

1.1 Aanleiding

Sinds maart 2020 wordt door de Corona Gedragsunit periodiek een vragenlijstonderzoek onder 50.000 respondenten uitgevoerd (inmiddels 11 rondes uitgevoerd), om zicht te krijgen op gedrag, draagvlak voor maatregelen en welbevinden gedurende de corona pandemie. Daarnaast worden interviews gehouden met een selectie van de respondenten. Tenslotte wordt sinds oktober elke 3 weken een verkorte versie van de vragenlijst met representativiteit op veiligheidsregioniveau uitgevoerd (totaal ca. 5000 respondenten), ten behoeve van publicatie op het Corona Dashboard van de rijksoverheid. Aangezien de inzet in deze pandemie is om data rondom corona met spoed openbaar te maken, en RIVM als beleid heeft "open, tenzij", is ingezet op een open data/science traject voor de Corona Gedragsunit.

1.2 Projectdoelstellingen

Op basis van heldere uitgangspunten, werkwijze en organisatie zorgdragen voor het beschikbaar maken van data en analyses voor derde partijen, met name wetenschappelijke onderzoekers in binnen- en buitenland.

Voor de lange termijn is de generieke doelstelling geformuleerd: op transparante wijze data van het onderzoek in de Corona Gedragsunit beschikbaar stellen ten behoeve van (wetenschappelijke) kennisverspreiding en -ontwikkeling. De doelgroep voor de data is breed: wetenschappers, burgers, media, beleid. Om e.e.a. behapbaar te houden is de korte termijn doelstelling: de data die gegenereerd worden in het kwantitatieve vragenlijstonderzoek, beschikbaar stellen ten behoeve van wetenschappelijk onderzoek. De doelgroep is daarmee ook kleiner.

1.3 Scope van het project

De focus van het project ligt op de data die is verzameld in het vragenlijstonderzoek. Echter, de werkwijze die ontwikkeld wordt, moet ook toepasbaar zijn voor interviewdata en data verzameld voor het Corona dashboard. Overigens zijn de dashboard data al op geaggregeerd niveau openbaar beschikbaar.

1.4 Projectresultaten

De volgende resultaten zijn onderdeel van wat we verder "de Open Science werkwijze" zullen noemen:

- Een heldere set uitgangspunten voor het beschikbaar stellen van AVG-proof data op recordniveau (antwoorden van 1 respondent staan op 1 regel per ronde), geaggregeerde data, of analyses;
- Hierbij wordt onderscheid gemaakt in het online beschikbaar maken van de data, of een specifieke uitsnede van de data verstrekken aan een derde partij naar aanleiding van een verzoek, of naar aanleiding van een vraag die we zelf gesteld hebben;
- Geaggregeerde data staan per 1 mei deels online op rivm.nl en worden in tranches daarna aangevuld;
- Overzicht van de data (per ronde of anderszins) die we beschikbaar stellen en waar we deze data opslaan;
- Repository van de meta-data;
- Publicatiebeleid (wat doen we zelf met de data, hoe en wanneer willen we welke data of analyses aan derden beschikbaar stellen, welke vragen willen we pro-

actief uitzetten, wat is de positie van WAR-leden en GGD'en, welke uitgangspunten rondom co-auteurschap hanteren we hierbij);

- Welke dataset gaan we beschikbaar stellen, met of zonder missings. Of we nu met of zonder missings data beschikbaar maken, in beide gevallen moet goed uitgelegd worden wat er wel en niet in de data zit en hoe het verzameld is;
- Een ingerichte organisatie voor vraagontvangst en -afhandeling van data verzoeken (loket) en een toetsingscommissie voor de verzoeken.
- Een compact handboek Open science, waarin wordt beschreven:
 - Vraagontvangst, -beoordeling en afhandeling; hierin zitten o.a. formats voor indienen van een verzoek, beoordelingscriteria en -toetsingsgroep, prioritering/planning van verzoeken intern RIVM, VWS en andere partijen, manier waarop data/analyse beschikbaar wordt gesteld, doorgeleiden vraag naar interne leverancier, terugleiding naar aanvrager;
 - Communicatie over onze data, via RIVMdata en andere mogelijke kanalen, zowel Nederlands als Engelstalig. Gaan we actief aanbieden, reactief afwachten, of gericht onderzoeksvragen uitzetten waar onderzoekers op kunnen 'intekenen'?
 - Bekostiging van de uitgifte van data;
 - Omgaan met AVG (inclusief het opstellen van een data sharing overeenkomst) en persoonlijke herleidbaarheid;
 - Overzicht houden over wat met de data gedaan wordt;
 - Risico inventarisatie publicatie van data.

