Inleiding

A. Beschrijving kenmerken gegevensverwerkingen

1. Voorstel

Deze DPIA is opgesteld ter beoordeling van de regionale surveillance applicatie ("clusterbuster").

De oorsprong van de applicatie is de roep vanuit verschillende GGD'en om betere informatievoorziening ten aanzien van de (regionale) surveillance, dat wil zeggen: inzicht in de ontwikkeling van COVID-19 omtrent lokale opleving en clustervorming.

De clusterbuster faciliteert de lokale surveillance en bestrijding van specifiek COVID-19, een meldingsplichtige infectieziekte, waartoe Nederland ook gehouden is onder Besluit EU/1082/2013 van de Europese Unie. En stelt de GGD'en in staat op adequate wijze uitvoering te geven aan hun wettelijke opdracht tot bewaking en bestrijding van de COVID-19 epidemie in hun eigen verzorgingsgebied.

De persoonsgegevens (deel A hoofdstuk 2) die worden verwerkt ten behoeve van de clusterbuster zijn afkomstig uit twee datasets. De eerste dataset is gebaseerd op OSIRIS-AIZ en de tweede dataset is een export uit het GGD-systeem ten aanzien van situations. De verwerkingen vinden plaats conform de doelen waarvoor ze zijn verzameld: namelijk de lokale en nationale bestrijding van infectieziekten (overeenkomstig één van de hoofddoelstellingen van OSIRIS-AIZ) en het verkrijgen en behouden van inzicht in het verloop van de COVID-19 epidemie en het geven van input voor eventuele aanscherpingen van maatregelen/beleid (conform een van de hoofddoelstellingen van de gegevensverzameling situation data direct uit het GGD systeem).

Binnen het RIVM hebben we Regionaal Epidemioloog Consulenten (REC'ers) en Regionaal Arts Consulenten (RAC'ers). Deze REC'ers en RAC'ers zijn betrokken bij de ontwikkeling van de applicatie, als vertegenwoordiging van de eindgebruikers. De eindgebruikers van de applicatie zijn artsen en epidemiologen binnen GGD'en die, vanuit hun medische achtergrond, bestuurders adviseren over de status quo en eventueel te nemen maatregelen.

2. Persoonsgegevens

We maken gebruik van de volgende gegevens, welke enkel aan de bron gepseudonimiseerde indirect herleidbare persoonsgegevens bevatten. Hieronder wordt ook weergegeven welke van deze aan de bron gepseudonimiseerde persoonsgegevens uiteindelijk in geaggregeerde vorm wordt gevisualiseerd via de clusterbuster.

OSIRIS-AIZ

Identificatienummer OSIRIS (bijzonder)

Het OSIRIS identificatienummer gebruiken we om op basis van de situationdata eigenschappen van de bij een situation betrokken patiënten op te halen.

Volledige postcode (PC6) patiënt (gewoon)

De PC6 gebruiken we om te achterhalen in welke wijk of gemeente een patiënt woont.

Deel postcode (PC4) patiënt (gewoon)
De PC4 gebruiken we voor geografische visualisaties van de bij een situation betrokken patiënten.

Naam GGD regio (gewoon)
De naam van een GGD regio gebruiken we om een melding aan een regio toe te verbinden.

Datum eerste ziektedag (bijzonder)
De datum van de eerste ziektedag gebruiken we voor extra inzicht in het verloop binnen situations.

Meldingsdatum GGD (bijzonder)
De datum waarop een melding bij de GGD is gedaan gebruiken we voor het weergeven van het verloop van het aantal meldingen. We gebruiken het daarnaast om weergaves temporeel te beperken.

Geboortedatum (gewoon)
De geboortedatum gebruiken we om een leeftijdscategorie aan patiënten toe te kennen en op basis daarvan meer inzicht te geven in de eigenschappen van een situation.

Geslacht (gewoon)
Het geslacht van een patiënt gebruiken we om meer inzicht te geven in de eigenschappen van een situation.

Beroep (gewoon)
Het beroep gebruiken we om meer inzicht te geven in de eigenschappen van een situation. We visualiseren alleen beroepen die de zorg of het onderwijs betreffen.

Bewoner instelling (bijzonder)
Of een patiënt een bewoner van een instelling is gebruiken we om meer inzicht te geven in de eigenschappen van een situation.

Setting van de besmetting (bijzonder)
De setting van de besmetting gebruiken we om weer te geven hoe de waarschijnlijke locatie waar patiënten de besmetting hebben opgelopen zijn verdeeld over plaats en tijd.

Situationdata

Nummer situation (bijzonder)
Het nummer van een situation geven we weer zodat gebruikers op basis van dat nummer in hun eigen systemen aanvullende informatie kunnen opzoeken. Ter verduidelijking: dit nummer betreft geen individueel patiëntdossier in het GGD systeem, maar het nummer dat door de GGD is toegekend aan een situation ('cluster' of 'outbreak'). Op basis van dit situationnummer kan door de GGD zelf binnen het betreffende GGD systeem worden gevonden welke patiënten er betrokken zijn.

Deel postcode (PC4) situation (gewoon)
De PC4 gebruiken we voor het weergeven van de (bij benadering) locatie van een situation.

Setting van een situation (bijzonder)
De setting, dit betreft het soort locatie of evenement (waaronder, maar niet beperkt tot, 'werkplek', 'ziekenhuis' of 'transport'), van een situation gebruiken we om meer inzicht te geven in de eigenschappen van een situation.

