

aanpak begeleidingsethiek

Verslag workshop

Aanpak Begeleidingsethiek

Synthetische data in het kader van ELAN-VIPP

Leids Universitair Medisch Centrum

ECP Platform voor de InformatieSamenleving

LUMC Leids Universitair Medisch Centrum

Voorwoord

De toenemende digitalisering brengt ingrijpende veranderingen met zich mee, in allerlei sectoren. Nieuwe mogelijkheden voor communicatie, monitoring en analyse roepen vragen op. Wat is de positie van de mens in deze omgeving, hoe staat het met data, met privacy, wat doen algoritmes en wat willen gebruikers eigenlijk? Begeleidingsethiek is ethiek die zich specifiek met dit soort vragen bezighoudt. Hij is gebaseerd op techniekfilosofie, die uitgaat van eeuwenoude verwevenheid tussen mens en technologie.

De Aanpak Begeleidingsethiek is een concrete aanpak waarin betrokkenen met elkaar in dialoog gaan over de effecten van de nieuwe technologie én de waarden die daarbij in het geding komen. Dat zijn vaak waarden gelieerd aan autonomie van de gebruiker, efficiëntie van het proces, transparantie van het algoritme, privacy, et cetera.

Tijdens de sessie komen verschillende stakeholders als gebruikers, ontwikkelaars, beleidsvormers en beslissers met elkaar in gesprek. Na de dialoog hebben de deelnemers ethische handelingsopties gegenereerd, waarvan verschillende vaak direct opgepakt kunnen worden. Er is gezamenlijk gekeken welke waarden we in het digitale domein belangrijk vinden en hoe we die willen verankeren en borgen in digitale processen en handelingen. Dit alles om ook bij verdere digitalisering op het vertrouwen van de samenleving en de participanten kunnen blijven rekenen. Dit is niet in één stap te realiseren het is een continu proces waarin deze workshop een schakel is.

Workshop aanpak begeleidingsethiek

Initiatiefnemers: Marco Spruit (LUMC)
Moderatoren: André Krom (LUMC), Maartje Bouwens (ZWconnect) en Marleen Eijkholt (LUMC)

Op vrijdag 25 augustus 2023 organiseerde LUMC op initiatief van Marco Spruit (LUMC) een sessie begeleidingsethiek over de casus 'Synthetische data in het kader van ELAN-VIPP'. Aan de workshop namen 14 deelnemers (zie bijlage voor specificatie) deel onder leiding van André Krom (LUMC), Maartje Bouwens (ZWconnect) en Marleen Eijkholt (LUMC).

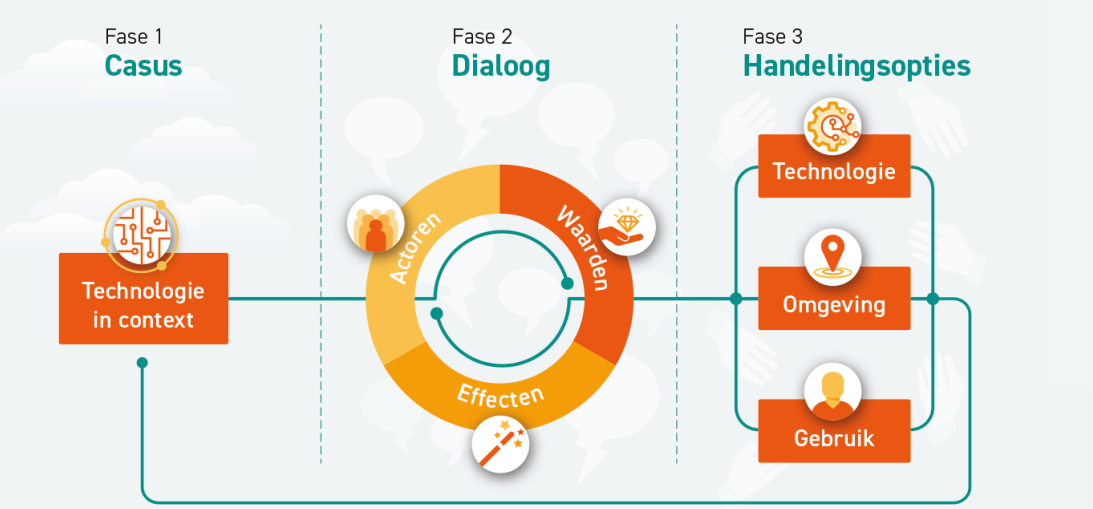
De doelstelling is tweeledig:

