Technische specificaties dataverwerking HPZone

# Inleiding

Data vanuit HPZone is niet direct inzichtelijk met externe software. Er bestaat momenteel geen API, exporteermogelijkheden zijn beperkt tot CSV/XML/Excel. Om epidemiologisch onderzoek uit te voeren is directe toegang tot de data nodig. Dit document beschrijft daarom een opzet om geëxporteerde data vanuit HPZone om te zetten naar een database, welke gebruikt kan worden voor o.a. dashboarding in Power BI en analyses in R.

# Overzicht

Data wordt binnen HPZone in drie categorieën geregistreerd:

* Case – een melding van één infectieziektegeval bij één individu
* Situation – een uitbraak of melding van meerdere gevallen
* Enquiry – (telefonische) vraag gesteld door een externe partij

Hierbij is het relevant om te vermelden dat de onderliggende verwekker of infectieziekte niet éénduidig wordt geregistreerd tussen de verschillende types. Zo heeft een case altijd een diagnose, maar niet per se een infectie of een verwekker. Een situation, daarentegen, heeft altijd een verwekker, maar geen infectie of diagnose. Een enquiry heeft wel een infectie, maar deze worden anders genoemd dan binnen cases. Hierdoor is een netwerk aan vertaaltabellen nodig. Daarnaast zijn er verscheidene infecties die leiden tot één klinisch beeld, waardoor er eveneens een groepering nodig is. Zo zijn er bijvoorbeeld drie typen paratyfus, die voor de surveillance als één gezien kunnen worden. Daarom is er een extra groepeertabel toegevoegd, waarnaar de andere vertaaltabellen kunnen verwijzen. (Hieronder *infectieziekte* genoemd.) Zie het diagram in figuur 1 voor een globaal overzicht.

# Kruistabellen

Een aantal elementen komen overeen tussen de verschillende registraties. Deze worden dan ook zoveel mogelijk gedeeld in losse tabellen, zodat deze later makkelijk kunnen worden aangepast bij wijzigingen binnen HPZone. De gedeelde tabellen worden hieronder verder toegelicht. Een aantal van deze tabellen kunnen rechtstreeks gekopieerd worden uit de datadump van HPZone. Deze is echter niet compleet. Daarom heeft een projectgroep, bestaande uit Loes Jaspers (arts IZB, GGD NOG), Lieke Winkeler (vpk IZB, GGD NOG) en Aart Dijkstra (epi, GGD NOG), deze handmatig uitgewerkt en ingevuld. Relevante ziekteverwekkers, infecties en diagnoses zijn gegroepeerd op al dan niet meldingsplichtige ziektes middels de groeperingstabel *infectieziekte*.

## Groepering

De groepering van infectieziekten is een ingewikkeld proces. De verschillende indelingen binnen HPZone verwijzen niet goed naar elkaar, en zijn daarnaast niet volledig ingedeeld in meldingsplichtige ziektes. Zo heeft de **infectie** Hepatitis B bijvoorbeeld een link naar de **agent** Hepatitis B virus, maar geen link naar een diagnose, aangezien er meerdere klinische beelden worden meegenomen in HPZone. In dit geval is de **agent** Hepatitis B handmatig gelinkt aan **infectieziekte** Hepatitis B, waardoor de **infectie** Hepatitis B ook leidt tot vermelding in de categorie van **infectieziekte** Hepatitis B. De **diagnose** Hepatitis A is wel gelinkt aan de **agent** Hepatitis A, waardoor de **infectieziekte** Hepatitis A kan worden aangenomen, en de **infectie** Hepatitis A, welke gelinkt is aan de **agent**, automatisch ook deze status krijgt. In de datadump van een case komt de waarde ‘Hepatitis A’ echter niet voor bij de diagnoses; deze heet daar ‘Acute hepatitis A’. Zoals deze voorbeelden illustreren is dit proces lastig te omschrijven, waardoor een exacte kopie van de pilotdatabase waarschijnlijk het meest efficiënt is. Deze indeling is als bijlage toegevoegd aan dit document. Daarnaast worden de volgende correcties uitgevoerd tijdens het inladen van de cases:

Tabel 1: Reguliere expressies voor het verwerken van inconsistente notatie binnen HPZone. **Let op:** de syntax is geschreven in R, wat betekent dat een backslash (\) tweemaal voorkomt. Binnen een andere taal dienen deze slechts éénmaal voor te komen, bijvoorbeeld \1 versus \\1. Let ook op het gebruik van \\L in de vervangende waarde bij hepatitis, deze wordt niet in iedere taal ondersteund.