Aantal 'linked cases' van een situation (gewoon)
Het aantal 'linked cases', ofwel het aantal meldingen dat is gekoppeld aan een situation, van een situation gebruiken we om meer inzicht te geven in de eigenschappen van een situation, alsmede

om de eindgebruiker de mogelijkheid te geven een visualisatie te filteren op de grootte van een situation.

Niet alle persoonsgegevens worden in de applicatie gepresenteerd, hieronder staan de persoonsgegevens die wel in de applicatie zijn ontsloten.

OSIRIS-AIZ (visualisatie)

De data wordt getoond in geaggregeerde overzichten (naar wijk, gemeente of GGD-regio). Voor de volgende variabelen wordt op 'regelniveau' een overzicht gegeven.

Eerste ziektedag
In een overzicht met 'linked cases', behorende bij een situation.

Leeftijdscategorie
In een overzicht met 'linked cases', behorende bij een situation.

Zorgmedewerker
Een binaire variabele (wel of geen zorgmedewerker) staat in een overzicht met 'linked cases', behorende bij een situation.

Setting van de besmetting
In een overzicht met 'linked cases', behorende bij een situation.

Situationdata (visualisatie)

De data wordt getoond in geaggregeerde overzichten (naar situation). De volgende variabelen zijn in te zien.

Nummer situation
In een geografisch overzicht met situations.

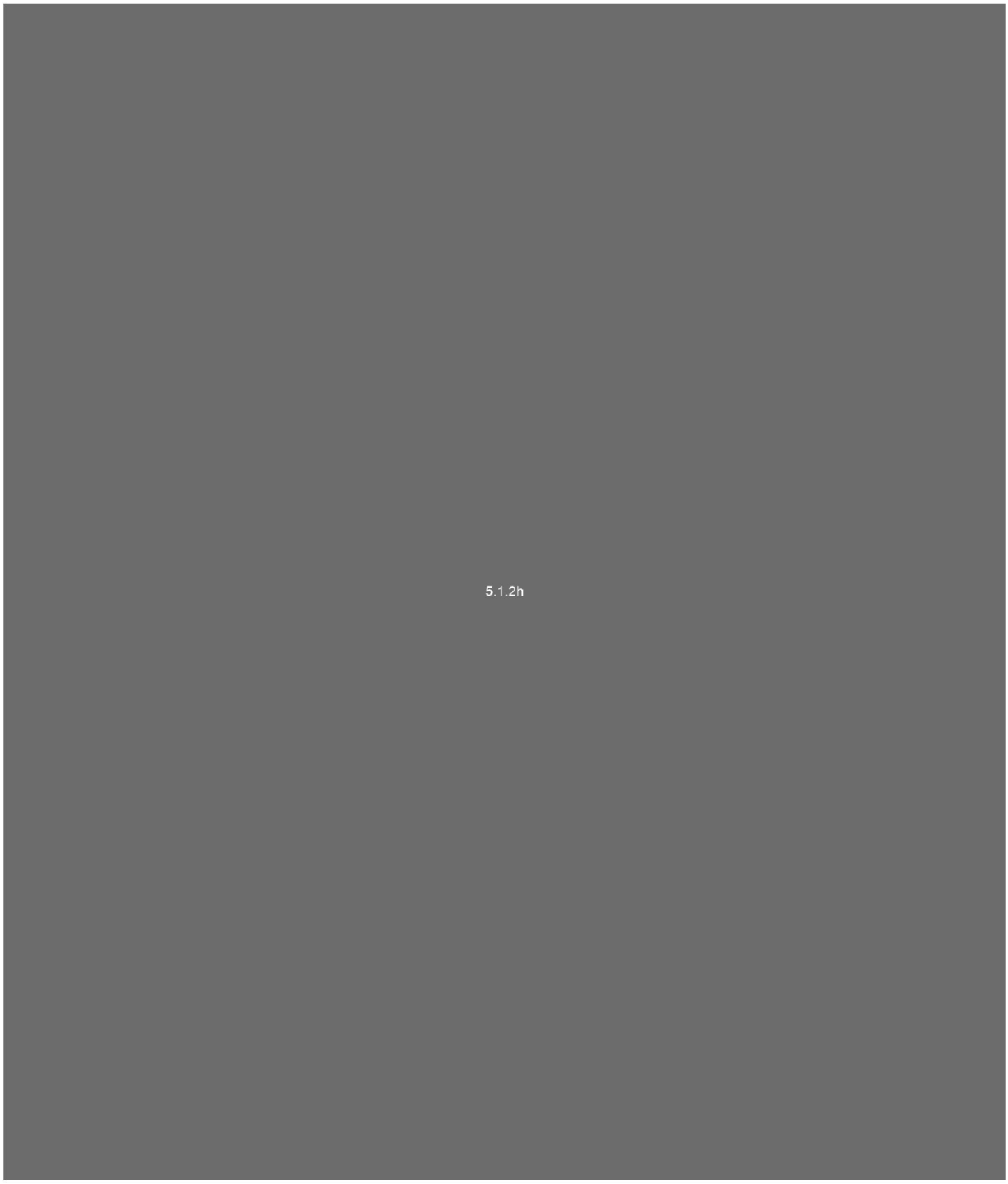
Setting van een situation
In een geografisch overzicht met situations.