- Leren tijdens de workshop
  - Welke effecten, waarden en actoren worden genoemd
  - Welke handelingsopties zijn er om de toepassing te verbeteren
- Leren over de Aanpak Begeleidingsethiek
  - Hoe kan de organisatie deze methodiek gebruiken rondom ethische kwesties en technologie-implementaties

Het verslag bevat de volgende elementen:

- Toelichting op de Aanpak Begeleidingsethiek
- Weerslag van de workshop

Aanpak begeleidingsethiek



De aanpak bevat de volgende fasen:

Fase 0 Introductie

Introductie over de doelstelling en een toelichting op het model en het gedachtengoed

Fase 1 Toelichting

Hoe ziet de technologie eruit en in welke context wordt deze toegepast?

Fase 2 Dialoog

- Een korte ronde waarin de deelnemers aan de workshop de betrokken actoren benoemen
- Brainwrite waar deelnemers mogelijke effecten benoemen en bespreken
- Benoemen van waarden die een rol spelen bij die effecten

Fase 3 Handelingsopties

In subgroepjes gaan de deelnemers op zoek naar handelingsopties vanuit de technologie, de omgeving en het individu.

Fase 1

Technologie en context

ELAN is het domein-overstijgende en regionale dataplatform van het Leids Academisch Medisch Centrum (LUMC). ELAN staat voor Extramuraal LUMC Academisch Netwerk. ELAN ondersteunt gezondheidsbeleid en onderzoek. Daartoe verzamelt ELAN via ongeveer 120 huisartspraktijken in de regio Haaglanden gepseudonimiseerde (a) medische data, (b) GGZ-informatie, (c) WMO-informatie, en (d) gegevens over de publieke gezondheid van honderdduizenden door hun huisarts geïnformeerde inwoners. Met behulp van deze data worden in ELAN zogeheten “voorspelmodellen” ontwikkeld, voor bijvoorbeeld het inschatten van iemands risico op een beroerte, mede op basis van diens leeftijd en geslacht. (e.g. Van Os, 2023)

ELAN is in Nederland een unieke bron voor domein-overstijgend zorgonderzoek en beleidsvorming. Een strategisch doel is om voorspelmodellen die in ELAN worden ontwikkeld terug “naar de bron” te brengen. Op die manier zouden partners van ELAN die de data aanleveren daar ook in hun dagelijkse praktijk profijt van kunnen hebben. Dit zijn o.a. het LUMC, de Universiteit Leiden, de gemeente Den Haag, het Haaglanden Medisch Centrum, het HagaZiekenhuis, het Reinier de Graaf Gasthuis, de GGD-Haaglanden, de Parnassia Groep en HADOKS (Haagse huisartsen). Daarnaast biedt ELAN grote potentie voor gebruik in onderwijsprogramma’s zoals Population Health Management, doordat studenten op realistische wijze de zorgcomplexiteit binnen hun eigen vertrouwde leefomgeving/populatie zouden kunnen exploreren.

Om deze omvangrijke databron met gevoelige gegevens veilig te kunnen gebruiken is vereist dat de toegang grondig gereguleerd is. Het uitgangspunt is om privacy te borgen en om het risico dat gevoelige informatie over personen of bedrijven wordt onthuld, zo klein mogelijk te maken. ELAN werkt daartoe via de zogeheten toegang-op-afstand (Remote Access, RA) omgeving van het Centraal Bureau voor de Statistiek (CBS). In de RA-omgeving kunnen Nederlandse universiteiten, wetenschappelijke organisaties, planbureaus en statistische instanties in een EU-land onder strikte voorwaarden zelf statistisch onderzoek doen.

In de praktijk vraagt het een aanzienlijke hoeveelheid tijd, geld en moeite om een ELAN data-analyse uit te kunnen voeren binnen de ELAN/CBS-RA omgeving. Bovendien mag de CBS-RA omgeving niet voor onderwijsdoeleinden worden gebruikt. Ook is “de weg terug” zeer ingewikkeld. Het CBS vereist dat eventuele “export” van data handmatig wordt gecontroleerd. Dat stelt bijvoorbeeld beperkingen aan het terug naar de bron brengen van voorspelmodellen die in

ELAN worden ontwikkeld naar partners die data aanleveren voor ELAN.

