



HERACLES

Van mythe naar realiteit de hobbelige weg naar een lerend gezondheidssysteem

Lessen van het HERACLES project uit de praktijk

Inhoudsopgave

Abstract	3
Inleiding	4
Data kwaliteit & organisatie	6
Techniek	8
Governance en organisatie	9
Aanbevelingen voor VWS (door de leerpunten van Heracles) als input voor de uitvoeringswet EHDS	12
Aanbevelingen voor de toekomst, voor de consortiumleden van HERACLES en alle onderzoekers in het algemeen	14
Conclusie en discussie	15
Auteurslijst	16

Abstract

In het HERACLES project ontwikkelen 13 verschillende organisaties waaronder data leveranciers, farmaceuten, oncologie- centra, technologiebedrijven en een zorgverzekeraar, een data infrastructuur om data op een veilige en controleerbare herbruikbaar te maken. Het project is in oktober 2022 van start gegaan en het unieke van HERACLES is dat gebruik wordt gemaakt van databronnen die gecombineerd inzicht geven in het volledige traject dat een patiënt doorloopt en waarbij gebruik wordt gemaakt van echte patiënt data. Het project gebruikt technieken die 'trust-by-design' zijn, data bij de gegevenshouder/bron laat en analyse over meerdere bronnen mogelijk maakt.

De belangrijkste conclusies van het HERACLES project zijn:

1. **Data kwaliteit is essentieel** en dat begint bij het vastleggen bij de bron. Dit is belangrijk voor onderzoek met individuele datasets en helemaal belangrijk voor onderzoek met gecombineerde data uit meerdere bronnen. Door de data - die aan de bron zo veel mogelijk eenmalig en gestructureerd worden vastgelegd - te standaardiseren volgens de internationale definities, wordt hergebruik van data makkelijker, efficiënter en tegen een acceptabele administratieve last. Let op: de administratieve last mag niet bij de zorgverlener terechtkomen.
2. **Veel data zijn niet op de patiënt gericht** maar proces gedreven georganiseerd en daardoor slecht toepasbaar voor een lerend zorgsysteem. Doordat de klinische data beperkt wordt vastgelegd vanuit het patiënt perspectief is het lastig tot onmogelijk om te leren van het traject wat een patiënt doorloopt: van huisarts, apotheek tot behandelingen en nazorg;
3. **De dekkinggraad is al snel onvoldoende** als voor dezelfde groep patiënten uit meerdere bronnen data gecombineerd wordt, waarbij niet elke bron een 100% dekking heeft. Dit lijkt triviaal maar is praktisch gezien een probleem waar we in het HERACLES project tegen aangelopen zijn en minder dan tien procent uiteindelijk van de informatie overhielden. Dit probleem is waarschijnlijk lastig op te lossen, maar in HERACLES werd dit pas duidelijk na een aanzienlijke hoeveelheid werk. Door dit overzicht sneller beschikbaar te maken, kunnen we in de toekomst veel inspanning besparen;
4. **Data koppelingen zijn een juridische en organisatorische uitdaging**, en dat ondanks de technische oplossingen zoals data spaces en Privacy Enhancing Technologieën, die gecontroleerde toegang met behoud van privacy mogelijk maken. HERACLES heeft gedemonstreerd hoe nieuwe dataspace standaarden gecombineerd kunnen worden met encryptie technologie zodat de gevoelige informatie bij de bron blijft en toch gebruikt kan worden voor analyse doeleinden. Hiermee kan technisch gezien een data soevereine en schaalbare oplossing worden geboden maar dat vergt investeringen door de datahouders, voor wie de business case moeilijk te maken valt. Bovendien leidt een gebrek aan kennis over de juridische aspecten van gebruikte data-minimalisatie en data-sleutels tot inertie zelfs indien er grote welwillendheid is bij de betrokken partijen;
5. **Meer informatie over de beschikbare data** is nodig om vroegtijdig te kunnen bepalen hoeveel de data set kan bijdragen aan het onderzoek. Vindbaarheid van data, is een kernaspect van de FAIR principes en blijft tot op heden beperkt tot een beschrijvingen van de data en de voorwaarden voor gebruik. Door vooraf indicatief aan te geven hoeveel data er beschikbaar komt bij het combineren van datasets uit verschillende bronnen, kan veel tijd en moeite bespaard worden als de verwachte omvang minimaal blijkt te zijn, als gevolg van de eerste vier punten. Bovendien is het belangrijk om de context van de verschillende bronnen te kennen, zodat de gegevens correct geanalyseerd kunnen worden voor het onderzoek.

Naast HERACLES werken ook andere initiatieven zoals de European Health Data Space (EHDS) via zogenoemde HDABs (Health Data Access Body) en Health-RI aan oplossingen om secundair gebruik van zorgdata in Nederland mogelijk te maken. Hoewel de potentie van gegevens enorm is, blijft deze onvoldoende benut door versnippering, technische obstakels en verouderde governance-structuren. Derhalve is het doel van dit Interim Advies rapport om de geleerde lessen vanuit HERACLES te delen zodat behalve het HERACLES project ook andere projecten hiervan kunnen profiteren.

