

De balans tussen openheid en privacy voor gezondheidsdata die verzameld worden via citizen science: verschillende perspectieven



Ria Wolkorte



Lieke Heesink



Michelle Kip

Bij het Citizenlab willen wij graag onderzoek doen waarin onderzoekers en burgers samenwerken, en waarin iedereen zijn eigen expertise in kan brengen. Dit wordt ook wel citizen science (burgerwetenschap) genoemd. Dit onderzoek kan op verschillende manieren worden gedaan, zoals bijvoorbeeld met:

- kwalitatieve methodieken (zoals interviews, groepsgesprekken of co-creatiesessies);
- kwantitatieve methodieken (gegevensverzameling op een website, dragen van activity trackers);
- een combinatie van bovenstaande (zoals vragenlijsten met open en gesloten vragen).

De burgers in het project hebben hierbij regelmatig een dubbele rol; ze denken mee met het onderzoek (in de rol van co-onderzoeker) maar leveren of verzamelen ook gegevens over zichzelf of over hun eigen gezondheid of welzijn (in de rol van onderzoeksdeelnemer).

Data

Het is belangrijk dat de **veiligheid en privacy van data** is gewaarborgd, helemaal omdat in onze onderzoeken gegevens over de gezondheid van mensen wordt verzameld. Ook vinden wij het heel belangrijk dat de deelnemers aan het onderzoek altijd **weten waar hun data voor gebruikt wordt of kan worden**, zodat deelnemers goed kunnen besluiten om daar **wel of geen toestemming** voor te geven en daarmee dus wel of niet aan het onderzoek mee te werken.

FAIR: wat is FAIR data?

Binnen citizen science wordt het heel belangrijk gevonden om **open en transparant te werken**, wat ook betekent dat je zo open mogelijk omgaat met de data die je verzamelt tijdens een onderzoek. Om dit te bereiken proberen we in citizen science projecten te werken volgens de principes van **FAIR data**. FAIR omvat richtlijnen voor het beschrijven, opslaan en publiceren van wetenschappelijke data. Een belangrijk doel van FAIR data is **om wetenschappelijke data geschikt te maken voor hergebruik**, bijvoorbeeld zodat andere onderzoekers deze gegevens ook in een ander onderzoek kunnen gebruiken. Een ander voordeel van FAIR data is dat gecontroleerd kan worden of het onderzoek op de juiste manier is uitgevoerd. Daarbij wordt de regel gehanteerd: **zo open mogelijk, maar gesloten waar nodig**. De overwegingen hierbij zijn afhankelijk van het soort data. Daardoor kunnen er andere keuzes worden gemaakt rondom het openbaar maken van data over luchtkwaliteit vergeleken met bijvoorbeeld data over de gezondheid van een persoon.

Metadata

Binnen de principes van FAIR wordt niet alleen gesproken over data, maar ook over metadata. **Metadata is data over data**. Een lijst met daarin telefoonnummers en namen is de data zelf, de omschrijving van deze lijst ("*in deze lijst staan 4 telefoonnummers en de bijbehorende namen*")

noem je metadata. Maar ook bijvoorbeeld de naam van de auteur van de data, of het aantal pagina's waaruit de data bestaat is metadata (zie figuur 1).

Je kunt er binnen de principes van FAIR bijvoorbeeld voor kiezen om niet de data van het onderzoek zelf openbaar te maken, maar alleen de metadata; dan weten anderen wel dat de data bestaat en welke informatie er in staat, maar kunnen zij deze niet zien.