1.5 Uitgangspunten

- RIVM en GGD zijn **verwerkingsverantwoordelijke** van de data en worden gezien als 'intern'. Anderen worden gezien als extern. De WAR wordt in deze gezien als partner en VWS/NCTV als opdrachtgever in dit verband, niet als derde partij die een dataverzoek indient. Besproken moet worden welke eventuele uitzonderingen we voor deze partijen maken in geval van verzoeken.
- De GGD'en die partner zijn in het vragenlijstonderzoek dienen ook een rol te hebben in het inrichten en uitvoeren van de open data werkwijze. Hierbij moet ook helder zijn welke (gedeelde)verantwoordelijkheid zij dragen bij het delen van de data.
- We voldoen aan het doel van ZonMw voor het zoveel als mogelijk openbaren van de data. ZonMw vraagt dit voor door haar gesubsidieerde onderzoeken (de gedragsunit heeft de eerste 4 maanden het vragenlijstonderzoek en interviews uitgevoerd in een ZonMw opdracht);
- We voldoen aan de AVG, voor wat betreft herleidbaarheid van gegevens ('statistical disclosure control'). De risico-acceptatie van eventuele disclosure is een lijnverantwoordelijkheid. N.B. De respondenten aan het vragenlijstonderzoek hebben geen toestemming gegeven voor het gebruik maken van persoonlijke data door derde partijen!;
- We maken gebruik van al bestaande werkwijzen binnen het RIVM rondom open data;
- We bieden Engelstalige informatie en werkwijzen, zodat ook internationale onderzoekers onze data weten te vinden;
- De organisatie rondom Open Science van de Corona Gedragsunit moet uiterlijk zomer 2021 operationeel zijn en geborgd worden in de opvolger van de gedragsunit.

1.6 Besluitvorming

- Data beschikbaar stellen:

1. Geaggregeerde data worden online gezet. Gebruik van deze data is niet gebonden aan een overeenkomst, maar de data wordt wel met een CC-BY licentie gepubliceerd, waarbij bij gebruik van deze data het RIVM als bron genoemd dient te worden.
De data wordt tranchegewijs beschikbaar gesteld, waarbij bijvoorbeeld eerst vergelijkbare data als de WP11 data worden geopenbaard, of de vragen die structureel in alle rondes zijn uitgevraagd.
2. De dataverzoeken worden in behandeling genomen, indien het aanvraagformulier volledig is ingevuld. Er worden dus geen verzoeken op voorhand uitgesloten.
3. AVG-proof ruwe data wordt op verzoek verstrekt, geframed voor een specifieke vraagstelling. Door middel van statistical disclosure control zal ervoor gezorgd worden dat herleidbaarheid van de deelnemer verwaarloosbaar klein is. Hierbij kan een samenwerkingscontract afgesloten worden, om te waarborgen dat zorgvuldig met de data omgegaan wordt en voorafgaand aan publicatie instemming op de inhoudelijke analyse wordt verkregen van RIVM/GGD.
4. Ruwe data met daarin potentiële herleidbaarheid tot een deelnemer wordt op verzoek intern verstrekt, geframed voor een specifieke vraagstelling.
5. Analyseresultaten worden op verzoek verstrekt, onder voorwaarde dat verwerking beperkt blijft tot gebruik van de resultaten zoals aangereikt en onder de voorwaarde dat we geïnformeerd worden als op basis van de data gepubliceerd gaat worden.

6.

7. Volgorde van verzoeken:

1. Interne en VWS/NCTV verzoeken voor data gaan voor op externe verzoeken.
2. Externe verzoeken om data worden als volgt geprioriteerd: verzoek om data om een door ons geïnitieerde vraag te beantwoorden, verzoek om een specifieke dataset, verzoek om een analyse.
3. Bij het behandelen van de verzoeken heeft de rol/functie van de aanvrager geen invloed op de prioritering van het verzoek. Echter, de waarde die het voorgenomen onderzoek heeft in (het leren over) de pandemiebestrijding speelt wel een rol in de prioritering.

4.

5. Wetenschappelijke publicaties

1. Ons eigen beleid voor wetenschappelijke publicaties bepaalt niet zozeer welke data we beschikbaar stellen aan derden, maar wel wanneer. Als we eerst zelf bepaalde analyses willen hebben uitgevoerd, kan ervoor gekozen worden deze data nog niet openbaar beschikbaar te stellen. Het is daarom essentieel dat we een helder overzicht hebben over de eigen voorgenomen publicaties en wanneer deze tot een resultaat gaan leiden.
2. Naast de voorgenomen eigen wetenschappelijke publicaties bieden we derden de mogelijkheid een vraag die wij hebben, op te pakken, de daarbij benodigde (geanonimiseerde) data te analyseren en een publicatie op te leveren.
3. Indien we data en analyses verstrekken aan derden (op verzoek, of naar aanleiding van een vraag van ons), dan dienen RIVM/GGD als bron vermeld te worden en worden afspraken gemaakt over acknowledgement of co-auteurschap.
4. Indien partijen gebruik maken van open data op rivm.nl, dient een bronvermelding opgenomen te worden (data bevat een CC-BY licentie).
5. Verzoeken vanuit WAR en andere expertteams dienen apart beoordeeld te worden. Kunnen zij eigenstandig publiceren, of altijd in

samenwerking met ons?

