelf toegankelijk gemaakt wordt. Voorbeelden van metadata zijn het aantal patiënten per leeftijdscategorie, of een overzicht van welke regio's met hoeveel patiënten zijn vertegenwoordigd. Het data portal voorziet in een mogelijkheid om direct een data-aanvraag op te stellen, inclusief specifieke query. Na versturing wordt een dergelijke data-aanvraag behandeld, eventueel goedgekeurd, en de gevraagde data gedeeld met de onderzoeker.

4.3.1. Metadata

De metadata bevat de volgende informatie:

- a. De structuur van de data;
- b. Beschrijvingen van de structuur met als doel variabelen inzichtelijk te maken;

- c. Het aantal rijen dat beschikbaar is per dataset.
- d. De datum waarop de data verzameld is.
- e. Eigenaarschap van de data.

4.3.2. Honorering

Onderzoekers dienen hun data-aanvraag in via een online formulier. Hierop vermelden zij de motivering en onderzoeksvragen van hun onderzoek, en specificeren zij de benodigde data. De ELAN-huisartsen stuurgroep, bestaande uit o.a. huisartsen, epidemiologen en dataspecialisten, beoordeelt deze aanvraag en neemt een besluit t.a.v. honorering.

4.3.3. Uitgifte

Als een dataverzoek gehonoreerd is, wordt de onderzoeker een Data Transfer Agreement (DTA) gestuurd. Deze overeenkomst stelt duidelijke voorwaarden aan het gebruik van de ELAN-data:

- De data wordt alléén gebruikt voor de onderzoeksvraag zoals gesteld in de aanvraag
- Data wordt alleen verstrekt d.m.v. met een wachtwoord beveiligde bestanden en/of verzonden via beveiligde verbindingen
- Data mag niet verder verspreid worden
- De onderzoeker is verantwoordelijk voor de technische en fysieke beveiliging van de data, opdat alleen geautoriseerde gebruikers hier gebruik van kunnen maken
- ELAN levert de gegevens 'zoals ze zijn', en geeft geen garanties t.a.v. volledigheid of geschiktheid voor een bepaald gebruik.
- LUMC/PHEG/ELAN is niet verantwoordelijk voor eventuele schade voor de onderzoeker ten gevolge van gebruik van de data.
- De onderzoeker moet jaarlijkse updates verstrekken over de voortgang van het onderzoeksproject
- Verstrekte data wordt gebruikt in samenwerking met leden van een LUMC/PHEG studiegroep, publicaties worden vooraf beoordeeld door tenminste een huisarts-lid van deze groep
- In publicaties wordt minimaal vermeld dat *'ELAN-GP-data were provided by the Extramural LUMC academic Network (ELAN) of the Leiden University Medical Center (LUMC) after approval by the GP steering committee'*
- Als de analyses afgerond zijn, worden alle datasets en syntaxen terug geleverd aan de LUMC/PHEG studiegroep om zo reproduceerbaarheid te waarborgen

De onderzoeker retourneert het ondertekende DTA, waarna de opgevraagde subset door een datamanager wordt samengesteld. Daadwerkelijke uitgifte aan de onderzoeker gebeurt bij voorkeur via een dataplatform, waarop de onderzoeker zijn of haar analyses kan uitvoeren. De data op dit platform wordt na afronding van het onderzoek verwijderd. Als een dergelijk dataplatform niet beschikbaar is voor een bepaald onderzoek, voldoet de beveiligde LUMC-omgeving, waar m.b.v. vooraf geïnstalleerde software statistische analyses kunnen worden uitgevoerd. Deze omgeving is alleen voor per studie geautoriseerde onderzoekers toegankelijk, en ook hier wordt de data na afloop van het onderzoek verwijderd.