|  |  |
| --- | --- |
| Zoekterm | Vervangende waarde |
| ^(Acute|Chronic) hep\\w+ (\\w) | Hepatitis \\2, \\L\\1 |
| Hepatitis (A|E).\* | Hepatitis \\1 |
| Streptococcal septicae.\* | Streptococcal Group A septicaemia |
| Streptococcal septic art.\* | Streptococcal Group A septic arthritis |
| Streptococcal (toxic shock syndrome|meningitis) | Streptococcal Group A \\1 |
| Streptococcal pneumonia \\(Group A\\) | Streptococcal Group A pneumonia |
| Haemophilus influenzae, spec.\* | Haemophilus influenzae infection, specified |
| Malaria$ | Malaria, unspecified |
| (Avian) (Influenza) | \\2, \\1 |

## Agent

Een agent (biological agent) verwijst binnen HPZone naar een ziekteverwekker. Dit betreft in de meeste gevallen een bacterie of virus. Deze worden centraal, voor alle GGD’en gelijk, opgeslagen in HPZone. Binnen HPZone zijn deze soms verbonden aan een diagnose, vooral wanneer een bepaalde bacterie altijd tot één ziektebeeld leidt. Dit is echter niet altijd het geval. Wanneer er binnen HPZone een standaarddiagnose is ingesteld, en deze diagnose valt onder de meldingsplichtige ziektes, dan kan de infectieziekte worden afgeleid van deze diagnose. De tabel is dan ook opgebouwd door een datadump vanuit HPZone samen te voegen met de datadump van diagnoses. Verdere indeling is handmatig doorgevoerd. (Zie inleidende paragraaf van dit hoofdstuk.)

## Context

De context waarin een melding zich afspeelt duidt op de fysieke locatie van de situatie. Dit kan bijvoorbeeld thuis zijn (‘Gezinssituatie’), op de BSO (‘Kinderopvang’) of in het AZC (‘Asielzoekerscentra’). Contexten kunnen binnen HPZone worden toegevoegd, aangepast en verwijderd, en worden gedeeld tussen de verschillende GGD’en. Het kan daardoor nodig zijn om bij een wijziging binnen HPZone de tabel aan te passen. Deze tabel wordt dynamisch samengesteld door de voorkomende waarden in de dataset. Plat gezegd wordt bij iedere datadump gekeken of de genoemde waardes voorkomen in de tabel, en toegevoegd wanneer dit niet het geval is.

## Diagnose

Diagnoses worden binnen HPZone centraal vastgelegd. Deze zijn te downloaden als datadump. Sommige diagnoses hebben een gelinkte meldingsplichtige ziekte en/of agent. Deze zijn overgenomen. Diagnoses waar dit niet het geval is zijn handmatig bijgewerkt door de projectgroep. (Zie inleidende paragraaf van dit hoofdstuk.)

## Infectie

Infecties worden eveneens centraal vastgelegd en zijn gelinkt aan een agent, maar niet aan een diagnose. Deze link is, waar mogelijk, uitgebreid door de indeling van de agents en diagnoses over te nemen. Verder is deze tabel handmatig aangevuld. (Zie inleidende paragraaf van dit hoofdstuk.)

## Medewerker

Medewerkers kunnen een case, situation of enquiry ontvangen danwel verwerken. De terminologie hierin verschilt per type. Bij een case wordt gesproken over een ‘case manager’ en een ‘investigating officer’, bij een situation over een ‘manager’ en een ontvanger (‘entered by’), bij een enquiry gaat het om een ontvanger (‘originally taken by’) en een verwerker (‘handled by’). Voor de interne consistentie heet de verwerker (‘investigating officer’ bij een case) altijd medewerker en de tweede persoon manager of ontvanger, afhankelijk van de context.

## Onderwerp

Onderwerpen komen alleen voor bij vragen (enquiry), en zijn anders gespeld dan infecties of diagnoses, ook als deze wel overeen (zouden moeten) komen. Deze zijn eveneens handmatig gegroepeerd per infectieziekte. (Zie inleidende paragraaf van dit hoofdstuk.) Binnen HPZone wordt dit ‘Specific topic’ genoemd.

## Scenario

Een scenario komt alleen voor bij een situation. Dit omschrijft een minder specifieke indeling van de gemelde infectieziekte. Bij een uitbraak van norovirus, bijvoorbeeld, zou het scenario ‘Gastroenteritis’ zijn, en de agent ‘Norovirus’. Deze tabel wordt dynamisch samengesteld door de voorkomende waarden in de dataset. Plat gezegd wordt bij iedere datadump gekeken of de genoemde waardes voorkomen in de tabel, en toegevoegd wanneer dit niet het geval is.