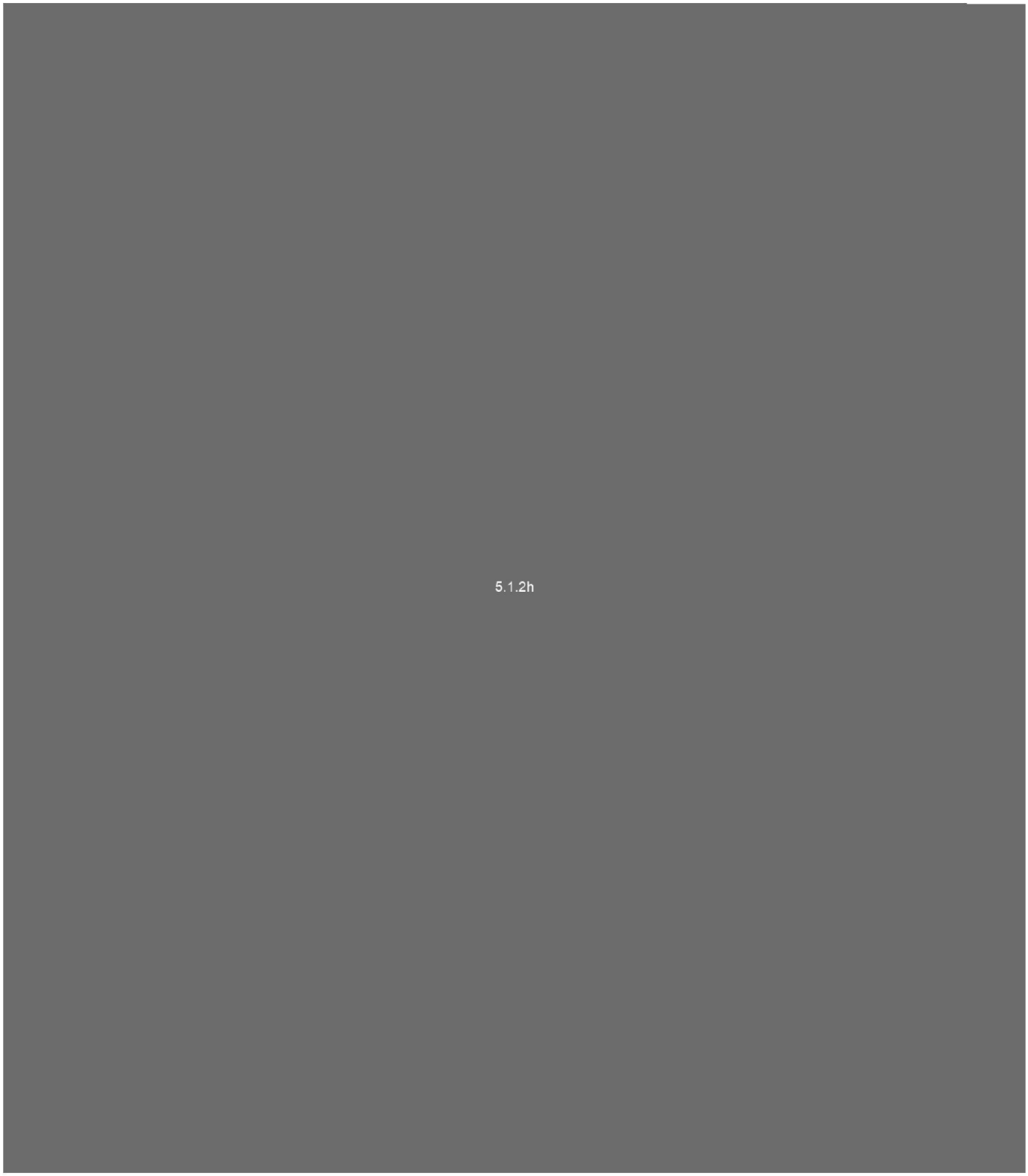
Deel postcode (PC4) van een situation
In een geografisch overzicht met situations.

Aantal 'linked cases' van een situation
In een geografisch overzicht met situations.

3. Gegevensverwerkingen

Genoemde persoonsgegevens (deel A hoofdstuk 2) zijn afkomstig uit twee datasets. De eerste dataset is gebaseerd op OSIRIS-AIZ en de tweede dataset is een export uit het GGD-systeem ten aanzien van situations.





4. Verwerkingsdoeleinden

De clusterbuster faciliteert de lokale surveillance en bestrijding van een meldingsplichtige infectieziekte, te weten COVID-19, waartoe Nederland ook gehouden is onder Besluit EU/1082/2013 van de Europese Unie.

De clusterbuster verschaft artsen en epidemiologen binnen GGD'en die, vanuit hun medische achtergrond, bestuurders adviseren over de status quo en eventueel te nemen maatregelen, inzicht in de ontwikkeling van COVID-19, omtrent lokale opleving, clustervorming of (grotere) uitbraken. De applicatie faciliteert real-time (dagelijks) inzicht in de ontwikkeling van het aantal COVID-19 meldingen, verdeeld over plaats en tijd.

5. Betrokken partijen

Verwerkingsverantwoordelijke

Het RIVM is verwerkingsverantwoordelijke voor de clusterbuster. Vanuit de afdeling Signalering en Surveillance binnen het Centrum Epidemiologie en Surveillance, onderdeel van het Centrum Infectiebestrijding, werken Data Scientists aan de verwerking van de persoonsgegevens, waarbij rechten/autorisaties individueel worden ingesteld.