6. Zie de tabel in de bijlage waarin alle mogelijke verzoeken en relevante aspecten worden opgesomd.
- 7.
8. Organisatie
9.
 1. Het loket en de toetsingscommissie die ingericht worden om dataverzoeken in behandeling te nemen, wordt bemenst vanuit de gedragsunit, maar moet na afronding van het corona programma ook geborgd worden.
 2. De toetsingscommissie bevat vertegenwoordigers vanuit data management, privacy expert, GGD en inhoudelijke experts uit WP3/evt. WP11.
 - 3.
 - 4.
 5. Consequenties keuzes
 - 6.
 7. Bij het minimale scenario, waarbij we alleen geaggregeerd data online zetten en vragen die naar aanleiding daarvan opkomen afhandelen, is sowieso nodig:
 - Data management voor online data op rivm.nl (vanuit WP3 organiseren)
 - Loket: vragen behandelen, overzicht over lopende verzoeken, data/analyse uitgeven, agenda Toetsingscie. voorbereiden
 - Bedrijfsvoering: contracten, afspraken, facturering
 - Toetsingscie. voor beoordelen verzoeken (inschatting maandelijkse bijeenkomst)
 - Kwaliteitsmanagement: standaardwerkwijze en formats, dataveiligheid, privacy office, zorgdragen voor toets op resultaten van interne/externe vraagsteller
- Hoeveel fte dit vraagt is moeilijk in te schatten, omdat het niet te voorspellen is hoeveel vragen we te behandelen krijgen, als we een geaggregeerde dataset openbaar beschikbaar maken.
-
- Versterking wordt gezocht op data management, waarbij inhoudelijke kennis van de data ook belangrijk is. Dus eigenlijk nog meer WP3 inzet nodig.
-
- Daarnaast is goede informatievoorziening noodzakelijk over onze eigen lopende/geplande publicatietrajecten. Hiervoor is een linking pin vanuit WP3/11 nodig.
- Vanuit de eigen publicatieplannen(organisatie) is ook de invulling van de onderzoeksvragen nodig die we (inclusief datasets) willen gaan uitzetten. Dit vergt ook enige communicatie inspanning om deze vragen goed te 'pluggen', zodat geïnteresseerde partijen zich hiervoor gaan melden bij het loket met een concreet onderzoeksvoorstel.

Tabel: uitgangspunten omgaan met data uitgifte

Type bestand	Gebruiker intern (RIVM/GGD) of extern? (Controller/verwerker bepaalt doel en middel gegevensverzameling)	Data sharing agreement nodig i.v.m. privacy?	Samenwerkings contract nodig? (afbakenen wat iemand met de data mag doen en hoe wij betrokken zijn)	Afspraak auteurschappen?	Afstemming overlap met eigen publicatieplannen?
1. Geaggregeerde data die online is gepubliceerd, onder CC-BY licentie (online bron)	Intern	Nee	Nee	Nee	Nee
2. Een specifieke dataset wordt verschaft, waarbij AVG-gevoelige zaken zijn verwijderd (geaggregeerd of uit gehoekt)	Extern	Nee	Nee (wel voorwaarden op site opnemen)	Nee	Nee (alleen timing wanneer online)
3. Een specifieke dataset wordt verschaft, waarbij wel AVG-gevoelige zaken worden meegegeven	Intern	Nee	Mogelijk (standaard uitgangspunten van toepassing maken)	Ja	Ja
4. Een analyseresultaat wordt verschaft zonder identificeerbare gegevens	Extern	Nee	Mogelijk	Ja	Ja
5. Een analyseresultaat wordt verschaft zonder identificeerbare gegevens	Intern	Nee	Mogelijk (standaard uitgangspunten van toepassing maken)	Ja	Ja
6. Een analyseresultaat wordt verschaft zonder identificeerbare gegevens	Extern niet mogelijk* (informed consent dekt dit niet)	Nvt*	Nvt*	Nvt*	Nvt*
7. Een analyseresultaat wordt verschaft zonder identificeerbare gegevens	Intern	Nee	Mogelijk (standaard uitgangspunten van toepassing maken)	Ja	Ja
8. Een analyseresultaat wordt verschaft zonder identificeerbare gegevens	Extern	Nee	Mogelijk	Ja	Ja

* De positie van de WAR-leden moet hierin nog bepaald worden. N.B. Deze werkwijze geldt voor verzoeken om data of analyses, maar ook als we pro-actief vragen gaan uitzetten, om anderen onze data te laten analyseren.

-