Aan de technologie ligt het niet. Die is beschikbaar en wordt, zij het beperkt, ook al in de praktijk toegepast. De waarde van de beschikbare data hangt grotendeels af van hoe deze is opgeslagen en daarin is nog veel te winnen. Als voor medische gegevens de data kwaliteit en vindbaarheid verbetert, patiëntgericht opgeslagen en ontsloten wordt via een gemeenschappelijke koppelmethode dan pas kunnen medische gegevens echt waarde opleveren voor een lerend zorgsysteem en zodoende voor patiënten, de zorg en de maatschappij.

Inleiding

In februari 2022 verscheen het OECD rapport 'Towards an integrated Health Information System in the Netherlands' waarin het gezondheidsinformatiesysteem van Nederland in kaart werd gebracht. De OECD onderzocht dit omdat verschillende landen zich zorgen maakten of ze de juiste data infrastructuur hadden om gezondheidsstatistieken te produceren, en om gezondheidsuitkomsten en de kwaliteit van zorg goed te kunnen monitoren. Nederland was het eerste land dat onder de loep werd genomen daarbij was het oordeel over het gezondheidsinformatiesysteem niet mals ([zie OECD rapport](#)). De belangrijkste barrière voor een dergelijk systeem dat werd gevonden is de fragmentatie van het Nederlandse zorgsysteem die een institutionele barrière vormt voor gegevensuitwisseling en deling. Verder werd het gebrek aan afstemming en een gemeenschappelijke interpretatie van wet- en regelgeving als uitdaging gezien van veilige gegevensdeling.

Op het moment van verschijnen van het OECD rapport was het HERACLES project al in ontwikkeling. Het doel van dit project is het toegankelijk maken van gefedereerde, gekoppelde oncologisch databronnen om zo inzichten te creëren die de impact van kanker kunnen reduceren. HERACLES vertrekt vanuit dezelfde constatering dat (oncologische) databronnen versnipperd zijn over verschillende organisaties (en soms ook binnen organisaties) (zie ook box 1). Daarnaast focust HERACLES zich op technieken die 'trust-by-design' zijn; data bij de gegevenshouder laat en analyse over deze bronnen wel mogelijk maakt. Uitgangspunt binnen HERACLES is daarmee data soevereiniteit; de gegevenshouders van de data houden controle over hun eigen databron. EHDS gaat meer ruimte geven om data makkelijker te kunnen hergebruiken voor onderzoek, waarin het uitgangspunt zal verschuiven van datasoevereiniteit naar datasolidariteit. In het HERACLES project brengen verschillende organisaties zoals data leveranciers, farmaceuten, een zorgverzekeraar, oncologie- centra, tot technologiebedrijven, data op een veilige en controleerbare manier samen. Het project is in oktober 2022 van start gegaan en het project loopt tot september 2025. Het unieke van HERACLES is dat gebruik wordt gemaakt van databronnen die gecombineerd inzicht geven in de volledige patiënt-reis en waarbij gebruik wordt gemaakt van echte patiënt data.

Box 1 HERACLES: welke problemen willen we oplossen?

Door een sterk gefragmenteerd zorgsysteem in Nederland worden de meeste verzamelde gezondheidsgegevens versnipperd vastgelegd en opgeslagen. Gelukkig zijn voor hergebruik van deze versnipperde gezondheidsgegevens meerdere technische oplossingen beschikbaar. Echter, deze technieken kunnen alleen worden ingezet als de gezondheidsgegevens ook daadwerkelijk voldoen aan de FAIR-principes (Findable, Accessible, Interoperable, Reusable). Om de gezondheidsgegevens van een patiënt aan elkaar te kunnen koppelen is toegang tot die gegevens via een koppelsleutel een belangrijk voorwaarde. Nieuwe wetgeving zoals EHDS en Wegiz vergemakkelijken hopelijk die benodigde goedkeuring en daarmee de toegang tot gezondheidsgegevens voor onderzoekers en andere belanghebbenden. Naast betere toegang is er ook behoefte aan een gestandaardiseerd systeem voor koppelsleutels om de verschillende gegevens te koppelen aan een unieke patiënt.

Om hergebruik van gezondheidsgegevens voor onderzoek te vergemakkelijken onderzoekt HERACLES:

- *De fragmentatie van patiëntgegevens:* gezondheidsgegevens van individuele patiënten worden gefedereerd opgeslagen in/over verschillende systemen en databases. Hierdoor zijn heldere afspraken en beschikbare technologie nodig om een compleet beeld te krijgen van zorgtrajecten, bijvoorbeeld om te onderzoeken welke factoren een voorspellende waarde hebben voor bepaalde behandeluitkomsten. HERACLES onderzoekt hoe dit in de praktijk toch kan werken.
- *De techniek:* HERACLES onderzoekt of en hoe Data Spaces en PET technologieën (zie box 2) hergebruik van gezondheidsgegevens kunnen vergemakkelijken.
- *Het koppelen van de verticaal gepartioneerde data:* de verticaal verdeelde data - in deze context betreft het patiëntgegevens die in de loop van de tijd op verschillende locaties zijn verzameld – zijn van grote waarde voor bijvoorbeeld het in kaart brengen van mogelijkheden tot vroeg-diagnose, chronische ziekten, preventieprogramma's of de effecten van diagnostiek en medicatie. HERACLES onderzoekt de governance en technische oplossingen om verticale data koppelingen mogelijk te maken.