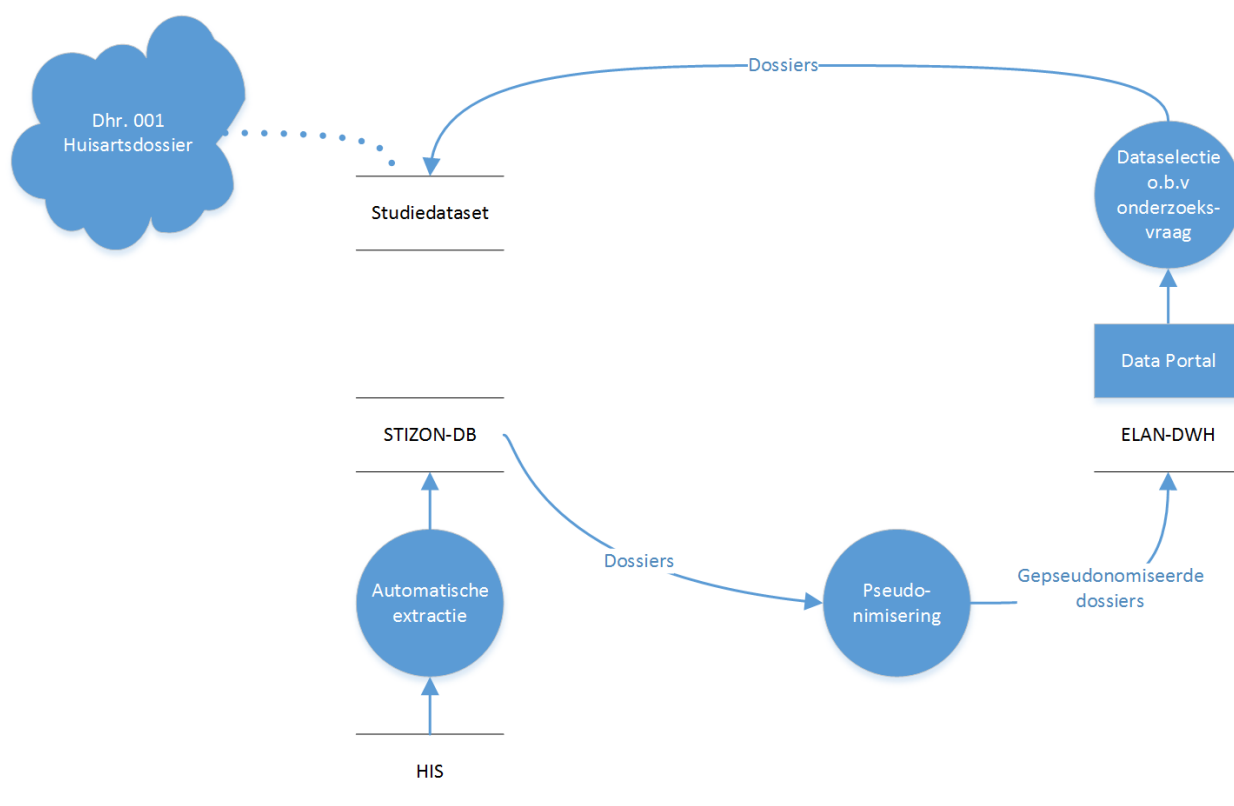
4.3.4. Archivering

Data die gebruikt wordt voor onderzoek moet uit juridische overwegingen vijftien jaar lang bewaard worden. In het geval van subsets die door ELAN worden uitgegeven, wordt data gedurende deze periode opgeslagen in de Data Vault van het LUMC. Deze data zijn niet toegankelijk zonder uitdrukkelijke toestemming van de ELAN-huisartsen stuurgroep. Voor informatie over de veiligheid van de opslag, zie het hoofdstuk Veiligheid.

5. Onderzoek

Met data uit het ELAN-GP datawarehouse kunnen verschillende soorten onderzoek gefaciliteerd worden. In deze sectie wordt de route voor een aantal veelvoorkomende vormen van onderzoek uiteengezet.

