Pseudonimsering in COVI

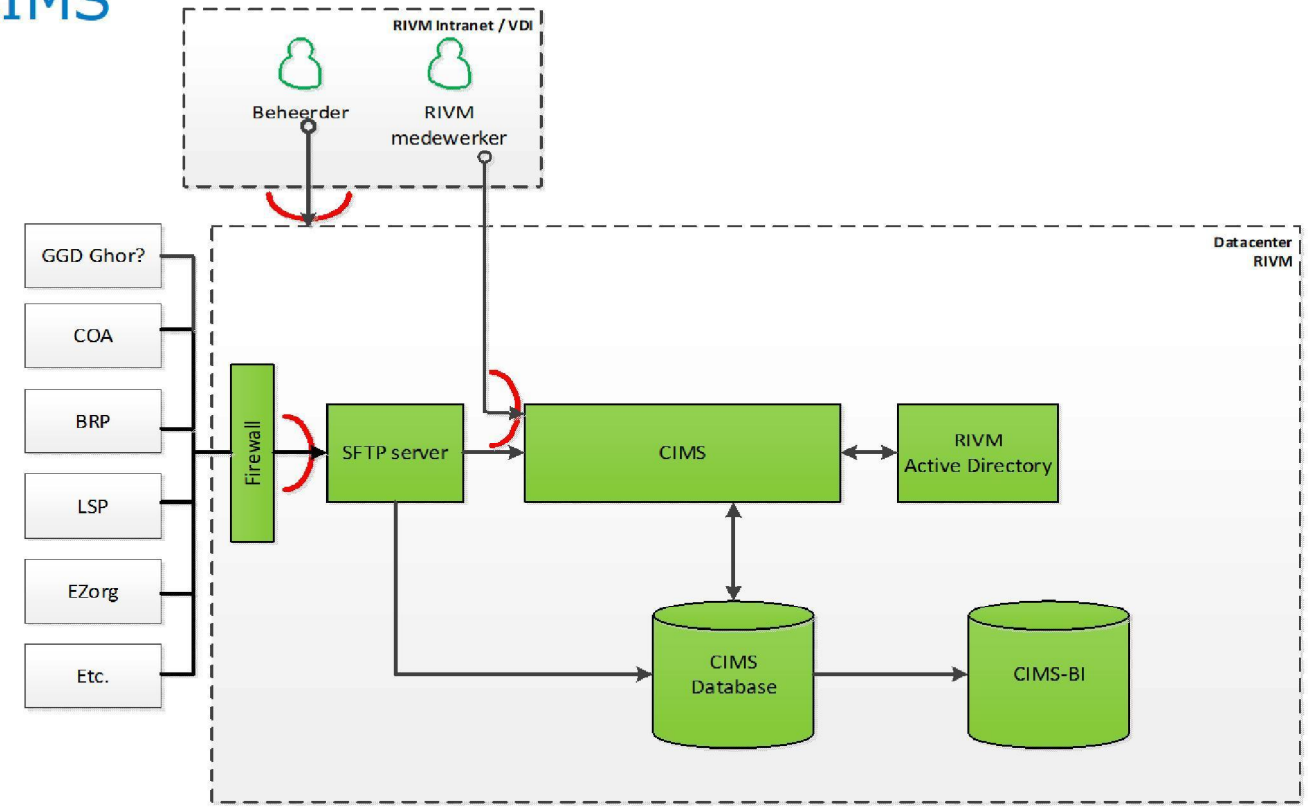
Overwegingen t.a.v. van inzet pseudonimisering en anonimisering in de rapportage- en monitoringvoorzieningen rond CIMS (BI-CIMS)