Om de doelmatigheid en efficiëntie van de zorg te kunnen bestuderen is het nodig om de zorg en de uitkomsten van die zorg, die vastgelegd zijn in gezondheidsdata van individuele patiënten, te volgen over langere tijd. Daarvoor is het belangrijk te weten hoe data over organisaties zijn verdeeld. In data management termen of het verticaal of horizontaal gepartioneerd is (zie box 2).

Box 2 data partitionering

Voor gegevensbeheer en de analyses spelen het type van partitionering — horizontaal en/of verticaal — een cruciale rol indien datasets uit verschillende bronnen moeten worden gecombineerd:

- *Horizontale partitionering*, waarbij twee of meer organisaties dezelfde soort gegevens (parameters) registreren. Het samenvoegen, wat geanonimiseerd plaats kan vinden, is gewenst om informatie over een groter aantal patiënten te krijgen. Dit laatste is heel relevant voor onderzoek naar zeldzame ziektes, maar ook als vergelijkbare data over veel bronnen is verspreid;
- *Verticale partitionering*, waarbij verschillende soorten gegevens (parameters) van eenzelfde groep individuen verdeeld zijn over verschillende bronnen. Bijvoorbeeld als de gezondheidsgegevens zijn opgeslagen bij huisartsen, binnen een ziekenhuis, de Sociale Verzekeringsbank en een zorgverzekeraar. Om een historisch overzicht en/of patient-journey te kunnen reconstrueren, gaat de in voorbereiding zijnde oplossing voor een gemeenschappelijke sleutel (“identifier”) zeer welkom zijn. Deze oplossing zal het kunnen relateren van de verschillende soorten gegevens aan dezelfde patiënt vergemakkelijken.

Het HERACLES-project loopt nu twee jaar en is onderweg op verschillende barrières gestuit. De lessen die we hierbij hebben geleerd, willen we graag ook buiten dit project middels dit rapport delen met name met partijen die het hergebruik van data in Nederland bevorderen, met name VWS, Health-RI, HDAB-NL, DHD, NICTIZ, NFU en NVZ.

De lessen in dit rapport zijn gebaseerd op interviews met de consortiumleden van HERACLES. Voor een goed overzicht zijn de lessen opgedeeld in drie categorieën: data kwaliteit & organisatie, techniek en governance & organisatie.

Data kwaliteit & organisatie

Doorgaans worden “aan de bron” de gezondheidsgegevens decentraal vastgelegd om de medisch specialistische zorg te organiseren voor de patiënt. Deze gegevens worden ook gebruikt voor de declaraties van de geleverde inspanningen. Daarnaast bestaan er in Nederland inmiddels vele registers met gezondheidsgegevens die meestal handmatig vastgelegd worden. In de praktijk betekent dit dat gezondheidsgegevens vrijwel altijd meervoudig moeten worden ingevoerd in meerdere systemen en meerdere databases, die onderling ook verschillend zijn opgezet met soms zelfs verschillende omschrijvingen van dezelfde data-elementen. Naast de registers worden er in Nederland voor onderzoek medische gegevens vertaald naar declaratiegegevens om vervolgens daaruit eindpunten af te leiden. Al deze acties hebben effect op de data kwaliteit. Om als onderzoeker de data goed in te kunnen zetten, moet worden begrepen hoe de data tot stand is gekomen en daarom moeten deze stappen goed worden vastgelegd. Vervolgens begint het proces voor een onderzoeker met het selecteren van de data die geschikt is voor hun studie, waarbij de kwaliteit en relevantie zorgvuldig worden beoordeeld.

In het HERACLES project is gekozen voor twee oncologische use-cases (zie onder) die vereisen dat de gezondheidsgegevens van patiënten uit verschillende databronnen dienen te worden gekoppeld. Deze use-cases zijn gekozen omwille van één van de doelstellingen van het project. Zijn gefedereerde data over verschillende organisaties herbruikbaar voor maatschappelijk relevante onderzoeksvragen? Het samenbrengen van deze data, die in diverse formaten, soms ongecodeerd, zijn opgeslagen, kunnen tot nieuwe waardevolle inzichten leiden. Vooral toegang verkrijgen tot de juiste data en daarna de koppelingen tussen de verschillende bronnen leggen, bleken uiteindelijk het meest tijdrovend. Verkorting van dit proces, is binnen project HERACLES op het moment van het schrijven van deze tussentijdse evaluatie ook nog niet gelukt.

Les 1. Zorg dat de data écht FAIR zijn.