Dataset			Metadata	
Naam	Telefoonnummer	Emailadres	Metadata	
Linda	06-45685275	linda@gmail.com	Aantal respondenten	4
Ron	06-25632563	ron@gmail.com	Inhoud	Namen
Jamila	06-85274196	jamila@gmail.com		Telefoonnummers
Rene	06-85479328	rene@gmail.com		Emailadressen

Figuur 1. Voorbeeld van een dataset en een deel van de bijbehorende metadata

Open data en opslag van data

Data van wetenschappelijk onderzoek wordt in toenemende mate opgeslagen in een online **repository** (opslagplaats). Dit is een centrale plaats, op een website, waar data opgeslagen wordt. Hier kunnen onderzoekers data vinden die door anderen zijn verzameld, maar die zij misschien ook voor hun eigen onderzoek kunnen gebruiken. Zo hoeven ze deze data niet zelf opnieuw te verzamelen. **Iedereen kan zonder te betalen en zonder in te loggen in een repository zoeken naar datasets.**

In zo'n repository kun je data van een onderzoek met verschillende niveaus van openheid plaatsen:

- **'open data'**. Open data verwijst naar datasets die vrij toegankelijk zijn. Dat betekent dat iedereen toegang heeft tot deze data. Iedereen kan deze datasets downloaden zonder reden of betaling.
- **'toegang alleen op aanvraag'** betekent 'opvraagbaar met een goede reden'. Hierbij wordt in de repository alleen de metadata van een onderzoek vermeld, waarbij wordt aangegeven dat onderzoekers die geïnteresseerd zijn in de dataset contact op kunnen nemen met de originele onderzoekers. Dan kan worden besproken of de aanvragers toegang mogen krijgen tot de (geanonimiseerde) data en onder welke voorwaarden.
- **'alleen metadata'**. Een andere optie is om in deze repository alleen de metadata van een onderzoek te vermelden. Hierbij hebben anderen nooit de mogelijkheid om de data in te zien of er mee te werken. Een belangrijke reden om te kiezen voor 'alleen metadata' is omdat de dataset zelf persoonlijke gegevens bevat, waarbij de dataset zijn waarde verliest wanneer deze persoonlijke gegevens worden verwijderd. Daarnaast kan hiervoor gekozen worden als deelnemers geen toestemming hebben gegeven voor het (openbaar) delen van de data.

Anonimisatie van gegevens

Als gegevens van mensen worden verwerkt in een wetenschappelijk artikel of opgeslagen worden in een repository, is het van belang dat deze gegevens anoniem zijn. Dat houdt in dat deze gegevens **niet meer terug te leiden zijn** naar de mensen die hebben meegedaan aan het onderzoek. Om data te anonimiseren zijn er verschillende mogelijkheden. Je kunt bijvoorbeeld een

geboortedatum veranderen in een categorie (bijvoorbeeld de categorie 30-40 jaar, in plaats van 21 juli 1983). Of je kunt gevoelige informatie (zoals namen of medische geschiedenis) weglaten uit de dataset.

Resultaten van 2 groepsgesprekken

Om te besluiten hoe we binnen het Citizenlab in de toekomst omgaan met data, zijn we het gesprek aangegaan. Aan het eerste gesprek hebben 4 mensen met reuma deelgenomen, die allen al eerder hadden meegedaan aan onderzoek van het Citizenlab. Aan het tweede gesprek hebben 2 **mensen met reuma deelgenomen, een citizen science onderzoeker, een ethicus, en een data steward** (een data steward houdt zich bezig met de juiste verzameling, opslag en verwerking van data). Over het algemeen lagen de meningen van de deelnemers erg dicht bij elkaar, hoewel er soms kleine individuele verschillen werden benoemd.

Belang van het delen van data

De deelnemers benoemen dat het **delen van onderzoeksdata belangrijk is**. Daardoor kan onderzoek efficiënter worden; bepaalde dataverzameling hoeft hierdoor niet opnieuw plaats te vinden. Dit verlaagt de belasting voor deelnemers en mogelijk ook kosten voor onderzoek. Ook kan de bijdrage van een deelnemer hierdoor grotere impact hebben omdat het voor verschillende onderzoeken wordt ingezet.

