To: 5.1.2e [5.1.2e @rivm.nl]; 5.1.2e [5.1.2e @rivm.nl]
From: 5.1.2e
Sent: Wed 2/3/2021 9:32:57 AM
Subject: RE: Issues m.b.t. specificaties geanonimiseerde data
Received: Wed 2/3/2021 9:32:58 AM

Hier kan je b.v. een 'salted hash' voor gebruiken. Een hash is een vingerafdruk van een stuk tekst (technisch een reeks bitjes). Je kan van de tekst wel een hash berekenen, maar je kan nooit vanuit de hash de oorspronkelijke tekst ontsleutelen. Je kan dit extra veilig maken door aan een identificerende tekst een 'salt' toe te voegen, dus een tekst die alleen de persoon die de hast berekent weet.
Vanuit die vingerafdruk kan je nooit terug naar de oorspronkelijke waarde, tenzij je de oorspronkelijke waarde, de willekeurige tekst en het hashingmechanisme weet.
Gebruik je b.v. het interne technische id van de persoon in het systeem (dus niet het BSN!!) plus een willekeurige tekst dan zijn dit gegevens die wij als RIVM niet hebben en we kunnen daarmee dus nooit terug naar een persoonsgegeven. Sterker nog, al weet je het hashing mechanisme en de willekeurige tekst wel, dan kunnen we alleen een reeks interne technische sleutels berekenen….. Je komt dus nooit bij een persoonsgegeven uit waarmee wij data kunnen koppelen.
Je kan bij een volgende prik via dezelfde berekening weer hetzelfde nummer berekenen, maar dus nooit terug naar het id.
Deze eenwegsversleuteling wordt alleen berekend op het moment dat er een export wordt gemaakt.

Wij doen aan onze kant hetzelfde in BI-CIMS. Dus aan onze data kant kunnen we de ID die daar zichtbaar is nooit meer relateren aan de ID die in het bestand staat, want die slaan we domweg niet op.
Ook al wéét je het hash algorithme en de willekeurige tekst die we bij het RIVM gebruikt hebben om de sleutel om te zetten, zonder de originele waarde kan je de RIVM sleutel niet berekenen….

Op die manier hebben we ook aan de kant van het RIVM zekerheid dat we de privacy geborgd hebben en voldoet het RIVM aan haar verplichtingen.
Dat op de correcte manier de code aan de kant van de leverancier bepaald wordt lijkt me een verantwoordelijkheid aan de kant van de leverancier.

Met vriendelijke groet,

5.1.2e
5.1.2e

SSC-Campus
Rijksinstituut voor Volksgezondheid en Milieu
Ministerie van Volksgezondheid, Welzijn en Sport
Postbak 86 | Postbus 1 | 3720 BA Bilthoven
T +31 5.1.2e
E-mail : 5.1.2e @rivm.nl
http://www.ssc-campus.nl

From: 5.1.2e <5.1.2e @rivm.nl>
Sent: woensdag 3 februari 2021 10:13
To: 5.1.2e <5.1.2e @rivm.nl>; 5.1.2e <5.1.2e @rivm.nl>
Subject: RE: Issues m.b.t. specificaties geanonimiseerde data

Sorry 5.1.2e, zo was het zeker niet bedoeld. Ongelukkige woorden van mij.

Leg me dan nog één ding uit: hoe weet een HIS of EVS dat het meegezonden ID van de 1e prik weer gebruikt moet worden bij de 2e prik. Dat moet dat toch lokaal zijn vastgelegd?

Groet,
5.1.2e

From: 5.1.2e <5.1.2e @rivm.nl>
Sent: woensdag 3 februari 2021 09:38
To: 5.1.2e <5.1.2e @rivm.nl>; 5.1.2e <5.1.2e @rivm.nl>

Subject: RE: Issues m.b.t. specificaties geanonimiseerde data

Beste 5.1.2e

We hebben dit gisteren besproken en ik heb het daarna conform afspraak de impact overdacht en ben tot de conclusie gekomen dat het niet werkbaar is de sleutels uit de specificaties te verwijderen.
We weten dan namelijk niet meer of iets een eerste of tweede prik is en ik vermoed dat 5.1.2e zich niet bewust is van dit feit. Het gaat dus verder dan het niet kunnen relateren van de eerste en tweede prik:
zonder dit gegeven kan je geen vaccinatietoestand en vaccinatiegraad vaststellen.