FAIR-compliance in datasets: Om gegevens vindbaar en herbruikbaar te maken, moet de inhoud van datasets worden beschreven in machine-leesbare formaten. Belangrijke aspecten zijn:

- “Beloon” instanties om de data FAIR te maken voor onderzoek.
- Vertaal de relevante ingevoerde gestructureerde en ongestructureerde Landelijke Basisregistratie Ziekenhuiszorg data (volgens Thesaurus Zorg en Welzijn) naar een leesbare taal voor onderzoek door gebruik te maken van standaarden, zoals internationaal gestandaardiseerde classificatiesysteem van diagnoses (ICD10) en internationaal medisch terminologiestelsel (SNOMED CT).
- Beoordeel de gezondheidsgegevens op mate van kwaliteit voor hergebruik (compleetheid, consistentie, correctheid, uniciteit, consistentie, representativiteit, etc.)
- Gebruikslimieten en licenties.
- Metadata en datastructuren (zoals vocabularia en classificatieschema's).
- Bijbehorende gebruiksvoorwaarden om automatische toegang tot en overdracht van gezondheidsgegevens tussen partijen mogelijk te maken. Het liefst in een machine leesbaar formaat.

Les 2. De kwaliteit van de beoogde data wordt pas echt zichtbaar als er met echte data wordt gewerkt.

Uit de interviews kwam naar voren dat na visuele inspectie van de data er aan het licht kwam dat er mogelijk ‘gaten’ in de patiëntreis zaten. Dit was niet zichtbaar in de geaggregeerde data die daarmee waarschijnlijk geen valide resultaat zouden geven. Een voorbeeld: uit de data bleek regelmatig dat chemotherapie-toedieningen kort na elkaar plaatsvonden, wat in de praktijk onwaarschijnlijk is vanwege mogelijke toxiciteit. Zonder context (bijvoorbeeld uitstel vanwege afwijkende vitale functies of bloedwaarden) en inhoudelijke kennis is dit lastig te duiden en kan dit leiden tot verkeerde inzichten als de data toch wordt gebruikt. Het opschonen van de data leidde tot een verdere reductie in het aantal beschikbare datapunten voor analyse.

Bij het koppelen van verschillende databronnen zijn data stewards met domeinkennis hard nodig. Missende data kunnen dan bijvoorbeeld nog achterhaald worden, of mogelijke fouten worden gecorrigeerd of indien nodig gevalideerd. De belangrijkste aanbeveling is dat klinische data ook patiëntgericht wordt opgeslagen naast proces georiënteerde opslag. Dit behoeft niet alleen technische maar ook systeemwijzigingen. Essentieel voor de acceptatie van deze wijzigingen is, dat het niet leidt tot een hogere administratieve last voor het personeel in de zorg.

Les 3. Mate van versnippering van waardevolle gezondheidsdata bleek groter dan bij aanvang van het project gedacht.

Bij het creëren van deze basis-set kwamen twee problemen aan het licht. Allereerst ontbraken er toch essentiële data zoals volledige overzichten van verstrekte medicatie. Het tweede probleem was dat tot onze verrassing slechts een klein deel van de dataset—een beperkt aantal individuen—daadwerkelijk te aan elkaar te relateren was. Dit aantal was te klein om statistisch betrouwbare uitspraken te doen over de patiëntreis. Zo zijn er voor oncologie patiënten vele routes van medicatie uitgiftes (openbare apotheek, ziekenhuisapotheek, politheken) wat tot fragmentatie leidt en is voor een compleet beeld vele bronnen nodig. Veel organisaties werken hard aan het koppelen van gegevens met een versleuteld BSN. In dit geval zou dat niet direct betere resultaten opleveren, omdat niet alle partijen het BSN standaard opslaan en de beschikbare gegevens niet compleet zijn.

Begin niet bij organisaties of beschikbare databronnen, maar bij de vraagstelling en de “reis” die de meeste individuen doorlopen in het zorglandschap. Zo kan sneller worden bepaald welke data en bronnen nodig zijn en zinloze koppelingen worden voorkomen. Bestudeer zo vroeg mogelijk in het proces de kwaliteit van de data, zoals de dekking over de verschillende patiëntengroepen en mogelijke missende data.

Techniek

Parallel aan het traditioneel koppelen van databronnen zoals hiervoor beschreven, wordt in het HERACLES project ook gewerkt aan een technische oplossing om data op een data space te ontwikkelen. De kern daarbij is dat data bij de datahouders blijven en dat ofwel algoritmes in staat zijn de data te bezoeken ofwel enkel versleutelde data uit de bron worden gebruikt, waarbij gevoelige informatie wordt verhuld. Hiervoor worden International Data Spaces (IDS), een architectuur gericht op het veilig, vertrouwelijk en gecontroleerd delen van data tussen organisaties en verschillende Privacy Enhancing Techniques (PETs), bijvoorbeeld secure Multi Party Computation of Federated Learning, ingezet.

Box 3 – dataspace en Privacy Enhancing Technologies

Volgens het [Data Space Support Centre \(DSSC\)](#) is een dataspace een infrastructuur die data transacties tussen verschillende partijen binnen een data-ecosysteem mogelijk maken op basis van het governance-kader van die dataspace. Voor een veilige analyse zijn verschillende Privacy Enhancing Technologieën beschikbaar. Andere sectoren zoals de financiële dienstverlening maken hiervan al jaren gebruik.

Voor het HERACLES project betekent dit het volgende:

- Gegevens worden op een beveiligde manier geanalyseerd.
- Alleen versleutelde gegevens en/of geaggregeerde data worden beschikbaar gemaakt voor hergebruik.
- Er wordt by-design afgedwongen dat er controle en privacybescherming is, terwijl transparante en controleerbare gegevenstransacties worden gewaarborgd.