## Geografische data

Vanuit alle registraties kan een postcode worden gehaald, welke omgezet kan worden naar een gemeentecode en -naam. Aangezien dit een algemene handeling betreft wordt deze in dit document niet verder uitgewerkt. De gebruikte code is echter wel aanwezig in het meegeleverde script.

# Verwerking data

HPZone biedt geen mogelijkheid om de data direct te downloaden of uit te wisselen via een API. Exporteren middels een CSV/XML/Excel-bestand is de enige optie. Aangezien de open tekstvelden kunnen zorgen voor fouten in de verwerking bij een CSV is Excel eigenlijk de enige functionele optie. Dit gebeurt momenteel handmatig, voor zowel cases als situations als enquiries. Dit betekent eveneens dat aanpassingen in de data pas worden overgenomen als de relevante tijdsperiode wordt gedownload. Aangezien de coronaperiode voor een explosie in het aantal gevallen heeft gezorgd, welke de exportfunctie niet meer aankan, kan slechts een beperkte periode worden gedownload. In de praktijk is de selectie ‘in dit jaar’ meestal voldoende. (In januari-maart wellicht ook voorgaand jaar meenemen.)

Na het downloaden van de drie relevante datadumps kunnen deze ingevoegd worden in de database. Let hierbij op het mogelijk al bestaan van de casuïstiek in de database; pas deze aan waar nodig. Het script waarmee deze dataverwerking momenteel wordt uitgevoerd is toegevoegd als bijlage.

# Uitvoer

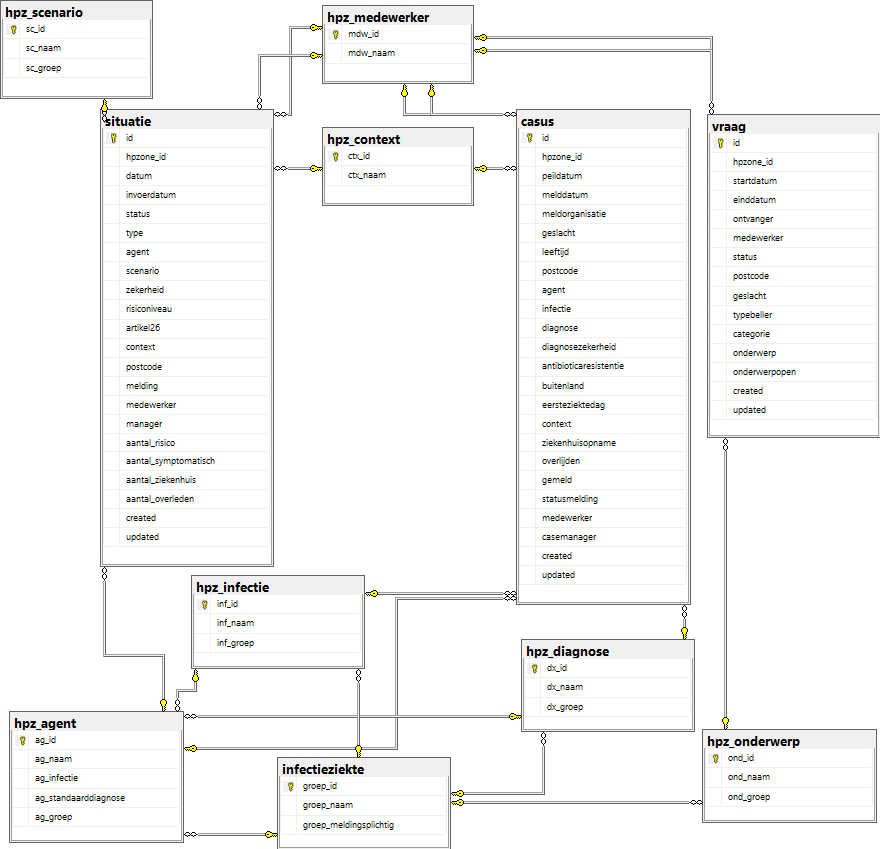
Zoals eerder beschreven dient de database beschikbaar te zijn vanuit verschillende applicaties. Voor gebruik door artsen en verpleegkundigen is Power BI de meest logische kandidaat, voor epidemiologen zal R een waardevolle toevoeging zijn. Om de wirwar van onderliggende tabellen te kunnen ontwijken en ter preventie van latere aanpassingen aan producten vanwege veranderingen in de onderliggende datastructuur is er een drietal views meegeleverd, zie de bijlage en tabel 2.