Verstrekkers

De gepseudonimiseerde data ontvangt het RIVM van alle GGD'en in Nederland, via enerzijds OSIRIS-AIZ en anderzijds direct via het GGD COVID-19 systeem (zie hiervoor deel A, hoofdstuk 3). De GGD'en ontvangen de COVID-19 meldingen van artsen, laboratoria en instellingen uit hun regio, hebben de sleutel van de gepseudonimiseerde data en zijn conform artikel 29 lid 2 Wpg verantwoordelijk voor het tijdig verwijderen van de NAW-gegevens behorend bij een Osiris- of situationnummer, zoals alleen bekend bij de GGD (zie voor toelichting van deze nummers deel A, hoofdstuk 2 en voor toelichting bewaartermijn, hoofdstuk 10). De GGD'en zijn zelfstandig verwerkingsverantwoordelijke over de data in hun eigen verzorgingsgebied en werken in een eigen systeem.

Ontvangers

Een aantal professionals IZB (infectieziektenbestrijding), oftewel artsen en epidemiologen binnen GGD'en die, vanuit hun medische achtergrond, bestuurders adviseren over de status quo en eventueel te nemen maatregelen. Niet alle gegevens die worden verwerkt worden gepresenteerd in de applicatie. Het beperkte aantal eindgebruikers ontvangt geaggregeerde overzichten (naar wijk, gemeente of GGD-regio en situation). De variabelen die wel in geaggregeerde overzichten worden getoond staan vermeld in hoofdstuk 2 en 3.

6. Belangen bij de gegevensverwerking

Voor de verwerkingsverantwoordelijke (RIVM) is het van belang te voldoen aan wettelijke opdracht tot facilitering van de lokale surveillance en bestrijding en de nationale surveillance en vroegwaarschuwing van de meldingsplichtige ziekten, waaronder COVID-19, waartoe Nederland ook gehouden is onder Besluit EU/1082/2013 van de Europese Unie. Het RIVM is daarnaast aangewezen als nationaal coördinatiepunt onder de Internationale Gezondheidsregeling, dat de risicobeoordeling uitvoert of een lokaal of nationaal gezondheidsincident aangemerkt moet worden als een potentieel gezondheidsrisico van internationaal belang (public health event of international concern) en dientengevolge door Nederland onverwijld aan de WHO gemeld dient te worden.

Voor de GGD'en (verstrekkers en eindgebruikers van de clusterbuster) is een adequate uitvoering van hun wettelijke opdracht tot bewaking en bestrijding van de infectieziekten en bestrijding van epidemieën in hun verzorgingsgebied van belang. De nationale gegevensverwerking als onderdeel van de gehele ketenzorg (melding – epidemiologie – laboratoriumtypering – nationale

referentiegegevens - interventies) maakt onderbouwde risicobeoordeling en gerichte interventies door burgemeester of voorzitter van de veiligheidsregio mogelijk.

7. Verwerkingslocaties

De verwerkingen vinden alleen plaats in Nederland. Zie voor exacte locatie databases hoofdstuk 3.

8. Techniek en methode van gegevensverwerking

De verwerking van de datasets (zie deel A hoofdstuk 3) geschiedt middels de programmeertaal R en is grotendeels geautomatiseerd. De datasets worden ingelezen in R Studio en de benodigde variabelen worden opgeslagen in tabellen binnen een SQLite database. De webapplicatie zelf is ook ontwikkeld in R (R Shiny). Voor het ontsluiten van de data via de webapplicatie gebruiken we containertechnologie (middels het OpenShift containerplatform). Versie en revisiebeheer van de code wordt gedaan via het platform Gitlab. Alle hierboven genoemde software draait op servers van het RIVM.

Er is nergens sprake van geautomatiseerde besluitvorming, profilering of andere aan machine learning verwante technieken.

9. Juridisch en beleidsmatig kader

De volgende wet- en regelgeving (en beleid) heeft mogelijke gevolgen voor de gegevensverwerking:

- Besluit EU/1082/2013, art. 6 jo art. 2;
- Wet publieke gezondheid, art. 6c;
- Besluit publieke gezondheid;
- De Wet op het RIVM (artikel 3, lid 1, sub a);
- Besluit ex artikel 3, eerste lid, onderdeel a, van de Wet op het RIVM;
- Wet geneeskundige behandelovereenkomst, art. 7:458 BW;
- RIVM-CIb/LCI richtlijnen;
- Richtlijnen Werkrgroep Infectie Preventie;
- Archiefwet;
- Selectielijst RIVM 2004 – (Staatscourant Nr. 20886, april 2017).
-

10. Bewaartermijnen

De huidige COVID-19 crisis is aangemerkt als zogenaamde Hotspot. Kortgezegd houdt dit in dat de overheid verplicht is om COVID-19-gerelateerde gegevens en mogelijk ook COVID-19-gerelateerde indirect herleidbare persoonsgegevens permanent te bewaren. Op dit moment is vastgesteld dat pas bij het einde van de huidige crisis bepaald wordt wat exact onder de Hotspot regeling dient te vallen. Dit betekent dat de ten behoeve van de clusterbuster geprepareerde datasets voor onbepaalde tijd worden bewaard, totdat over de reikwijdte van de Hotspot is besloten.