5.1. Onderzoek met gepseudonimiseerde data

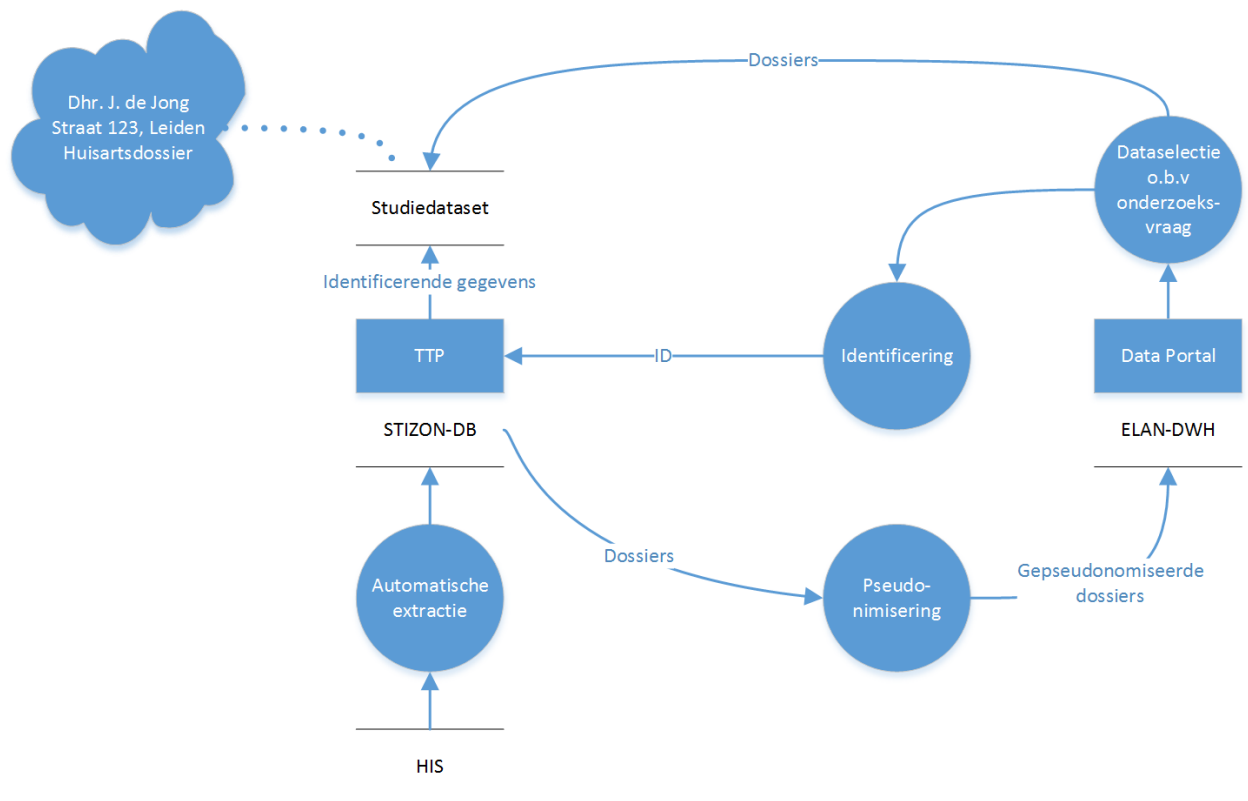


Figuur 4: Route voor onderzoek met gepseudonimiseerde data.

In Figuur 4 wordt de route voor data-uitgifte geïllustreerd voor onderzoek waarin gebruik wordt gemaakt van gepseudonimiseerde data. Data wordt geëxtraheerd uit de HIS'en, opgeslagen in de STIZON-database, en gepseudonimiseerd voor plaatsing in het ELAN-GP datawarehouse (zie de secties over Infrastructuur en Pseudonimiserings voor meer informatie).

De metadata wordt inzichtelijk gemaakt voor de onderzoeker in het data portal, van waaruit een data-aanvraag voor een onderzoek (incl. gespecificeerde query) ingediend kan worden. Na honorering wordt de subset behorende bij de aanvraag gedeeld met de onderzoeker. Deze studiedataset bevat gepseudonimiseerde data en is dus niet op huisarts- of patiëntniveau herleidbaar.

5.2. Onderzoek met geïdentificeerde patiënten of huisartsen



Figuur 5: Route voor onderzoek met geïdentificeerde patiënten of huisartsen.

De route voor onderzoek waarin patiënten of huisartsen identificeerbaar zijn, zoals cohortstudies of terugrapportage van onderzoeksresultaten op praktijkniveau wordt geïllustreerd in Figuur 5. Het begin van het proces is identiek aan het eerder beschreven proces met gepseudonimiseerde data. Als een data-aanvraag echter is goedgekeurd, wordt de studiedataset aangevuld met identificerende informatie. Dit proces verloopt via de TTP en wordt nader beschreven in de sectie over Depseudonimisering. Let wel dat voor deze vorm van onderzoek zowel toestemming op praktijkniveau (van de betrokken huisartsen) als toestemming op individueel niveau (informed consent van patiënten) nodig is.

5.2.1. Voorselectie huisartsen t.b.v. patiëntgebonden onderzoek.

Het ELAN datawarehouse kan gebruikt worden als steekproefbestand t.b.v. patiëntgebonden onderzoek: als startpunt voor het selecteren van potentiële deelnemers, die uit naam van de huisarts kunnen worden aangeschreven voor een specifiek project. Nadat een aanvraag is goedgekeurd, analyseert de datamanager van ELAN in het Data Portal welke huisartsen (huisartsnummers) geschikt zijn voor benadering door een realistische inschatting te maken van de te verwachte instroom. In dit geval zal ELAN zelf, zonder tussenkomst van de TTP, toestemming aan huisartsen vragen. Hiervoor krijgen de datamanagers van ELAN de beschikking tot een koppellijst met huisartsnummers en NAW-gegevens via STIZON. Indien van toepassing zal na toestemming van de huisarts depseudonimisering van patiënten via STIZON worden ingezet.