Tabel 2: Uitvoer van de datastructuur voor verwerking in analysesoftware, bijvoorbeeld Power BI of R. N.B.: De term ‘groep’ verwijst naar de groepering waarin de verwekker, infectie of diagnose zich bevindt. Dit correspondeert met de tabel *infectieziekte*.

|  |  |  |
| --- | --- | --- |
| vw\_cases |  |  |
| Veld | **Type** | **Verklaring** |
| id | Getal |  |
| hpzone\_id | Getal | ID in HPZone, landelijk uniek |
| peildatum | Datum | Eerste ziektedag, melddatum, of invoerdatum, in die volgorde |
| melddatum | Datum | Melddatum of invoerdatum, in die volgorde |
| geslacht | M/F/U | Geslacht in drie categorieën: man (M), vrouw (F), onbekend (U) |
| meldorganisatie | Tekst | Organisatie die de melding gedaan heeft. Dit kan zijn: Arbodienst / werkgever, AWBZ zorginstelling, GGD, Huisarts, Kinderopvang, Laboratorium, Onderwijsinstelling, Overig, Verloskundige, Ziekenhuis |
| leeftijd | Getal |  |
| postcode | Getal | Viercijferige postcode, zonder letters |
| gemeentecode | Getal | Code van de gemeente waarin de postcode zich bevindt, volgens CBS gebiedsindeling |
| gemeentenaam | Tekst |  |
| groep | Tekst | Groep waarin deze melding zich bevindt. Dit veld is een samenstelling van de onderliggende groepen van agent, infectie en diagnose, in die volgorde. |
| meldingsplichtig | 0/1 | Is de groep waaraan deze melding gekoppeld is meldingsplichtig? |
| agent | Tekst | Ziekteverwekker |
| agent\_groep | Tekst | Groep waarin de verwekker zich bevindt |
| agent\_meldingsplichtig | 0/1 | Is de groep waaraan deze verwekker gekoppeld is meldingsplichtig? |
| infectie | Tekst | Infectie |
| infectie\_groep | Tekst | Groep waarin de infectie zich bevindt |
| infectie\_meldingsplichtig | 0/1 | Is de groep waaraan deze infectie gekoppeld is meldingsplichtig? |
| diagnose | Tekst | Gestelde diagnose |
| diagnose\_groep | Tekst | Groep waarin de diagnose zich bevindt |
| diagnose\_meldingsplichtig | 0/1 | Is de groep waaraan deze diagnose gekoppeld is meldingsplichtig? |
| diagnosezekerheid | Tekst | Niveau van bevestiging van de casus. Mogelijke waarden: Confirmed, Discarded, Possible, Probable |
| context | Tekst | Context waarin de casus zich afspeelt (thuis/werk/etc.) |
| antibioticaresistentie | Tekst | Als er sprake is van antibioticaresistentie, het type resistentie. Mogelijke waarden: ‘BRMO, other’, CPE, MRSA, VRE |
| buitenland | Tekst | Indien er sprake is van recente reis naar buitenland: het land in kwestie |
| eersteziektedag | Datum | Datum van eerste ziekteverschijnselen (let op: kan gegokt zijn) |
| overlijden | Datum | Indien de patiënt/cliënt inmiddels overleden is: overlijdensdatum |
| ziekenhuisopname | Datum | Indien de patiënt/cliënt is opgenomen in een ziekenhuis: datum van opname |
| gemeld | Datum | Indien de casus gemeld is aan het RIVM: meldingsdatum OSIRIS |
| statusmelding | Tekst | Status van de melding in OSIRIS. Mogelijke waarden: Definitief, Gefiatteerd, Gewist, Niet meldingsplichtig, Overgedragen |
| medewerker | Tekst | Medewerker die de casus behandelt (meestal een vpk) |
| casemanager | Tekst | Medewerker die de casus overziet (meestal een arts) |
| created | Datum en tijd | Datum waarop deze casus is toegevoegd aan de database (let op: intern, dus wanneer het script gedraaid is) |
| updated | Datum en tijd | Datum waarop deze casus voor het laatst is aangepast in de database (let op: intern, dus wanneer het script gedraaid is) |