Conform artikel 29 tweede lid van de Wpg wordt de sleutel van de gepseudonimiseerde data, en de herleidbare persoonsgegevens, bij de verstrekkende GGD na 5 jaar na opname in de registratie vernietigd; dit betekent dat de NAW-gegevens behorend bij een Osirisnummer of nummer van een situation (zie voor toelichting deel A, hoofdstuk 2), zoals alleen bekend bij de GGD, worden vernietigd. Dit is de verantwoordelijkheid van de GGD. De gepseudonimiseerde persoonsgegevens vervallen daarmee automatisch en hetgeen resteert zijn gegevens niet-herleidbaar tot een persoon.

Data die de eindgebruikers in de applicatie zelf kunnen raadplegen gaat terug tot maximaal vijf weken. Deze data wordt, zoals aangegeven onder hoofdstuk 17, dagelijks vervangen.

B. Beoordeling rechtmatigheid gegevensverwerkingen

11. Rechtsgrond

De verwerking is rechtmatig op grond van artikel 6, lid 1(e) AVG: de gegevensverwerking (inclusief logginggegevens) is noodzakelijk voor de vervulling van een taak van algemeen belang. Deze taak van algemeen belang is vastgesteld in art. 3 lid 1 en 2, van de Wet op het RIVM juncto artikelen 6c, 7, 28 en 50 van de Wet publieke gezondheid (monitoring, nationale surveillance, onderzoek, vroegwaarschuwing en response op uitbraken en epidemieën van infectieziekten) alsook het Besluit ex artikel 3, eerste lid, onderdeel a, van de Wet op het RIVM.

12. Bijzondere persoonsgegevens

De verwerking is noodzakelijk voor redenen van algemeen belang op het gebied van de volksgezondheid. De regionale surveillance applicatie wordt namelijk in het leven geroepen om te helpen in de lokale bestrijding van infectieziekten, door artsen en epidemiologen bij GGD'en inzichten te verschaffen ten behoeve van adequate uitvoering van hun wettelijke opdracht tot bewaking en bestrijding van de infectieziekten en bestrijding van epidemieën in hun verzorgingsgebied.

Voor het doel van het onderzoek is de verwerking van persoonsgegevens noodzakelijk. Onder deze persoonsgegevens vallen tevens bijzondere persoonsgegevens, namelijk gegevens over gezondheid waarvoor in beginsel een verbod op verwerking geldt.

Op de verwerkingen in het kader van dit onderzoek is de uitzondering ex artikel 9, tweede lid onder i van de AVG van toepassing. Er is sprake van:

- een grond in Unierecht of lidstatelijk recht (artikel 23 sub a Uitvoeringswet AVG in relatie met EU Besluit EU/1082/2013 art. 6 jo art. 2, artikel 3 Wet op het RIVM en artikel 6c Wpg);
- waarin passende en specifieke maatregelen zijn opgenomen ter bescherming van de rechten en vrijheden van de betrokkene (de pseudonimiseringsplicht ex artikel 6c.3 Wpg en het beroepsgeheim in art. 7:457 BW en artikel 88 Wet BIG).
-

13. Doelbinding

De persoonsgegevens zijn verzameld voor het doel zoals beschreven onder hoofdstuk 4.

Er is geen sprake van doelwijziging. De verwerkingen die plaatsvinden ten behoeve van de clusterbuster vinden plaats conform de doelen waarvoor ze zijn verzameld: namelijk de lokale en nationale bestrijding van infectieziekten (overeenkomstig één van de hoofddoelstellingen van OSIRIS-AIZ) en het verkrijgen en behouden van inzicht in het verloop van de COVID-19 epidemie en het geven van input voor eventuele aanscherpingen van maatregelen/beleid (conform gegevensverzameling situation data direct uit GGD systeem).

14. Noodzaak en evenredigheid

De voorgenomen gegevensverwerking is proportioneel omdat:

1. De clusterbuster noodzakelijk is om adequate informatievoorziening ten aanzien van de (regionale) surveillance te verschaffen, dat wil zeggen: inzicht in de ontwikkeling van COVID-19 omtrent lokale opleving en clustervorming;
2. Het RIVM door middel van deze gegevensverwerking kan voldoen aan haar wettelijke opdracht tot facilitering van de lokale surveillance en bestrijding van COVID-19;
3. De clusterbuster de GGD'en in staat stelt op adequate wijze uitvoering te geven aan hun wettelijke opdracht tot bewaking en bestrijding van de COVID-19 epidemie in hun eigen verzorgingsgebied;
4. Waar mogelijk dataminimalisatie wordt toegepast, uitsluitend aan de bron gepseudonimiseerde persoonsgegevens worden verzameld en verwerkt, welke uiteindelijk op geaggregeerd niveau worden gevisualiseerd in de clusterbuster ten behoeve van een beperkt aantal eindgebruikers binnen de regionale GGD'en, waarbij de eindgebruikers uitsluitend toegang krijgen tot de visualisaties die betrekking hebben op hun eigen regio.