Ook heb ik in voorstel 2 ook een methode beschreven waarbij deze sleutels van de leveranciers niet opgeslagen wordt in onze systemen! Dus zowel in de systemen van de leverancier als in onze systemen kan je deze sleutels niet terugvinden en dus niet gebruiken om deze aan een persoon te relateren. Het wordt alleen gebruikt in de communicatie met het RIVM en om de prikken van een persoon te koppelen, verder niet. Het is daarmee geen pseudoniem van een persoon, maar een koppeling van de prikjes. Niet alleen m.i., dit is reeds bevestigd is door de privacy specialisten van RIVM en VZVZ, zonder de maatregel dat we deze sleutel in de RIVM systemen niet terug laten komen.

Kortom, volgens mij heb ik volkomen integer en conform afspraak gehandeld, stellen dat ik me 'er niet zoveel van aan lijk te trekken' vind ik een tamelijk ongepaste reactie.

Met vriendelijke groet,

5.1.2e
5.1.2e

SSC-Campus
Rijksinstituut voor Volksgezondheid en Milieu
Ministerie van Volksgezondheid, Welzijn en Sport
Postbak 86 | Postbus 1 | 3720 BA Bilthoven

T +31 5.1.2e
E-mail : 5.1.2e @rivm.nl
http://www.ssc-campus.nl

From: 5.1.2e < 5.1.2e @rivm.nl>
Sent: dinsdag 2 februari 2021 22:17
To: 5.1.2e < 5.1.2e @rivm.nl>; 5.1.2e < 5.1.2e @rivm.nl>
Subject: RE: Issues m.b.t. specificaties geanonimiseerde data

Ik heb met 5.1.2e zijn reactie doorgesproken.
Ik heb 5.1.2e op twee dingen gewezen waar hij zich niet zoveel van aan lijkt te trekken, nl.

a. 5.1.2e zegt dat het toevoegen van een willekeurig ID aan de verzending alleen maar kan door dat ID toe te voegen aan de persoonsgegevens van de patiënt/client, anders kun je hetzelfde ID niet reproduceren t.b.v. de 2e prik. Daarmee is sprake van pseudonimisering en valt het derhalve onder de AVG; betekent dit een zwaarder regime, zoals DPIA?
b. 5.1.2e hecht niet zoveel waarde aan het heel precies kunnen koppelen van 1e en 2e prik

En ik zeg het nog maar eens: dit is een bijstroom. Het is data waarover de persoon in kwestie heeft gezegd dat hij deze niet wil delen met het RIVM. Naar mijn mening is de enige rechtvaardiging voor het verkrijgen van enige data uit deze categorie de wens om de vaccinatiegraad te kunnen vaststellen, een beetje geografisch (want eerste 2 cijfers postcode) en naar leeftijd.

Groet,
5.1.2e

From: 5.1.2e < 5.1.2e @rivm.nl>
Sent: dinsdag 2 februari 2021 20:37

To: 5.1.2e <5.1.2e @rivm.nl>; 5.1.2e <5.1.2e @rivm.nl>
Subject: RE: Issues m.b.t. specificaties geanonimiseerde data
Importance: High

Beste 5.1.2e ,

Vanmiddag heb ik overleg gehad met 5.1.2e . Dit ging in eerste instantie over een ander onderwerp, maar we hebben ook de aanlevering van anonieme gegevens besproken.
We delen de zorg dat we in de anonieme stroom waarschijnlijk onvolledige data binnenkrijgen waar we weinig mee kunnen. Deze zorg wordt ook gedeeld door 5.1.2e .
Het gaat goed als het percentage opt-in boven de 95% blijft, daaronder zullen we toch meer gegevens (AGB Code, woonplaats) moeten weten om de monitoring te kunnen doen.