Belang van privacy, toestemming en anonimisatie

Volgens de deelnemers **staat privacy voorop bij deelname aan onderzoek**. Data mag alleen buiten het betreffende onderzoek worden gedeeld en/of hergebruikt wanneer deelnemers hier expliciet over zijn geïnformeerd en zij hiervoor toestemming hebben gegeven. Het is daarbij erg belangrijk dat de data geanonimiseerd wordt. Dit betekent dat persoonsgegevens, zoals namen, emailadressen, en andere herkenbare gegevens worden verwijderd of vervangen door bijvoorbeeld een nummer of een algemene beschrijving. Dit houdt dus in dat persoonsgegevens en andere herleidbare gegevens nooit worden gedeeld buiten het onderzoek waarvoor de data in eerste instantie is verzameld.


Data binnen het Citizenlab

Er bestaan **verschillende soorten data**, variërend van bijvoorbeeld interviews waarin veel (persoonlijke, mogelijk gevoelige) informatie wordt gedeeld, naar continue registratie van een hartslag (op basis van de data niet herleidbaar naar een persoon). De deelnemers geven aan dat er bij kwalitatieve data veel aandacht moet zijn voor goede anonimisatie. Wanneer de data goed geanonimiseerd is, zou daarna met alle soorten data op eenzelfde manier omgegaan mogen worden.

De deelnemers gaven aan dat de voorkeur uitgaat naar

- het **publiceren van metadata in een repository**.
- **toegang tot de geanonimiseerde database kan worden verkregen door een wetenschapper met een legitieme onderzoeksvraag**; hierbij is het aan de onderzoekers om te bepalen wat een legitieme vraag is. Daarvoor zouden eventueel criteria opgesteld kunnen worden.

Vooraf informeren

Het is volgens alle deelnemers erg belangrijk dat **voorafgaand aan deelname** duidelijk wordt uitgelegd wat er met de data gebeurt. Op deze manier kunnen mensen een **geïnformeerde keuze** maken over eventuele deelname aan een onderzoek. Daarbij moet altijd de afweging worden gemaakt tussen **volledige duidelijkheid en de leesbaarheid en lengte** van de uitleg. Er zijn verschillende suggesties gedaan om dit te doen. Zo zou de uitleg kort en beknopt kunnen, waarbij bijvoorbeeld een knop “  ” kan worden gebruikt om deelnemers de mogelijkheid te geven om uitgebreidere informatie te lezen. Ook kan worden gedacht aan een kort filmpje met uitleg, en daarnaast kan de mogelijkheid geboden worden om meer informatie te ontvangen van de onderzoekers.

Conclusie

Bij het opslaan, verwerken en eventueel delen van de data is het vooral heel belangrijk om open en transparant te zijn over het onderzoek. Zo kunnen eventuele deelnemers een goed geïnformeerde afweging maken over eventuele deelname. Daarbij wordt het belang van het delen van onderzoeksdata onderkend, maar zal privacy altijd gewaarborgd moeten zijn.

Op basis van alle informatie uit deze gesprekken, en aanvullende gesprekken met data stewards, privacy specialisten en specialisten mensgebonden onderzoek zal door ons een handleiding worden geschreven over het bewust omgaan met data die wordt verkregen in onderzoek. Hierin zal onder andere worden beschreven:

- welke informatie deelnemers vooraf moeten ontvangen
- dat data moet worden geanonimiseerd en waar onderzoekers daarbij op moet letten
- de data kan worden vertaald naar het Engels, om door een grotere groep onderzoekers hergebruikt/bekeken te kunnen worden
- welke opties er zijn wat betreft openheid wanneer data wordt opgeslagen in een repository en wat deze opties inhouden
- hoe je goede metadata van de dataset kunt beschrijven, zodat anderen weten wat voor data er verzameld is
- hoe je data veilig kan delen met andere onderzoekers

Deze handleiding kan altijd worden aangepast op basis van nieuwe inzichten, en uiteraard kan er – waar wenselijk en onderbouwd – van worden afgeweken binnen projecten.

Voor meer informatie of deelname aan vervolgonderzoeken, kunt u contact met ons opnemen via r.wolkorte@utwente.nl.