Deze uitdagingen staan centraal in het VIPP-ELAN project. VIPP-ELAN staat voor een Virtuele Patiënten en Populatie dataset uit ELAN. De probleemstelling van het project is hoe zorgonderzoek kan worden versneld en hoe zorgonderwijs kan worden geïnnoveerd m.b.v. “synthetische data” en zonder onthullingsrisico. Synthetische data simuleren kenmerken van relaties tussen personen en objecten (bijvoorbeeld een school of een wijk), waardoor de realiteit kan worden nagebootst zonder dat de persoon of het object geïdentificeerd wordt (bron: CBS).

Sinds einde 2021 wordt onder leiding van de afdeling Public Health en Eerstelijnsgeneeskunde (PHEG) van het LUMC reeds aan dit VIPP-ELAN project gewerkt met de partners CBS en Syntho. Syntho is een Nederlandse startup die software ontwikkelt en exploiteert voor het maken van synthetische data. In het ELAN-VIPP project worden verschillende technische manieren om synthetische data te maken met elkaar vergeleken. Ook wordt onderzocht hoe een synthetische dataset kan worden ontwikkeld die tenminste even uitlegbaar en betrouwbaar is als de oorspronkelijke ELAN dataset zonder dat het risico op onthulling groter wordt.

Fase 2

Dialoog

In deze tweede fase gaan de deelnemers in gesprek over wie er betrokken zijn bij de inzet en het gebruik van synthetische data in het kader van VIPP-ELAN. Ook buigen zij zich over de positieve en negatieve effecten van deze inzet, en benoemen zij belangrijke waarden waar rekening mee gehouden moet worden bij de inzet van een dergelijke toepassing.

Actoren:

Bij de actoren is de vraag wie er betrokken is of geraakt wordt door de case. De deelnemers aan tafel vertegenwoordigen al een deel van die actoren. Ze noemen de volgende betrokkenen (alfabetisch):

- Ambulance medewerkers
- Andere zorginstellingen
- Banken (in relatie tot verzekeraars)
- Beleidsmakers (o.a. ministerie van VWS, Europese Commissie)
- Burgers
- Centraal Bureau voor de Statistiek (CBS)
- Data administrators / managers
- Data bezitters (bijv. patiënten)
- Deelnemers vanuit ELAN (data leveranciers)
- Gemeenten
- Gemeentelijke Gezondheidsdiensten (GGD)
- Samenleving/ iedereen
- Leverancier
- Media
- Onderzoekers (/ hypothese makers)
- Ontwikkelaars
- Patiënten
- Privacy officer
- Professionele zorgfederaties
- Reviewers
- Rijksinstituut voor Volksgezondheid en Milieu (RIVM)
- Scholen
- Secretaresses
- Security
- Software ontwikkelaars
- Studenten
- Subsidieverstrekkers (bijv. ZonMw)
- Tijdschriften
- Toezichthouder

- Triagisten
- Universiteit Leiden
- Verzekeraars
- Wetgever
- Zorginstituut Nederland (ZiN)
- Zorgverleners

De vraag aan de deelnemers was om in het vervolg van de sessie ook te proberen vanuit deze perspectieven hun inbreng te geven.

Effecten:

De vraag is hier welke positieve en negatieve effecten de invoering en het gebruik van synthetische data hebben. Er is bij enkele effecten specifiek aangegeven in hoeverre dit het onderwijs (O), onderzoek (OZ) of kwaliteit van zorg (K) raakt.

