



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

A. van Leeuwenhoeklaan 9
3721 MA Bilthoven
Postbus 1
3720 BA Bilthoven
www.rivm.nl

KvK Utrecht 30276683

T 030 274 91 11
S.I.S. @rivm.nl

Datum
4 februari 2021

Ons kenmerk

Uw kenmerk

Behandeld door
5.1.2e

Kopie aan

Bijlage(n)
DPV_210 Aanlevering
geanonimiseerde
vaccinatiedata V1.2x.doc

memo

Borging anonimiteit bij geanonimiseerde
vaccinatiedata RIVM

In het document *DPV_210 Aanlevering geanonimiseerde vaccinatiedata 1.2* wordt de aanlevering van geanonimiseerde covidvaccinatiegegevens aan het RIVM gespecificeerd. Hierbij wordt de data in de vorm van een csv-bestand (comma separated values) aangeleverd.

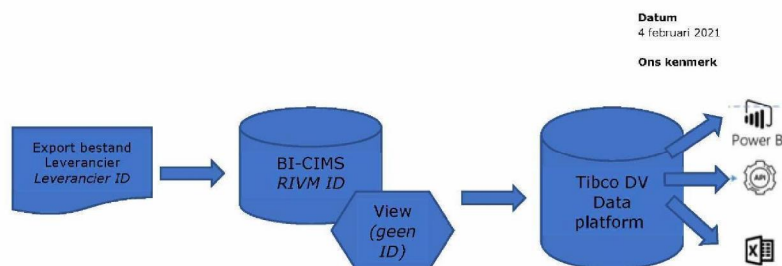
Deze specificatie maakt het op verzoek van de leveranciers mogelijk om deze data elke dag incrementeel aan te leveren.

Aangezien het voor de monitoring door het RIVM noodzakelijk is dat bekend is welke vaccinaties aan één persoon toegediend zijn is een technische sleutel toegevoegd aan de specificatie die niet aan een natuurlijk persoon te relateren is.

Deze code is per leverancier uniek en wordt alleen gebruikt voor de communicatie met het RIVM.

Voor de verwerking van de bestanden is een methodiek bedacht waarbij het niet nodig is om deze technische sleutel van de leveranciers één op één op te slaan in de database van het RIVM (BI-CIMS). Hierdoor wordt het onmogelijk om aan de hand van de gegevens in de BI-CIMS database terug te gaan naar de persoonsgegevens in de systemen van de leverancier.

Een schematische weergave van de import en verdere verwerking van de data staat in Figuur 1.



Figuur 1 schematische weergave verwerking anonieme vaccinatiegegevens

De volgende stappen worden doorlopen tijdens de import:

1. De leverancier van de data levert het exportbestand aan via de voorgeschreven route. Dit bestand bevat per toegediende vaccinatie een technische sleutel (Leverancier ID) waarmee de vaccinaties van één persoon aan elkaar gerelateerd kunnen worden.
2. De data wordt verwerkt. Hierbij wordt de kwaliteit van de data gecontroleerd. De technische sleutel wordt vervangen door een intern RIVM ID en de anonieme vaccinatiegegevens worden opgeslagen in de BI-CIMS-database in een aparte tabel voor de anonieme import. Omdat dit RIVM ID een andere code is dan aangeleverd is door de leverancier kan er vanuit de BI-CIMS database nooit teruggegaan worden naar het systeem van de leverancier om met behulp van deze sleutel een persoon op te zoeken.
3. De gegevens worden verzameld in een view waarbij de gegevens per gevaccineerde in één rij komen te staan: geboortjaar (>90 wordt als 90 weergegeven), geslacht, eerste twee cijfers postcode, selectie criterium, gegevens eerste toediening (datum, HPK-code, Batchnummer) en gegevens tweede toediening (datum, HPK-code, Batchnummer). Deze view bevat verder geen sleutels meer.
4. De view wordt ontsloten via het Tibco Dataplatform en de rapportageomgeving. In het platform is geen toegang tot zowel de Leverancier ID als de interne RIVM ID.

De RIVM sleutel wordt bepaald door middel van een 'salted hash'. Een hash is een digitale vingerafdruk van een bestand of een stuk tekst en uniek voor het bestand of de tekst waarvan de hash berekend wordt: de kans dat er twee bestanden zijn die dezelfde hash code hebben is minder dan 1 op 4 miljard. Hierdoor kan je ervan uit gaan als je de hash code van

twee bestanden of teksten vergelijkt en deze identiek zijn dit ook geldt voor de inhoud van het document.

Datum
4 februari 2021

Omdat de hash code niet de informatie bevat van het origineel, kan je vanuit een hash nooit het oorspronkelijke bestand of de oorspronkelijke tekst reconstrueren.

Ons kenmerk

Je zou wél voor een reeks willekeurige teksten de hashcode kunnen berekenen en kunnen vergelijken met de waarden in de database: door een berekende hash te vergelijken met deze in de database zou je op die manier via een repeterend proces kunnen zoeken naar een match. Om dit te voorkomen wordt de oorspronkelijke technische sleutel van de leverancier uitgebreid met een 'salt' voordat de hash bepaald wordt. Dit is een stukje tekst dat alleen bekend is in de import. Zonder deze 'salt' is het onmogelijk een correcte RIVM ID te berekenen.

Nadat het RIVM ID gebruikt is om de vaccinaties van één persoon aan elkaar te relateren wordt deze niet meer gebruikt en is deze verder niet meer beschikbaar binnen de het data platform en de rapportageomgeving.

Voordelen van deze aanpak:

- De leveranciers ID is een code welke in het reguliere systeem van een leverancier niet te relateren is aan een persoon omdat deze alleen in de communicatie met het RIVM gebruikt wordt. Het wordt alleen gebruikt in het bestand dat aan het RIVM wordt aangeleverd.
- Het RIVM ID is de vingerafdruk van het leveranciers ID. Zowel in de overige systemen van het RIVM (b.v. CIMS) als de leveranciers is dit verder niet bekend en ook niet te relateren aan een natuurlijk persoon. We kunnen daarmee wel de vaccinaties van een persoon aan elkaar relateren, maar nooit terug naar het leveranciers ID.
- Het RIVM ID wordt alleen gebruikt bij de eerste stap van de import in de BI-CIMS-database. De view waarbij de gegevens van een persoon op één regel gecombineerd wordt bevat dit gegeven niet meer. Deze view wordt gebruikt voor de verdere verwerking. Hierdoor is zowel het Leveranciers ID als het RIVM ID niet beschikbaar op de plek waar de data daadwerkelijk wordt gebruikt: het dataplatform en de rapportageomgeving.