CIMS | 18-12-2020 | versie 0.2 | ter bespreking

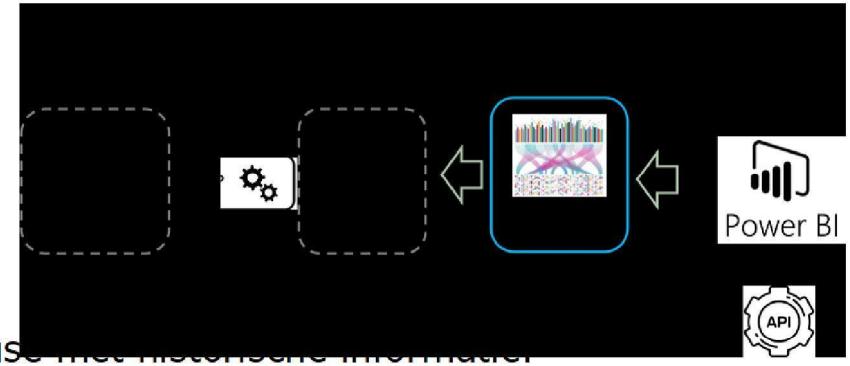
Ter inleiding

- Deze presentatie is bedoeld om het vraagstuk rond pseudonimisering in BI-CIMS te verkennen en de keuzes hieromtrent inzichtelijk te maken.
- De uiteindelijke keuzes zullen opgenomen worden in de PSA COVID Informatie Rapportage.
- Deze PSA heeft versie 0.5 bij het maken van deze presentatie. De informatie hierin wordt bekend verondersteld.
- Voor alle duidelijkheid: DIT IS EEN CONCEPT.

CIMS


RIVM Intranet / VDI
Beheerder
RIVM medewerker
Datacenter RIVM
GGD Ghor?
COA
BRP
LSP
EZorg
Etc.
Firewall
SFTP server
CIMS
RIVM Active Directory
CIMS Database
CIMS-BI

BI-CIMS



Hieronder enkele relevante eigenschappen van de beoogde oplossing:

- BI-CIMS is een datawarehouse met historische informatie.
- Daarnaast is BI-CIMS een schaduwkopie van CIMS waarop batchtaken uitgevoerd worden.
- De voeding van BI-CIMS bestaat uit een vrijwel volledige kopie van CIMS, inclusief persoonsinformatie. Er is nog geen anonimisering of pseudonimering ingebouwd.
- De CIMS-machine voor evaluatie en planning van op basis van vaccinatieschema's werkt op basis van informatie in CIMS, maar levert zijn resultaat af in BI-CIMS.
- Het RIVM heeft DataVirtualisatie (DV) als oplossing gekozen voor beheer, ontsluiting en geïntegreerd bevragen van data. De inrichting van DV moet nog gebeuren voor CIMS.

Anonimisering en Pseudonimiseren

- De CIMS-database bevat (bijzondere) persoonsgegevens. Dit is inherent aan haar doel: vaccinatieregister. De belangrijkste hiervan zijn
  - Identificerende nummers (BSN, A-nummer, Zorgnummer, RNI-deelnemersnummer)
  - Naamgegevens, adresgegevens
  - Geboortedatum
  - Geslacht
- BI-CIMS moet allereerst geaggregeerde informatie opleveren. Verwerking van de voornoemde persoonsgegevens moet daarbij indien enigszins mogelijk voorkomen worden. Anonimisering en pseudonimisering kunnen daarbij helpen. In beide gevallen worden unieke referenties naar personen vervangen door betekenisloze nieuwe referenties. Bij anonimisering is dit onomkeerbaar; bij pseudonimisering is de wel omkeerbaar. Daarnaast moeten de overige persoonsgegevens zodanig aangepast worden dat hun combinatie niet meer herleidbaar is tot individuele personen. Bijvoorbeeld naar:
  - Pseudoniem, geboortejaar en -maand, geslacht en postcode (cijfers of geheel).
- Van belang is dat bij pseudonimisering en anonimisering steeds dezelfde vertaling naar een unieke betekenisloze referentie plaatsvindt, zodat informatie gerelateerd aan dezelfde persoon wel bij elkaar houden kan worden. Bij voorkeur gebeurt dit over bronsystemen heen.
- RIVM heeft nog geen centraal beleid of service voor pseudonimisering. Het voorgaande punt geeft aan dat dit wel wenselijk is. DV kan wel aansluiten op een dergelijke dienst, maar voorziet er zelf ook niet in. Bij voorkeur richt RIVM daarom op zeer korte termijn een dergelijke dienst in. In eerste instantie kan dit eenvoudig met alleen BSN's. Later zou dit uitgebreid kunnen worden met indificaties voor personen die geen BSN hebben.