Deze toepassing binnen de zorg biedt belangrijke voordelen op het gebied van gegevensbeheer, privacybescherming, soevereiniteit en schaalbaarheid. Tegelijk vraagt het om het gebruik van nieuwe technologieën voor analyses en het maken van nieuwe afspraken tussen samenwerkende partijen binnen een datasamenwerking. Het succes is sterk afhankelijk van de mate van harmonisatie en standaardisatie van alle datahouders, omdat geen enkele partij controle krijgt over de gegevens van anderen. Dit staat in contrast met de traditionele werkwijze, waarbij gegevens volledig in handen komen van een partij die de analyses uitvoert en eigen, vaak unieke standaardisaties kan opleggen die niet overeenkomen met de internationale coderingen of andere initiatieven.

Les 4. Infrastructuurleveranciers moeten de ruimte krijgen om onderscheidende diensten te leveren, terwijl harmonisatie essentieel blijft om schaalbaarheid en eenduidigheid te garanderen.

Privacy Enhancing Technologies (PETs) (zie box 3) bieden veelbelovende mogelijkheden voor nieuwe toepassingen, maar bevinden zich nog in een vroeg ontwikkelingsstadium binnen de zorg in vergelijking met technologieën die gebaseerd zijn op centrale dataverzameling en -analyse. Verschillende PET leveranciers zijn nu hun diensten aan het ontwikkelen. Datagebruikers zijn aan het onderzoeken of en waarvoor deze diensten behulpzaam kunnen zijn. Deze gebruikers willen zo min mogelijk 'last' hebben van techniek, maar zijn vaak wel gebaat bij een privacy-by-design oplossing door onderzoek te kunnen doen wat zonder PETs niet mogelijk zou zijn. Het kunnen gebruiken van verschillende oplossingen afhankelijk van de vraag vereist dat de PET oplossingen vergelijkbare input gebruiken en vergelijkbare output leveren. Deze harmonisatie van verschillende PET-oplossingen zou nu moeten worden ingezet.

Bij het ontwerpen van een data-infrastructuur voor de European Health Data Space (EHDS), waar eenvoud essentieel is, wordt het potentieel van de PETs slechts beperkt beschreven, ondanks hun toegevoegde waarde op het gebied van privacybescherming en schaalbaarheid. Wij pleiten daarom voor versneld onderzoek naar de inzetbaarheid van de PETs, met een nadruk op oplossingen die interoperabel zijn. Voor de PET's is eenheid van taal een van de belangrijkste voorwaarde. De harmonisatie daarvan kan alleen worden afgedwongen door de stelselverantwoordelijke.¹

Governance en organisatie

Gefedereerde ecosystemen vormen de basis van moderne data spaces, zoals het European Health Data Space. Deze ecosystemen en onderliggende wetgeving bieden organisaties de mogelijkheid om samen te werken en data te delen voor onderzoek, terwijl ze zelf controle behouden over hun gegevens. Dit model brengt complexe governance-uitdagingen met zich mee voor de implementatie van EHDS-wetgeving. Vooral datatoegang, toestemmingsbeheer, naleving van regelgeving en het behoud van ethische waarden vereisen een robuuste aanpak om schaalbaarheid en efficiëntie te waarborgen.

¹ Adviezen over verdere harmonisatie van een internationaal erkend onderzoek model voor Federated Learning zoals "Observational Medical Outcomes Partnership Common Data Model" (OMOP CDM) kunnen ingewonnen worden bij het internationale netwerk van ontwikkelaars genaamd Observational Health Data Sciences and Informatics (OHDSI).

Bestaande raamwerken vertrouwen vaak op handmatige processen die gevoelig zijn voor incompleetheid, fouten en vertragingen, wat de noodzaak benadrukt van geavanceerdere, geautomatiseerde oplossingen die onderzoek voor innovatie, kwaliteit, capaciteit en beleidskeuzes kunnen voeden. Benodigd is een digitale infrastructuur die geautomatiseerde governance-mechanismen bevat en die data spaces ondersteunt om continue naleving te garanderen. Ook over governance heeft HERACLES verschillende lessen geleerd.

Les 5. Don't Fix what's not broken: Partijen gaan toch liever 1-op-1 relaties aan dan toetreden tot een data space.

HERACLES partners zijn samen gaan werken aan een decentrale dataoplossing vanuit de visie dat dit een goed idee is. Toch blijken partijen bij een specifieke use-case snel terug te pakken op de klassieke manier van toegang, namelijk door onderlinge afspraken te maken. Als bestaande oplossingen werken in de praktijk, voelt men geen noodzaak om te investeren in nieuwe technologie. Dat brengt ons bij het volgende punt:

Les 6. De olifant in de kamer - Worden data spaces wel gebruikt?

Ook in andere Europese landen worden data platforms gecreëerd om data hergebruik te faciliteren zoals ook de EHDS (European Health Data Space) ten doel heeft. Het Zwitserse data-initiatief [SPHN](#) heeft de zaken goed op orde zoals ze hebben laten zien in een HERACLES-webinar. Helaas bleek uit deze presentatie ook dat deze infrastructuur nog nauwelijks gebruikt wordt. Ditzelfde blijkt uit het [Sitra rapport](#).