Positieve effecten

- Nieuwe innovatieve toepassingen bedacht (K)
- Versnelde innovatie van AI (OZ)
- Diepgaander onderwijsaanbod
- Betere professionals
- Snellere voortgang van studenten (O)
- Betere kwaliteit van zorg (K)
- Onderwijsinnovatie (O)
- Regionale samenwerking
- Data compleet maken (O/OZ/K)
- Gevaar tot data lek vermindert
- Domein overstijgend werken
- Aanvulling voor evidence based medicine
- Dynamisch, mooi om mee te spelen, flexibel
- Onderzoek doen zonder beperkingen van privacy
- Inzicht
- Studenten kunnen met ‘echte’ data leren werken
- Minder beperkingen in het beheren van de data (gebruiksgemak/ toegankelijkheid)
- Betere zorguitkomsten
- Betere gepersonaliseerde zorg

Negatieve effecten

- Toegenomen angst voor onthullingsrisico (K)
- Gevaar tot datalek (risico)
- Regionale samenwerking
- Kan leiden tot bias (onwaarheden in dataset)
- Je kan niet alle relaties afbakenen dus ook niet alle relaties onderzoeken (OZ)
- Persoonlijke zorg verminderd (luisteren)
- Mensenrechten/zorgvuldigheid
- Risico van blackboxing
- Ontstaan van verkeerde hypotheses
- Klinische data kloppen niet met praktijk
- Kwaliteit van onderzoek o.b.v. synthetische data kan omlaag gaan
- Hoe doe je onderzoek o.b.v. synthetische data --> vertrouw je conclusies uit synthetische data
- Hogere drempel om met synthetische data te werken (begrijpen van de data zelf)
- Schaalbaarheid (beperkingen)

- Minder beperkingen in het beheren van de data (gebruiksgemak / toegankelijkheid)
- Snellere doorlooptijd (OZ)
- Betere modellen
- Betere zorguitkomsten (K)
- Toegankelijkheid van data voor beleid
- Raakvlakken met praktijk vanuit onderwijs --> mogelijk invloed
- Als katalysator voor projectinnovatie (OZ)
- Onderzoeksprojecten sneller, meer en beter
- Hypotheses vormen a.d.h.v. synthetische data (ontdekken van nieuwe vragen)
- Synthetische data gebruiken om een bias te verminderen (vb. ongebalanceerde data)
- Synthetische data kan de algemene discussie van datagebruik in de maatschappij stimuleren
- Open beschikbaar maken zodat partijen die niet bij de 'echte' data kunnen er wel mee aan de slag kunnen
- Betere, gepersonaliseerde zorg (K)
- Geen toestemming meer nodig van patiënten (OZ)
- Aggregatie van verschillende datasets --> identificeerbaar worden van data
- Onbegrip over gebruik kan vertrouwen schaden (wantrouwen) (privacy, gebruiksdoelen)
- Risico: eventueel onduidelijk of een dataset uit hypothetische data bestaat of niet
- Kan je onderscheid maken tussen echte en neppe data?
- Teveel vertrouwen in de kwaliteit en bruikbaarheid van synthetische data
- Gaat vaak om individuen en daar is synthetische data minder geschikt voor (K)
- Meer *technology push* en niet meer beredeneerd vanuit behoefte (OZ)

Waarden:
Na het inventariseren van mogelijke effecten is de groep deelnemers opgedeeld in 3 subgroepen. Het totaal aantal effecten is gedeeld door 3 en verdeeld over de subgroepen. Elke subgroep heeft eerst een top-3 uit hun lijst van effecten geselecteerd en daar vervolgens waarden bij geïdentificeerd.

Effect	Waarden
Toegenomen angst voor onthullingsrisico (K) (-)	Vertrouwen Betrouwbaarheid Uitlegbaarheid Transparantie Rechtvaardigheid (*justice*) Inclusiviteit Relaties
Gevaar tot data lek (risico) (-)	Vertrouwen Betrouwbaarheid Veiligheid Privacy Relaties
Gevaar tot data lek vermindert (+)	Vertrouwelijkheid (*confidentiality*) Privacy Weerlegbaarheid (*plausible deniability*) Institutioneel vertrouwen Reputatie Zelfbeschikking
Data compleet maken (O/OZ/K) (+)	Kwaliteit Betrouwbaarheid (van bijv. onderzoekers) Volledigheid Overeenstemming (over bepaalde standaarden) Democratie Inclusiviteit Meeting the needs of others
Hoe doe je onderzoek o.b.v. synthetische data → vertrouw je conclusies uit synthetische data? (-)	Betrouwbaarheid (van uitkomsten) Validiteit (van onderzoeksuitkomsten) Vertrouwen (in het onderzoeksproces)
Minder beperkingen in het beheren van de data (+)	Gebruiksgemak Toegankelijkheid Innovatie Doelmatigheid (kosteneffectiviteit)