Functionele behoefte

- Uit de functionele behoefte blijkt dat pseudonimisering (i.p.v. anonimisering) gewenst is. Na rapportage en analyse kan het wenselijk zijn nader onderzoek te doen naar individuele gevallen uit een bepaalde doelgroep. Depseudonimisering is daarvoor nodig, uiteraard omgeving met de nodige waarborgen daarvoor.
- De functionele behoefte is geschetst in de PSA.
  1. Hierbij staan rapporten genoemd die volledig met gegevens uit CIMS te realiseren zijn (opkomst – voor zover gedaan vanuit CIMS, vaccinatiegraad)
  2. Ook staan rapporten genoemd die gegevens uit CIMS combineren met informatie uit Osiris (effectiviteit) en Lareb (veiligheid)
- De rapporten bij 2. worden niet gemaakt door DVP. In beide gevallen levert DVP op basis van aangeleverde persoonsgegevens de vaccinatiegegevens terug aan de beherende organisatie. Hiervoor bestaat een grond. Dit is voor het RVP een handmatige activiteit. Gezien de verwachte aantallen bij Covid-19 moet dit geautomatiseerd gebeuren.
- Daarnaast moet vastgesteld welke andere functies op BI-CIMS draaien die ook persoonsgegevens nodig hebben.

Oplossingsrichtingen

Pseudonimisering in	CIMS	BI-CIMS	DataVirtualisatie
	Pseudoniem in CIMS geregistreerd bij persoon	Pseudoniem in DWH van BI-CIMS geregistreerd bij persoon	Pseudoniem toegevoegd in DataVirtualisatie
Vraagt aanpassing van	CIMS ETL naar BI-CIMS	BI-CIMS Views voor ontsluiting naar DV	Inrichting
Verwerking Persoonsgegevens (m.n. BSN)	CIMS	CIMS BI-CIMS	CIMS BI-CIMS DataVirtualisatie
Gevolgen voor CIMS-machine	De-pseudonimisering signaallijsten nodig.	Geen	Geen
Levering vaccinatiegegevens aan andere systemen	Alleen vanuit CIMS mogelijk	Uit CIMS en BI-CIMS mogelijk	
Eigen rapportages	Geen, indien doelgroepen nog steeds identificeerbaar	Geen, indien doelgroepen nog steeds identificeerbaar	Geen

Oplossing lange termijn

Deze oplossing is de oplossing 'onder architectuur'. Hij heeft de volgende kenmerken:

- CIMS bevat de volledige actuele status van een persoon. De CIMS-machine moet zijn resultaat dus in CIMS plaatsen (i.p.v. BI-CIMS).
- Een RIVM-brede functie is beschikbaar voor het toekennen van pseudoniemen aan personen op basis van de volledige set bekende persoonsgegevens.
- In CIMS wordt aan elk persoon een pseudoniem toegekend.
- In BI-CIMS en DV wordt uitsluitend met het pseudoniem gewerkt. De overige gegevens alleen voor zover nodig om groepen te bepalen (Pseudoniem, geboortejaar en -maand, geslacht en postcode (cijfers of geheel).
- Aanlevering van vaccinatieinformatie buiten het RIVM vindt plaats op basis van bevraging van CIMS met behulp van persoonsgegevens (voorbeeld: Lareb). De functies hiervoor worden gerealiseerd op CIMS.
- Gecombineerde rapportages met vaccinatieinformatie door organisatieonderdelen van het RIVM worden gemaakt in DataVirtualisatie. Matching van personen gebeurt op basis van pseudoniemen.