Daarnaast is bij de gegevensverwerking sprake van subsidiariteit, omdat:

1. Zonder deze gegevens de doelstellingen niet kunnen worden bereikt, omdat het opschoon- en minimalisatieproces van de datasets nodig is om tot de betreffende visualisaties op geaggregeerd niveau te komen, zoals beschreven onder deel A, hoofdstuk 3;
2. Deze gegevens niet op een andere, minder ingrijpende wijze verzameld kunnen worden, aangezien data wordt gepseudonimiseerd aan de bron, wordt opgeschoond en volgens het vier-ogen principe op geaggregeerd niveau wordt gevisualiseerd aan een beperkte groep eindgebruikers (tevens de verstrekkers van de originele datasets).
3.
4. Geconcludeerd kan worden dat de verwerkingen in het kader van de clusterbuster:
1. De belangen van betrokkenen niet onevenredig schaden;
2. Beperkt zijn tot persoonsgegevens die noodzakelijk zijn voor de doeleinden van de voorgenomen gegevensverwerkingen zoals beschreven onder A, hoofdstuk 4; en
3. Niet op een minder ingrijpende wijze verkregen kunnen worden.
4.

15. Rechten van de betrokkenen

Rekening houdend met het doel van de verwerking en de pseudonimisering aan de bron van ingebrachte gegevens, zijn met toepassing van artikel 41 lid 1 onder e en artikel 44 van de Uitvoeringswet AVG, de artikelen 15, 16 en 18 van de AVG niet van toepassing (recht op inzage, rectificatie en gegevenswissing).

Als de verwerking van persoonsgegevens noodzakelijk is voor de uitvoering van een taak van algemeen belang, heeft betrokkene geen recht op bezwaar en zal het verzoek worden afgewezen (art. 21 lid 6 Algemene Verordening Gegevensbescherming).

Als het RIVM persoonsgegevens verwerkt die door het RIVM niet herleidbaar zijn tot een betrokkene, zoals hier het geval is vanwege peudonimisering aan de bron, zijn de rechten van betrokkenen niet van toepassing en zal het verzoek overeenkomstig artikel 11 AVG worden afgewezen.

In het theoretische geval dat iemand zulke gegevens aanlevert dat identificatie wél mogelijk is, dan onderzoekt het RIVM altijd of uitvoering kan worden gegeven aan deze rechten. Hierbij zal tevens worden beoordeeld of een wettelijke beperking op de rechten van toepassing is.

C. Beschrijving en beoordeling risico's voor de betrokkenen

16. Risico's

Hier worden allereerst enkele algemene risico's beschreven en daarna, in lijn met de beschrijving van de gegevensverwerkingen (hoofdstuk 3), worden de risico's per processtap beschreven.

Algemene risico's

Los van de risico's per processtap geldt voor elke processtap een aantal 'standaard' risico's die zich kunnen voltrekken als men niet volgens de wettelijke kaders en RIVM-protocollen werkt. Voorbeelden zijn beveiligingsincidenten/datalekken, het niet tijdig archiveren dan wel verwijderen van data, onrechtmatige verdere verwerking van persoonsgegevens, een onzorgvuldig autorisatiebeheer en een verlies van grip op data governance. Die risico's zijn inherent aan elke gegevensverwerking en dus elke processtap.

Om deze risico's tot een aanvaardbaar niveau terug te brengen zijn maatregelen getroffen; zie hiervoor onderdeel D, hoofdstuk 17 en bijlage B (QuickScan BIO) en bijlage C (informatiebeveiliging risicoanalyse).

Stap 1: overzetten situationdata 5.1.2h

Een risico bij overplaatsing van de situationdata is dat wanneer de data over een onbeveiligde verbinding zou worden verzonden, het risico bestaat dat kwaadwillenden de data onderscheppen. De kans dat dit zich voordoet is gemiddeld (zie voor mitigaties van dit risico hoofdstuk 17). De impact van dit risico op het subject van de data is laag, omdat de gegevens al aan de bron gepseudonimiseerd zijn.

Stap 2: datapreparatie door het R-team van situationdata

Een identificeerbaar risico binnen deze processtap is dat een of meerdere Data Scientists van het R-team tijdens het datapreparatieproces van het onbewerkte .RDS bestand willens en wetens of per ongeluk data naar buiten brengen.

De kans dat een van deze risico's zich voordoet is gemiddeld, waarvoor mitigerende maatregelen zijn gepresenteerd onder hoofdstuk 17 om deze kans zoveel mogelijk te minimaliseren; de impact is laag, omdat uitsluitend wordt gewerkt met aan de bron gepseudonimiseerde gegevens.

Stap 3: datapreparatie door het CB-team van situationdata

Op drie punten binnen deze processtap zijn risico's te identificeren. Bij het inlezen van de data, de bewerking van de data en de opslag van de data.

Inlezen van de data

Bij het inlezen van de data is het mogelijk dat oude data wordt ingelezen terwijl de veronderstelling bestaat dat het actuele data is. De kans dat dit risico zich voordoet is door de werkwijze van het R-team laag (zie hoofdstuk 17) en de impact op het subject van de data is tevens laag.

Bewerkingen van de data

Door programmeerfouten bestaat het risico dat variabelen foutief bewerkt worden, denk hierbij aan de onjuiste toekenning van een leeftijd in jaren op basis van de geboortedatum. In lijn met de inschatting van voorgaand risico (verkeerd inlezen van de data) is de kans dat dit risico zich voordoet laag. De impact op het subject van de data is tevens laag.

Opslaan van de data

Ten aanzien van de opslag van data bestaat het risico dat de data op een onjuiste locatie of onder

een verkeerde naam wordt opgeslagen. In lijn met de eerder genoemde risico's onder stap 3 is de kans dat dit risico zich voordoet laag. De impact op het subject van de data is tevens laag.