	Openheid (Open Science/ FAIR)[1] Transparantie
Onbegrip over gebruik kan vertrouwen schaden (wantrouwen) (privacy gebruiksdoelen) (-)	Integriteit Acceptatie Vertrouwen Privacy Reputatie Gezondheid
Betere zorguitkomsten (K) (+)	Gezondheid Betaalbaarheid Werkplezier Kwaliteit van leven
Betere, gepersonaliseerde zorg (+)	Vitaliteit Efficiëntie Vertrouwen

[1] FAIR staat voor Findability, Accessibility, Interoperability, and Reusability, https://www.go-fair.org/fair-principles/

Fase 3

Handelingsopties

In het proces van het opstellen van de handelingsopties zijn de waarden uit de vorige fase meegenomen. De deelnemers kregen een uitleg over de verschillende categorieën binnen de handelingsopties. Het kan bijvoorbeeld gaan over het ontwerp van de technologie, het aanpassen van de omgeving waarbinnen de toepassing functioneert en het gedrag van mensen. De deelnemers worden in drie werkgroepen verdeeld en gaan ieder aan de slag met een van deze drie gespreksonderwerpen. Dat leidt tot handelingsopties per categorie.

Technologie:

Effect	Handelingsopties
Toegenomen angst voor onthullingsrisico (K)	• Fysieke presentatie van data (graphics) • De stappen kunnen laten zien/ documenteren • Gebruik een methode die geen echte data gebruikt (rule-based model) • Vereenvoudigde kopie van het proces van de technologie, inclusief media
Gevaar van datalek (risico)	• Onderzoeken of er extra beveiliging kan zijn (extra encriptie)
Data compleet maken (O/OZ/K)	• Grote rekenkracht • Grote(re) hoeveelheid data • Verschillende methoden gebruiken en vergelijken
Hoe doe je onderzoek o.b.v. synthetische data --> vertrouw je conclusies uit synthetische data	• Onderzoek doen naar evaluatie van metrieken om dit te kwantificeren / karakteriseren
Gevaar tot data lek vermindert	• Tijdens ontwikkelproces reeds juridische borging (AVG-compliant) en informatiebeveiliging (technisch)
Onbegrip over gebruik kan vertrouwen schaden (wantrouwen) (privacy, gebruiksdoelen)	• Vast watermerk bij synthetische data

Omgeving:

Effect	Handelingsopties
Toegenomen angst voor onthullingsrisico (K)	• Veranderingen documenteren, documenteren hoe het model werkt en hierover afspraken maken • Proces documenteren → wat gebeurt er wanneer? • Protocollen opstellen die beschrijven hoe
Gevaar van datalek (risico)	• Protocollen opstellen: wat is er nodig om in een veilige omgeving te werken
Data compleet maken (O/OZ/K)	• Partners nodig die data leveren • Opt-in/ opt-out schema's voor gebruik van data
Hoe doe je onderzoek o.b.v. synthetische data --> vertrouw je conclusies uit synthetische data	• Integreren in opleidingen t.b.v. toekomstige professionals
Gevaar tot data lek vermindert	• Ontwikkelen van gedragscodes, door het veld gedragen (en toetsen)
Minder beperkingen in het beheren van de data (gebruiksgemak / toegankelijkheid)	• Openbaar maken van data en algoritmes (Open Science) • Bijsluiter met meta data en toelichting (o.a. koppelbaarheid)
Onbegrip over gebruik kan vertrouwen schaden (wantrouwen) (privacy, gebruiksdoelen)	• Certificering van data • Software certificering • Ontwikkelen van standaarden, bijv. ISO standaard • Afspraken maken over het genereren en gebruiken van data • Privacy criteria en evaluatie in proces opnemen
Betere zorguitkomsten (K)	• Opleiding/ cursus opzetten voor zorgverleners: wat is 'echt' en wat is 'nep' (qua data)

Mens:

Effect	Handelingsopties
Toegenomen angst voor onthullingsrisico (K)	• Training • Workshop over veiligheid • Bewustzijn creëren (*leniency*) • “Hoe te publiceren” • Wetenschapsjournalisten die de technologie kunnen vertalen naar de gemiddelde burger
Gevaar van datalek (risico)	• Werken in een veilige omgeving
Data compleet maken (O/OZ/K)	• Overhalen van partijen om data te delen
Hoe doe je onderzoek o.b.v. synthetische data --> vertrouw je conclusies uit synthetische data	• Communiceren nodig om bewustzijn/ competentie van publiek t.a.v. de technologie te stimuleren • Wetenschapscommunicatie
Gevaar tot data lek vermindert	• Responsible disclosure • Responsible stewardship • Aanspreken op en ‘opvoeden’ t.b.v. van morele verantwoordelijkheid
Onbegrip over gebruik kan vertrouwen schaden (wantrouwen) (privacy, gebruiksdoelen)	• Trainingen over werken met synthetische data • Certificaat kunnen aantonen • Data steward c.q. key user aanstellen als vraagbaak
Betere zorguitkomsten (K)	• Begrijpelijke termen gebruiken in gesprek met patiënten • Onderzoeker goed bewust maken van de (on)mogelijkheden van synthetische data

Terugblik en afronding

Aan het einde van de workshop wordt teruggekeken en besproken wat de Aanpak Begeleidingsethiek concreet heeft opgeleverd. De deelnemers geven aan dat het veel nieuwe inzichten oplevert en dat er in korte tijd veel wordt gerealiseerd. Daarnaast wordt de brede groep van deelnemers als meerwaarde ervaren. De combinatie van een duidelijke casus, enthousiaste deelnemers, en een aanpak die zorgt voor duidelijke uitkomsten in een korte tijd was dus een succes.

De deelnemers

1. **Beek, Sifra van de; Promovendus, sectie Ethiek en Recht van de gezondheidszorg, LUMC**
2. **Bloem, Maarten; Relatiebeheerder microdataservices; Onderzoeker, Centraal Bureau voor de Statistiek (CBS)**
3. **Bouwens, Maartje; Programmaleider innovatie, werkgeversvereniging ZWconnect**
4. **Braaksma, Barteld; Innovation Manager, Centraal Bureau voor de Statistiek (CBS)**
5. **Dijk, Friso van; Privacy-Officer, O & P Rijk, duaal promovendus**
6. **Dortmont, Koert van, data scientist; Accountmanager, CTcue, HagaZiekenhuis**
7. **Eijkholt, Marleen; Senior docent/ onderzoeker, Sectie Ethiek en Recht van de Gezondheidszorg, LUMC**
8. **Faiq, Mohammed Ammar; Data Science Researcher; Promovendus bij de LUMC-Campus Den Haag**
9. **Krom, André; Senior docent/ onderzoeker, Sectie Ethiek en Recht van de Gezondheidszorg, LUMC; duo coördinator ELSI LUMC voor Health-RI**
10. **Moester, Martiene; Senior Data Steward, Afd. Advanced Data Management; duo-coördinator ELSI LUMC voor Health-RI**
11. **Mook-Kanamori, Dennis; Huisarts, Klinisch epidemioloog bij het LUMC en de Health Campus Den Haag, hoofd van het kernteam van ELAN (Extramuraal LUMC Academisch Netwerk)**
12. **Spruit, Marco; Hoogleraar Translationele Datawetenschap in Populatiegerichte Zorg bij zowel het Dep. Publieke Zorg & Eerstelijnsgeneeskunde (PHEG, LUMC) en het Leiden Instituut voor Informatica (LIACS)**
13. **Vajda, Ildikó; Senior Adviseur Patiëntenbelang, Nederlandse Patiëntenfederatie**
14. **Vonk, Marijn; medeoprichter en Chief Product Officer, Syntho (start-up synthetische datasoftware)**

Moderatoren:
André Krom (moderator) (LUMC), Marleen Eijkholt (co-moderator) (LUMC), Maartje Bouwens (co-moderator) (ZWconnect)

aanpak begeleidings ethiek

ECP Platform voor de InformatieSamenleving

Verslag workshop
Aanpak Begeleidingsethiek
Synthetische data in het kader van ELAN-VIPP

Meer informatie over de Aanpak Begeleidingsethiek, waaronder dit verslag, vindt u op **www.begeleidingsethiek.nl**

Foto omslag: Pixabay