Nadere uitwerking pseudonimisering

- De volgende attributen worden tweeweg-gepseudonimiseerd:
  - Clientnummer
  - BSN
  - Alle technische sleutels van tabellen die éénduidig relateren naar Clienten, worden éénweg gepseudonimiseerd.
  - Overige gegevens overgenomen conform voorstel op attribuutniveau.
- CIMS-Keys databaseschema voor bewaring pseudonimiseringmateriaal.
- Apart schema voor depseudonimisering.
- Twee-weg pseudonimisering via AES-256. Sleutels in CIMS-keys-schema (zie volgende sheet).
- Eén-weg pseudonimisering via SHA-256 hashing. SALT-key in CIMS-keys-schema (zie volgende sheet).

	Sleutelkast	Hashing	Encryptie AES-256
Eénweg-pseudonimisering	Indien voorkeur niet mogelijk. Toegang tot depseudonimiseren blokkeren.	Meest geëigend. Voorkeur.	Minder geëigend. Toegang tot decryptie blokkeren
Tweeweg-pseudonimisering	Indien andere niet mogelijk.	Niet mogelijk	Voorkeur.
Eigenschappen	- Eenvoudig toepasbaar - Bij onthulling individuele gevallen geen risico totale set - Dataformaat ID's gelijk - Extra tabel met persoonsgegevens opgeslagen - Sleutels toegankelijk voor DBA.	- Geen herleiding terug naar bron mogelijk - Oracle-packages genereren geen uniek nummer. Niet bruikbaar. - MD5-hash werkt wel. MD5 echter verouderd. Rijksstandaard nu SHA-2 - Evt. SALT-key toevoegen - Number(8) -> varchar(40)	- Industriestandaard - Borgen sleutelbeheer. In CIMS-keys-schema. - Bij compromitering sleutelmateriaal onthulling alle data mogelijk - Number(8) -> varchar(40) - Evt. RSA

Functiescheiding

Rol	Functies

Sleutelbeheerder (!= BI-CIMS data analist)	- Depseudonimisering Clientnummer en BSN
	  - Individueel
	  - Batch
	- Toevoegen persoonsgegevens in batch na depseudonimisering
	- Beheer sleutelmateriaal (incl. reservekopieën)
	- Update sleutelmateriaal initiëren

BI-CIMS beheer	- Uitvoeren inlezen CIMS-gegevens - Bevragen pseudonimisering, één- en tweeweg
BI-CIMS data analist (!= Sleutelbeheerder)	- Rapporteren uit BI-CIMS
Technisch beheerders met hogere rechten - DBA - Ontwikkelaar GRIP	- Scheiding via database vault, indien mogelijk. - Beheermaatregelen RIVM voor gebruik beheerfuncties
Dataeigenaar	- Toestemming tot doorbreken pseudonimisering

Oplossing korte termijn

De hiervoor geschetste oplossing voor de lange termijn is niet uitvoerbaar binnen de korte tijdslijnen van CIMS/BI-CIMS, mede omdat dit ontwikkelingen buiten de aansturing van het CIMS-programma betreft. Om toch op korte termijn de benodigde functionaliteit te leveren, biedt het volgende de oplossing:

- CIMS genereert bij elke persoon een zelfgekozen pseudoniem. Dit pseudoniem wordt meegenomen richting BI-CIMS en DV. Er is slechts één functies met logging op read-operaties die de terugvertaling kan doen.
- Er komt een functie op BI-CIMS om vaccinatiegegevens op te vragen. Dit ten behoeven van Lareb en EPI (Osiris) Zodra het resultaat van de CIMS-machine is overgezet naar CIMS, wordt deze functies ook overgezet naar CIMS. Deze functie kan de volgende vormen hebben:
  - Webservice:
  - Bestandsimport en export: Lijsten met persoonsgegevens worden geïmporteerd, vervolgens gematched tegen de personen in CIMS, daar wordt de lijst mét vaccinatiegegevens geëxporteerd.
- In DV worden geen persoonsgegevens verwerkt, behalve degene die nodig zijn voor rapportage. Naar verwachting is dat: Pseudoniem, geboortejaar en -maand, geslacht en postcode (cijfers of geheel). BI-CIMS biedt hiervoor views aan.