Verder krijgt 5.1.2e nu al de data van de GGD GHOR binnen en we hebben er even in kunnen kijken. Deze set wijkt op een aantal punten af van wat wij gespecificeerd hebben en is meer in lijn met de oorspronkelijke specificatie in DPV_203 en DPV_210:

- Het gaat om alle gegevens (dus opt-in en niet opt-in), waarbij er een vinkje is dat aangeeft of de persoon opt-in is;
- AGB code wordt niet aangeleverd;
- Prik één en Prik twee worden op een regel weergegeven;
- Weeknummer ipv datum;
- Naast woonplaats wordt ook PC3 niveau geleverd (dat voldoet dus in feite niet aan de eisen);

Alle input combinerend wil ik de volgende wijzigingen voorstellen:

1. AGB code vervalt: deze is er waarschijnlijk door mij bijgezet zodat de data overeenkomt met de reguliere aanlevering. Dit zou in theorie handig kunnen zijn bij een recall, maar goed we weten ook aan wie we een batch hebben uitgeleverd. Staat niet in het oorspronkelijke kaderdocument van 5.1.2e DPV 203 (hebben we vanmiddag gecheckt), dus die kunnen we m.i. gewoon weglaten.
2. Naast Postcode 2 niveau worden alle aanleveringen van mensen boven de 90 jaar meegenomen als 90 jarigen (dus niet meer uitgesplitst per jaar, maar bijeen 'geveegd').
   Ik heb even naar de statistieken van CBS gekeken, per leeftijd gaat bij een aantal PC2 niveaus toch niet goed (n<10). Voor 95 plus wel, dus met 90+ zit je altijd veilig.
3. Sleutel om 'prikken' te relateren handhaven: deze sleutel is noodzakelijk omdat we zonder deze sleutel niet kunnen zien of een prik een eerste of een tweede prik is. Je weet dan nooit hoeveel mensen er überhaupt geprikt zijn of hoeveel mensen de reeks hebben afgemaakt....

We hadden in een eerder formaat conform DPV_203 de eerste, tweede en derde prik van een persoon op één regel in het formaat staan, i.c.m. het aanleveren van alle gegevens vanaf de start van de campagne bij elke levering. Dit werd door de leveranciers afgewezen en als onwerkbaar geacht. Ook een aanlevering waarbij zij (zoals BRBA wel doet) aangeven of het om de eerste of tweede prik gaat was óók niet werkbaar omdat in de systemen het om losstaande voorschrijvingen zou gaan, er moet dan teruggekeken worden in een grote bak met data. Kortom, werd ook afgekeurd.

Als we deze sleutel laten vallen kunnen we m.i. deze hele import net zo goed laten vallen, want naast het feit dat we niet weten om welke prik het gaat én zijn deze gegevens op geen enkele manier te combineren met de gegevens van de GGD.
NB zowel privacy specialisten van het RIVM én VZVZ hebben naar deze sleutels gekeken en zij zagen hier **geen** bezwaar in, mits de sleutel alleen voor de communicatie met het RIVM gebruikt wordt. **Voorstel 1:** we gaan deze discussie niet herhalen, maar we voegen de voorwaarde dat deze sleutel alleen gebruikt wordt in de communicatie richting het RIVM toe aan de beschrijving van de sleutels. **Voorstel 2** We passen op deze sleutel een zelfde algoritme toe als we bij de normale slag van CIMS naar BI-CIMS toepassen voordat we de data opslaan in BI-CIMS. Hierdoor komt de sleutel nooit in het data platform van de rapportageomgeving terecht en is behalve dat het in het importbestand staat ook voor het RIVM niet meer zichtbaar in de verdere verwerking. Hiermee voorkomen we m.i. privacy issues in de rapportageomgeving.