EHDS zorgt ervoor dat er nu veel geïnvesteerd wordt in systemen om hergebruik van gezondheidsdata mogelijk te maken. Van belang is om te leren van de al bestaande initiatieven en inzicht te krijgen in de vraag waarom deze infrastructuren tot nu toe nauwelijks gebruikt worden en of de introductie van wetgeving en de Health Data Access Bodies hier daadwerkelijk verandering in gaat brengen.

Les 7. Neem de algemene leiding en de relevante afdelingen van de betrokken organisaties vroegtijdig mee in de ontwikkeling (keuzes en afspraken) van een data space.

Zonder draagvlak en een heldere visie vanuit het management van betrokken organisaties in de datasamenwerking blijven initiatieven beperkt. Het is van belang om duidelijk te definiëren hoe gezondheidsgegevens de kwaliteit van zorg kunnen ondersteunen. Zonder deze richting bieden initiatieven geen duurzame oplossingen. Gecommitteerde vertegenwoordigers van de betrokken organisaties zijn cruciaal voor het slagen van projecten rondom data beschikbaarheid zoals HERACLES, maar kunnen niet zonder brede steun voor de juridische en technische bevoegden die nodig zijn voor het doen slagen van deze projecten. In het geval van het HERACLES wordt er gepoogd een werkende use-case bij enkele partners op te tuigen. Daar is te beperkt op de IT-afdelingen gerekend die onvoldoende tijd en budget blijkt te hebben en meer informatie nodig heeft om de technologie te implementeren.

Betrek in een vroeg stadium de IT-afdelingen van de gegevenshouders om processen en de IT te organiseren om snel en flexibel te experimenteren met technologie. Daarbij is het noodzakelijk dat het management van de organisatie het initiatief ondersteunt, de nodige capaciteit regelt voor de IT-afdeling om daadwerkelijk het werk te verrichten.

Les 8. Een schaalbare governance vereist dat zorgpartijen met verschillende belangen gaan harmoniseren in hun data space beleid.

In de governance van een data space dient er van alles te worden geregeld, zoals de inhoud van de dataset, de gebruiksbeperkingen, de licenties, de gegevensverzamelingsmethoden, de gegevenskwaliteit en de vindbaarheid van de gegevens in machine-leesbaar (FAIR) formaat.

Digitale infrastructuren moeten geavanceerde governance-mechanismen bevatten die naleving van regelgeving afdwingen (en misbruik voorkomen) en direct kunnen reageren op verzoeken. Denk aan slimme contracten die automatisch data-deelovereenkomsten uitvoeren die rekening houden met de heterogeniteit van partijen. De betrokkenheid van diverse stakeholders, zoals farmaceuten, technologiebedrijven en zorgverzekeraars, vereist een level playing field. Dit vraagt om standaarden die interoperabiliteit en samenwerking bevorderen en “vendor-lock-in” voorkomen. Om afhankelijkheid van specifieke leveranciers te vermijden, dienen ecosystemen gebaseerd te zijn op open standaarden en gemeenschappelijke architecturen. Dit voorkomt fragmentatie en maakt opschaling of wisselen van vendor eenvoudiger.

Les 9. Creëer in een geharmoniseerde data aanvraagprocedure.

Verschillende data leveranciers en data verwerkers hebben nu allemaal een eigen data aanvraagprocedure. Als data van verschillende leveranciers moeten worden gekoppeld betekent dit een onnodig lange procedure waar soms dingen dubbel of driedubbel worden gedaan. Een geharmoniseerde data aanvraagprocedure kan dit proces versnellen.

Met gebruik van interoperabele tools, zoals slimme contracten, kunnen gegevensdelings-overeenkomsten worden geautomatiseerd. Dit versnelt processen en verhoogt de transparantie binnen het ecosysteem.

Aanbevelingen voor VWS (door de leerpunten van HERACLES) als input voor de uitvoeringswet EHDS

Uit werkpakket 1 (onderzoek)

1. **Selecteer voor de onderzoeksvraag de belangrijkste datapunten en markeer deze binnen een passende registratie.** Voor HERACLES is de passende registratie de Landelijke Basisregistratie Ziekenhuiszorg (LBZ): beroepsverenigingen kunnen voor maatschappelijk relevante onderzoeksvragen een selectie maken van de LBZ-data die inzichten kunnen opleveren via vooraf vastgestelde maatschappelijk relevante onderzoeksvragen op basis van Population, Intervention, Comparison en Outcome (PICO) die:
 - a. Gericht zijn op de karakteristieken van de patiënten die verwezen worden voor medisch specialistische zorg en gekoppeld zijn aan de gestelde diagnose(s). Over wie gaat het en wat is het probleem?
 - b. Gericht zijn op de verrichtingen gedurende patiënt-reis in de zorg. Welke zorg is geleverd?
 - c. Gericht zijn op benodigde capaciteit voor de zorgactiviteiten. Hoe groot was de personele inspanning?
 - d. Gericht zijn op de belangrijkste behandelopbrengst of resultaat van de zorgactiviteiten. Wat heeft het opgeleverd?
2. **Vertaal de gemarkeerde data naar FAIR data:** Laat de partijen in de medisch specialistische zorg de relevante ingevoerde gestructureerde (en ongestructureerde) Landelijke Basisregistratie Ziekenhuiszorg data (volgens Thesaurus Zorg en Welzijn) zoveel als mogelijk geautomatiseerd omzetten naar één machine-leesbare taal voor onderzoek. Maak gebruik van standaarden, zoals bijvoorbeeld het internationaal gestandaardiseerde classificatiesysteem van diagnoses (ICD10) en internationaal medisch terminologiestelsel (SNOMED CT). Harmoniseer deze data eens per kwartaal naar een "Common Data Model" voor de Privacy Enhanced Technologies (PET's), zoals Multi Party Computation (MPC) en Federated Learning (FL).
3. **Data kwaliteit:** Stel de optie in om (minimale) kwaliteitsnormen vast te leggen voor gezondheidsgegevens die hergebruikt worden voor specifieke doeleinden (completeheid, consistentie, correctheid, uniciteit, consistentie, representativiteit, etc.).