|  |  |  |
| --- | --- | --- |
| vw\_situations |  |  |
| Veld | **Type** | **Verklaring** |
| id | Getal |  |
| hpzone\_id | Getal | ID in HPZone, landelijk uniek |
| datum | Datum | Datum waarop de uitbraak heeft plaatsgevonden of is gemeld |
| invoerdatum | Datum | Datum waarop de situation is aangemaakt in HPZone |
| status | Tekst | Status van de uitbraak. Mogelijke waarden: Open, Closed |
| type | Tekst | Toegewezen type van de uitbraak. Mogelijke waarden: Cluster, Exposure, Issue, Outbreak, Super Outbreak, Threat |
| agent | Tekst | Ziekteverwekker |
| agent\_groep | Tekst | Groep waarin de verwekker zich bevindt |
| agent\_meldingsplichtig | 0/1 | Is de groep waaraan deze verwekker gekoppeld is meldingsplichtig? |
| scenario | Tekst | Brede categorie waarin de uitbraak valt, bijvoorbeeld gastroenteritis. Specifiek type is af te leiden uit agent en agent\_groep |
| scenario\_groep | Tekst | Groep waarin dit scenario zich bevindt (momenteel ongebruikt) |
| zekerheid | Tekst | Mogelijke waarden: Laboratory Confirmed, Suspected |
| risiconiveau | Tekst | Risicowaarding op een schaal van 1 t/m 4. Let op: de waardes zijn niet numeriek, maar bevatten de tekst ‘Level X’. |
| artikel26 | 0/1 | Betreft het een artikel 26-melding? |
| context | Tekst | Context waarin deze uitbraak zich afspeelt, bijvoorbeeld thuis of kinderopvang |
| postcode | Getal | Viercijferige postcode, zonder letters |
| gemeentecode | Getal | Code van de gemeente waarin de postcode zich bevindt, volgens CBS gebiedsindeling |
| gemeentenaam | Tekst |  |
| melding | 0/1 | Is de casus gemeld bij het RIVM? |
| medewerker | Tekst | Medewerker die de casus behandelt (meestal een vpk) |
| casemanager | Tekst | Medewerker die de casus overziet (meestal een arts) |
| aantal\_risico | Getal | Aantal individuen dat risico loopt binnen deze casus (let op: niet consistent gebruikt) |
| aantal\_symptomatisch | Getal | Aantal individuen dat symptomen heeft binnen deze casus (let op: niet consistent gebruikt) |
| aantal\_ziekenhuis | Getal | Aantal individuen dat opgenomen is in een ziekenhuis binnen deze casus (let op: niet consistent gebruikt) |
| aantal\_overleden | Getal | Aantal overleden individuen binnen deze casus (let op: niet consistent gebruikt) |
| created | Datum en tijd | Datum waarop deze casus is toegevoegd aan de database (let op: intern, dus wanneer het script gedraaid is) |
| updated | Datum en tijd | Datum waarop deze casus voor het laatst is aangepast in de database (let op: intern, dus wanneer het script gedraaid is) |

|  |  |  |
| --- | --- | --- |
| vw\_enquiries |  |  |
| Veld | **Type** | **Verklaring** |
| id | Getal |  |
| hpzone\_id | Getal | ID in HPZone, landelijk uniek |
| startdatum | Datum | Datum van aanmaken in HPZone |
| einddatum | Datum | Indien gesloten: datum waarop casus is afgesloten |
| ontvanger | Tekst | Medewerker die de (telefonische) vraag heeft ontvangen |
| medewerker | Tekst | Medewerker die de vraag behandelt |
| status | Tekst | Mogelijke waarden: Closed, Open |
| postcode | Getal | Viercijferige postcode, zonder letters |
| gemeentecode | Getal | Code van de gemeente waarin de postcode zich bevindt, volgens CBS gebiedsindeling |
| gemeentenaam | Tekst |  |
| geslacht | Tekst |  |
| typebeller | Tekst | Type individu of organisatie die vraag gesteld heeft |
| categorie | Tekst | Mogelijke waarden: Hygiene/THZ, Infectieziekten, Omgeving en milieu, Overig, Reizigerszorg, SOA/Sense, TBC |
| onderwerp | Tekst | Onderwerp van de vraag. Vertoont overeenkomst met agent of infectie in de andere tabellen, maar vaak andere spelling. |
| onderwerp\_groep | Tekst | Groep waarin het onderwerp zich bevindt |
| onderwerp\_meldingsplichtig | 0/1 | Is de groep waaraan deze verwekker gekoppeld is meldingsplichtig? |
| onderwerpopen | Tekst | Extra informatie over de gestelde vraag, indien aanwezig |
| created | Datum en tijd | Datum waarop deze casus is toegevoegd aan de database (let op: intern, dus wanneer het script gedraaid is) |
| updated | Datum en tijd | Datum waarop deze casus voor het laatst is aangepast in de database (let op: intern, dus wanneer het script gedraaid is) |



Figuur 1: Diagram van een pilotversie van de in dit document omschreven database.