Stap 4: overzetten data 5.1.2h

Deze processtap valt binnen de scope van de DPIA 5.1.2h waarin risico's voor betrokkenen in kaart zijn gebracht en de geïmplementeerde maatregelen zijn beschreven om deze naar een aanvaardbaar niveau terug te brengen.

Stap 5: datapreparatie door het R-team van data 5.1.2h

Hier bestaat het risico dat een of meerdere Data Scientists van het R-team tijdens het datapreparatieproces van het onbewerkte .RDS bestand willens en wetens of per ongeluk data naar buiten brengen.

De kans dat een van deze risico's zich voordoet is gemiddeld, waarvoor mitigerende maatregelen zijn gepresenteerd onder hoofdstuk 17 om deze kans zoveel mogelijk te minimaliseren; de impact is laag, omdat uitsluitend wordt gewerkt met aan de bron gepseudonimiseerde gegevens.

Stap 6: datapreparatie door het CB-team van data 5.1.2h

Op drie punten binnen deze processtap zijn risico's te identificeren. Bij het inlezen van de data, de bewerking van de data en de opslag van de data.

Inlezen van de data

Bij het inlezen van de data is het mogelijk dat oude data wordt ingelezen terwijl de veronderstelling bestaat dat het actuele data is. De kans dat dit risico zich voordoet is door de werkwijze van het CB-team laag (zie hoofdstuk 17) en de impact op het subject van de data is tevens laag.

Een tweede risico bij het inlezen van de data is dat er onvoldoende sprake is van dataminimalisatie in een voorgaande processtap (stap 5). Hierdoor is het mogelijk dat leden van het CB-team meer gegevens te zien krijgen dan strikt noodzakelijk. De kans dat dit risico zich voordoet is hoog. De impact op het subject van de data is, doordat de data aan de bron gepseudonimiseerd is, laag.

Bewerkingen van de data

Door programmeerfouten bestaat het risico dat variabelen foutief bewerkt worden, denk hierbij aan de onjuiste toekenning van een leeftijd in jaren op basis van de geboortedatum. In lijn met de inschatting van voorgaand risico (verkeerd inlezen van de data) is de kans dat dit risico zich voordoet laag. De impact op het subject van de data is tevens laag.

Opslaan van de data

Ten aanzien van de opslag van data bestaat het risico dat de data op een onjuiste locatie of onder een verkeerde naam wordt opgeslagen. In lijn met de eerder genoemde risico's onder stap 6 is de kans dat dit risico zich voordoet laag. De impact op het subject van de data is tevens laag.

Stap 7: opslaan van de clusterbuster SQLite database 5.1.2h

Een identificeerbaar risico onder deze stap is dat de data foutief wordt opgeslagen (onjuiste locatie of onder een verkeerde naam). In lijn met de inschatting onder stap 3 is de kans dat dit risico zich voordoet laag. Ook de impact op het subject van de data is onder dit risico laag.

Een ander risico is dat onbevoegden toegang krijgen tot de data. Doordat het aantal medewerkers dat toegang heeft tot de betreffende locaties 5.1.2h beperkt is, is de kans dat dit risico zich voordoet laag. De impact op het subject van de gepseudonimiseerde data bij dit risico is tevens laag.

Stap 8: plaatsing van de clusterbuster SQLite database 5.1.2h en presentatie in applicatie

Identificeerbare risico's binnen deze verwerkingsstap zijn te plaatsen op twee punten. Het opslaan van de data en het presenteren van de data.

Opslaan van de data

De data zoals deze in stap 6 is geprepareerd door het CB-team wordt in de 5.1.2h omgeving geplaatst (zie voor meer toelichting op deze omgeving hoofdstuk 8). Enkele leden van het CB-team (maximaal 5) hebben toegang tot deze omgeving. Hiervandaan wordt de data geautomatiseerd ingeladen in 5.1.2h Een risico in deze stap is dat onbevoegden toegang krijgen tot de data. Daar de autorisaties zeer beperkt zijn is de kans dat dit risico zich voordoet laag. Het risico voor het subject van de gepseudonimiseerde data is tevens laag.

Presenteren van de data in de applicatie

Een risico bij de presentatie van de gepseudonimiseerde data is dat in extreme gevallen de combinatie van kenmerken herleidbaar zou kunnen zijn. De kans dat dit risico zich voordoet is laag en de impact op het subject van de data is gemiddeld. Daarnaast bestaat het risico dat onbevoegden toegang krijgen tot de applicatie. Door de zeer beperkte autorisaties is de kans dat dit risico zich voordoet laag. Door de aard van de inzichten in de applicatie is de impact van dit risico op het subject van de data tevens laag.

D. Beschrijving voorgenomen maatregelen

17. Maatregelen

Hier worden allereerst enkele algemene mitigerende maatregelen beschreven en daarna, in lijn met de beschrijving van de gegevensverwerkingen (hoofdstuk 3), worden de maatregelen per processtap beschreven die de hiervoor beschreven risico's tot een aanvaardbaar niveau terugbrengen.