Uit werkpakket 2 (techniek)

1. **Creëer een mogelijkheid om data te koppelen:** Maak het mogelijk (en eenvoudig) om de data van patiënten uit verschillende bronnen eenvoudig te kunnen koppelen.
2. **Beleg de verantwoordelijkheid voor data koppeling bij één of meerdere Health Data Access Bodies**
3. **Geautomatiseerde data gateways:** maak automatiseerde ondersteuning mogelijk voor meervoudig gebruik van de gestandaardiseerde data voor onderzoek (uit een gestandaardiseerde kopie van de levende database met de ge-pseudonimiseerde originele gezondheidsgegevens).

Uit werkpakket 3 (governance)

1. **Afsprakenstelsel:** Huidige juridische grondslagen zijn multi-interpretabel, waardoor veel te veel juridisch overleg nodig is tussen alle partijen voor goedkeuringen om onderzoek te kunnen doen met data uit de dagelijkse praktijk. Meerdere partijen maken afsprakenstelsels afhankelijk van project of samenstelling van deelnemende partijen. Standaard “formats” en duidelijke vertalingen van wetgeving zouden de verschillende (publieke/private) partijen die onderzoek doen enorm helpen.

Uit werkpakket 4 (toekomst)

1. **Stimuleer het gebruik van de PET's** (zoals Federated Learning en Multi Party Computation) als oplossing voor standaard vraagstellingen waarvoor meerdere gevoelige gezondheidsgegevens nodig zijn en gekoppeld dienen te worden om antwoord te verkrijgen op terugkerende onderzoeksvragen.

We verwijzen ook graag naar de [adviezen over FAIR data](#) die gegeven zijn in het kader van de pandemische paraatheid.

Aanbevelingen voor de toekomst, voor de consortiumleden van HERACLES en alle onderzoekers in het algemeen

Als Datahouder:

1. Heb een heldere en ambitieuze data en AI strategie en draag deze uit.
2. Zorg dat de data vindbaar zijn voor onderzoekers en dat helder is wat de kwaliteit van de dataset is en onder welke voorwaarden deze data beschikbaar zijn. Zorg dat de termen worden beschreven met ontologieën en voeg rijke meta-data toe, zodat helder is hoe de data is verkregen en de bruikbaarheid kan worden bepaald.
3. Betrek bij het koppelen van verschillende databronnen (binnen een organisatie) data stewards met domein kennis.
4. Zorg dat er inzicht komt wanneer data platformen wel en niet gebruikt worden en trek hier lessen uit om te komen tot een platform dat breed inzetbaar is.
5. Zorg dat partijen die nu een rol hebben in het beschikbaar maken van data voor onderzoek, beleid of zorg ook een duidelijke rol en verantwoordelijkheid hebben in een (door EHDS-implementatie) nieuw stelsel.
6. Zorg dat er op korte termijn governance afspraken gemaakt en zorg dat deze een goed technische implementatie kennen.
7. Automatiseer gegevensdelingsovereenkomsten door het gebruik van bijvoorbeeld slimme contracten om processen te versnellen en transparantie te verhogen.

Als datagebruiker:

1. Begin niet eerst met het ontsluiten van beschikbare databronnen maar bij de vragen die gesteld worden en waarvoor (op regelmatig basis) data uit verschillende bronnen nodig zijn.
2. Zorg dat er een goed beeld is van de dekking van de data over patiënt groepen en over de missings binnen de datasets.
3. Gebruik geen data die vanuit het perspectief van betaling verzameld zijn, maar gebruik de brongegevens die vanuit patiënt perspectief zijn vastgelegd (dus op basis van de diagnose en verrichtingen thesauri).
4. Betrek in een vroeg stadium de IT-afdelingen van de gegevenshouders en hun management in de ontwikkelingen van data platformen zodat het initiatief ondersteund wordt en de nodige capaciteit beschikbaar is om daadwerkelijk het werk te verrichten.
5. Betrek bij het koppelen van verschillende databronnen (tussen organisaties) data stewards met domein kennis.
6. Versnel onderzoek naar de inzetbaarheid van PETs, met een nadruk op oplossingen die interoperabel zijn, zodat deze waar nodig kunnen worden ingezet in de EHDS.
7. Blijf alle betrokken partijen informeren over de waarde van data voor zorg en gezondheid in het algemeen, en van een goede data-infrastructuur in het bijzonder. Leg uit dat privacy van patiënten geborgd wordt via organisatorische en technische maatregelen.
8. Zorg dat het duidelijk wordt wanneer een decentrale dataoplossing een goede oplossing is en wanneer niet.