4. Mutaties doorgeven: de suggestie van 5.1.2e om bij wijziging een mutatie aan te maken lijkt me aan onze kant werkbaar. Dat zou kunnen door een kolom "Toevoegen/Verwijderen" toe te voegen met de waarde "toevoegen" of "verwijderen". Ik weet echter niet of dit praktisch uitvoerbaar is aan de kant van de leveranciers.
   Alternatief is dat we net als bij GGD GHOR alle data anoniem aan laten leveren, dus ook de opt-in, met een vinkje dat aangeeft of iemand óók via de reguliere stroom aangeleverd wordt (Opt-in j/n). Op die manier hebben we altijd een compleet beeld voor de vaccinatiegraad zonder dubbeltellingen, door deze alleen op de anonieme stroom te baseren. Maar dan op PC2 niveau. Voor alle andere rapportages (bv op niveau gemeente of veiligheidsregio) kunnen we dan alleen baseren op de stroom mét persoonsgegevens. Dit is te verdedigen, we mogen die data niet hebben op dit niveau. Kortom, hier valt nog wel wat te kiezen morgen.

Met vriendelijke groet,

5.1.2e
5.1.2e

SSC-Campus
Rijksinstituut voor Volksgezondheid en Milieu
Ministerie van Volksgezondheid, Welzijn en Sport
Postbak 86 | Postbus 1 | 3720 BA Bilthoven

T +31 (5.1.2e
E-mail : 5.1.2e @rivm.nl
http://www.ssc-campus.nl

From: 5.1.2e < 5.1.2e @rivm.nl>
Sent: dinsdag 2 februari 2021 15:02
To: 5.1.2e < 5.1.2e @rivm.nl>; 5.1.2e < 5.1.2e @rivm.nl>
Subject: Issues m.b.t. specificaties geanonimiseerde data
Importance: High

5.1.2e, ik zet de issues rond geanonimiseerde data op een rij in de hoop dat we hier nu eindelijk snel tot vaststelling kunnen komen.

Eigenlijk is het één grote knoop: oplossing voor het ene issue creëert weer een ander issue.

De ingrediënten:
a. De dataset sluit niet uit dat data herleidbaar is tot een persoon: zie mijn notitie van 1 februari jl.
b. Leveranciers willen incrementeel aanleveren, maar kan tot dubbeltellingen leiden als een wijziging van opt in tussen 1e en 2e prik doorgevoerd wordt
c. LHV en NHG willen vasthouden aan de mogelijkheid om altijd terug te komen op een opt in, ook tussen 1e en 2e prik

Naar mijn idee komen we er alleen maar uit als we terug gaan naar de dataset en bij de noodzakelijke aanpassing proberen de punten b en c te honoreren.
Ik heb hier vanmorgen over gesproken met 5.1.2e
5.1.2e geeft aan dat het geboortejaar (> 90 jaar afgekapt) en de eerste twee cijfers van de postcode het allerbelangrijkste zijn. Zij is bereid in te leveren:

- Geslacht
- AGB code: kan die gemist worden 5.1.2e?
- De koppeling tussen 1e en 2e prik, want is de redenering van 5.1.2e : een identificerend nummer voor de vaccinatiereeks en een identificerend nummer voor de toediening zijn identificerende gegevens die in het register aan een persoon hangen en feitelijk een pseudonimisering is en daardoor onder AVG valt (zie ook mijn notitie van gisteren).

5.1.2e deed ook de suggestie om van wijziging opt in een 'mutatie' te maken, ter vookoming van dubbeltellingen. Dit laatste probeer ik me voor te stellen, maar moet een knappe kop toch eens beredeneren.

Stel we doen dit: zouden we dan alle drie de problemen oplossen?

Graag jullie mening.

Groet,
5.1.2e

5.1.2e
5.1.2e

Rijksinstituut voor Volksgezondheid
en Milieu
Ministerie van Volksgezondheid,
Welzijn en Sport

Programma Covid Informatie en Monitoringsysteem (CIMS)
A. van Leeuwenhoeklaan 9 | 3721 MA | Bilthoven
Postbus 1 | 3720 BA | Bilthoven als
M +31 5.1.2e
E 5.1.2e @rivm.nl
www.rivm.nl
De zorg voor morgen begint vandaag