Algemene maatregelen

Door de hele keten van verwerkingen met betrekking tot de clusterbuster wordt door een beperkt aantal geautoriseerde Data Scientists (maximaal 15, waarvan 10 in het R-team en 5 in het CB-team) uitsluitend met aan de bron gepseudonimiseerde persoonsgegevens gewerkt, welke uiteindelijk op geaggregeerd niveau worden gevisualiseerd in de clusterbuster ten behoeve van een beperkt aantal eindgebruikers binnen de regionale GGD'en, waarbij de eindgebruikers uitsluitend toegang krijgen tot de visualisaties die betrekking hebben op hun eigen regio. Iedereen die op enig moment toegang krijgt tot gepseudoninimeerde persoonsgegevens (op geaggreageerd niveau), is gehouden aan contractuele geheimhoudingsverplichtingen. Adequate data governance is voorts een doorlopend punt van aandacht (zie meer over data governance in hoofdstuk 17).

Er zijn informatiebeveiligingsmaatregelen getroffen om deze verwerkingen zo veilig mogelijk in te richten. Zie hiervoor bijlage B (QuickScan BIO) en bijlage C (informatiebeveiliging risicoanalyse). Uit de risicoanalyse zijn geen restrisico's naar voren gekomen.

Stap 1: overzetten situationdata 5.1.2h

Het overzetten van de data van de systemen 5.1.2h geschiedt via een SSH File Transfer Protocol (SFTP), hetgeen betekent dat er bij de transfer sprake is van end-to-end encryptie. Zie voor meer details over deze transfer van data bijlage A.

Stap 2: datapreparatie door het R-team van situationdata

In deze stap wordt gewerkt met maximaal 10 Data Scientists van het R-team die allen gehouden zijn aan contractuele geheimhoudingsverplichtingen. Autorisatiebeheer vindt plaats op individueel niveau. Adequate data governance is een doorlopend proces en punt van aandacht binnen het R-team.

Stap 3: datapreparatie door het CB-team van situationdata

Persoonsgegevens zijn voor het CB-team alleen bereikbaar via de virtuele omgeving van het RIVM, welke toegankelijk is na twee-factor-authenticatie. Daarbij geldt dat de Data Scientists uit het CB-team (maximaal 5 medewerkers) expliciete autorisatie moeten krijgen alvorens er toegang kan worden verkregen tot de bestanden 5.1.2h

Het CB-team werkt volgens de principes van dataminimalisatie. In deze verwerkingsstap wordt enkel een selectie van variabelen uiteindelijk opgeslagen in de SQLite database. Om er zorg voor te dragen dat de scripts die hiervoor ontwikkeld zijn geen fouten bevatten is er sprake van controle op de code middels het 4-ogen principe. Daarnaast is er sprake van versie- en revisiebeheer, zodat bij fouten in de code altijd navolgbaar is waar een fout zich heeft voorgedaan en deze op passende wijze hersteld kan worden.

Zodoende worden enkel de reeds geprepareerde variabelen (volgend op stap 2) verwerkt die noodzakelijk zijn voor het functioneren van de applicatie, worden ook alleen deze variabelen opgeslagen in de SQLite database en is de bewerking beperkt tot een selecte groep van Data Scientists binnen het CB-team die werken volgens principes van adequate data governance.

Stap 4: overzetten data 5.1.2h

Deze processtap valt binnen de scope van de DPIA 5.1.2h waarin risico's voor betrokkenen in kaart zijn gebracht en de geïmplementeerde maatregelen zijn beschreven om deze naar een aanvaardbaar niveau terug te brengen.

Stap 5: datapreparatie door het R-team van data 5.1.2h

Net zoals in stap 5, wordt in deze stap wordt gewerkt met maximaal 10 Data Scientists van het R-team die allen gehouden zijn aan contractuele geheimhoudingsverplichtingen. Autorisatiebeheer vindt plaats op individueel niveau. Adequate data governance is een doorlopend proces en punt van aandacht binnen het R-team.

Stap 6: datapreparatie door het CB-team van data 5.1.2h

Voor deze processtap gelden dezelfde maatregelen als voor stap 3.

Stap 7: opslaan van de clusterbuster SQLite database 5.1.2h

Ook voor deze processtap is er sprake van dezelfde maatregelen als in stappen 3 en 6.

Stap 8: plaatsing van de clusterbuster SQLite database in 5.1.2h en presentatie in applicatie

De 5.1.2h omgeving is een afgeschermde omgeving waarbij het niet alleen noodzakelijk is om allereerst in te loggen op de virtuele omgeving van het RIVM (met twee-factor-authenticatie), maar waarbij toegang tot de 5.1.2h omgeving voor de clusterbuster zelf ook beperkt is tot de medewerkers van het CB-team. Ook het 5.1.2h is enkel toegankelijk voor deze medewerkers. De SQLite database wordt 5.1.2h bovendien dagelijks vervangen, waardoor het niet mogelijk is om via deze omgeving oudere databases te raadplegen.

De applicatie zelf is enkel toegankelijk nadat een eindgebruiker is geautoriseerd voor de applicatie. De eindgebruiker krijgt vervolgens toegang tot de applicatie na twee-factor-authenticatie, welke

wordt ingeregeld door medewerkers van Identity Acces Management van het RIVM. De eindgebruiker krijgt vervolgens in de applicatie enkel de data te zien die van belang is voor zijn of haar verzorgingsgebied. De data in de applicatie gaat nooit verder terug dan vijf weken. De weergaven in de applicatie betreffen tot slot geaggregeerde overzichten van de gepseudonimiseerde data.