Conclusie en discussie

Veel partijen werken aan oplossingen om secundair gebruik van zorgdata in Nederland mogelijk te maken. Hoewel de potentie van gegevens enorm is, blijft deze onvoldoende benut door versnippering, technische obstakels en verouderde governance-structuren. We signaleren dat voor belangrijke vraagstukken rondom zorg en gezondheid, patiëntgegevens uit meerdere bronnen dienen te worden gecombineerd. Het realiseren van een lerend zorgsysteem vraagt om meer dan wetgeving en technische oplossingen: het vereist een gezamenlijke inspanning, cultuurverandering en een sterke focus op samenwerking tussen diverse partijen. De introductie van de EHDS en de aankomende uitvoeringswet, alsmede de oprichting van een Health Data Access Body zullen ongetwijfeld een belangrijke stimulans zijn in de beweging naar betere data beschikbaarheid. Echter, voor het effectief ontsluiten van gecombineerde patiëntgegevens uit meerdere bronnen, signaleren we een aantal belangrijke uitdagingen. Bij het ontwerp van de HDAB structuur en onderliggende data space, dient aandacht te komen voor de problemen voor vindbaarheid, toegang, ontsluiting, koppeling en kwaliteit van gekoppelde data.

Door robuuste governance en infrastructuur te combineren met een sterk commitment aan FAIR-principes, kunnen data spaces zoals ontwikkeld in het HERACLES-project een blauwdruk vormen voor het lerend zorgsysteem. Zo kunnen we niet alleen voldoen aan de behoeften van vandaag, maar ook de uitdagingen van morgen aangaan.

Deze data spaces zullen privacy moeten borgen door governance en techniek en beide behoeven (inter)nationale standaarden zodat datagebruikers toegang kunnen krijgen tot verschillende datasets met verschillende Privacy Enhancing Technologies. Daarnaast zal de data space zo georganiseerd moeten zijn dat data uit verschillende sets horizontaal en verticaal te koppelen zijn.

De technologie is beschikbaar maar de waarde van de beschikbare data hangt grotendeels af van hoe deze is opgeslagen en daarin is nog veel te winnen. Ten eerste door de data goed te beschrijven inclusief meta-data en ontologieën, maar vooral ook door te zorgen dat de klinische data voldoende wordt vastgelegd vanuit het patiënt perspectief, zodat er over de hele patiënt reis geleerd kan worden. Door de data - die aan de bron zo veel mogelijk eenmalig en gestructureerd worden vastgelegd - te standaardiseren volgens de internationale definities, wordt hergebruik van data makkelijker, efficiënter en tegen een acceptabele administratieve last. Het is daarbij van groot belang dat de extra administratieve last voor de zorgverlener wordt voorkomen, en idealiter zelfs verminderd. Data gaat dan pas echt tijd en waarde opleveren voor de patiënten, zorg en maatschappij. Vanzelfsprekend gebruiken we deze lessen om het HERACLES project te verbeteren. We nodigen ook van harte alle relevante veldpartijen uit om ervaringen uit te wisselen om gezamenlijk de juiste keuzes te maken om data deel infrastructuur succesvol te implementeren in de praktijk.

Auteurslijst

Name	Organisation
Andre Boorsma	TNO
Erik Cornelisse	TNO
Jildau Bouwman	TNO
Sarah van Drumpt	TNO
Simon Dalmolen	TNO
Brigitte Gijsen	IKNL
Dimitris Katsimpokis	IKNL
Gijs Geleijnse	IKNL
Pauline Vissers	IKNL
Michel van den Heuvel	Radboudumc
Olga Schuurbiers	Radboudumc
Merle Bootsma	Radboudumc
Nelleke Ottevanger	Radboudumc
Janneke Walraven	Radboudumc
Josine Kuiper	Pharmo
Ron Herings	Pharmo
Maarten Everts	Linksight
Martine van de Gaar	Linksight
Ton Peters	J&J
Sander Dalhuisen	J&J
Jan Pander	AstraZeneca
Roy Bleumink	Roche
Willem Willems	CZ
Stefan Willems	UMCG
Vincent de Jager	UMCG
Erik Kentie	SURF
Andries	Almende



HERACLES



Radboudumc



Johnson&Johnson



umcg:



HERACLES is een samenwerking tussen iKNL, Stichting Radboud UMC, UMC Groningen, AstraZeneca BV, Johnson & Johnson, Roche Nederland BV, Almende BV, Linksight BV, Pharmo Institute NV, SURF BV, CZ Zorgverzekeringen NV, Stichting Olijf en TNO.

Projectondersteuning door Topsector Life Sciences & Health.



Vragen? Neem contact op met Erik Cornelisse erik.cornelisse@tno.nl.