

Ce qu'est réellement une valeur dans DHIS2

Cinq coordonnées, pas un nombre. Élément de données, combinaison d'options de catégories, unité d'organisation, période et ensemble de données — et la différence entre un élément de données et un indicateur, d'où part l'essentiel de la confusion sur les chiffres de routine.

Pierre Robentz Cassion

Leçon 1 sur 8

Données de routine et DHIS2 — Comment la base est structurée

Ce que couvre cette leçon

- Une valeur a cinq coordonnées
- Élément de données — la chose qui est comptée
- Combinaisons de catégories — la ventilation intégrée à l'élément
- Élément de données contre indicateur
- Ensembles de données — le formulaire, et ce que « attendu » veut dire
- Demandez les métadonnées, pas seulement les données
- La suite

Une valeur a cinq coordonnées

Coordonnée	Dans l'extraction	Ce à quoi elle répond
Élément de données	antigen plus la mesure	Ce qui a été compté
Combinaison d'options de catégories	absente	Selon quelle ventilation
Unité d'organisation	facility_id	Où
Période	period	Quand
Ensemble de données	absent	Sur quel formulaire il a été collecté

Élément de données — la chose qui est comptée — En Python

```
import pandas as pd

vax = pd.read_csv("vaccination-coverage-2024.v1.csv", parse_dates=["period"])
print(sorted(vax["antigen"].unique()))
print(vax.groupby("antigen")["doses_administered"].sum())
```

Élément de données — la chose qui est comptée — En R

```
library(dplyr)
library(readr)

vax <- read_csv("vaccination-coverage-2024.v1.csv")

vax |> summarise(doses = sum(doses_administered), .by = antigen)
```

- Demandez lequel des deux — Un dénominateur stocké comme élément annuel peut être révisé rétrospectivement ; stocké. . .

Combinaisons de catégories — la ventilation intégrée à l'élément — Exemple

Data element: Penta3 doses administered

Category combo: Age (<1, 1+) x Sex (F, M)

Stored values: 4 per org unit per period

- **L'élément n'a aucune ventilation**, donc il y a une valeur et rien n'a été perdu.
- **L'export a agrégé les combinaisons d'options**, auquel cas la ventilation existe dans le système et votre fichier ne. . .

Combinaisons de catégories — la ventilation intégrée à l'élément — En Python

```
expected = (vax["facility_id"].nunique() * vax["period"].nunique()  
            * vax["antigen"].nunique())  
print(f"{expected} expected rows, {len(vax)} in the file")
```

```
c(expected = n_distinct(vax$facility_id) * n_distinct(vax$period) *  
          n_distinct(vax$antigen),  
  actual = nrow(vax))
```

- Un **élément de données** est stocké. Quelqu'un l'a saisi.
- Un **indicateur** est calculé. Il a un numérateur, un dénominateur et un facteur, et DHIS2 l'évalue au niveau que vous. . .

Élément de données contre indicateur — Exemple

Indicator: Penta3 coverage
Numerator: #{Penta3 doses administered}
Denominator: #{Surviving infants}
Factor: 100

- Un indicateur n'est pas stocké, donc il ne peut pas être faux dans la base — Si un taux de couverture est faux, c'est. . .
- L'indicateur est évalué au niveau demandé — et l'arithmétique est sommer-puis-diviser, non diviser-puis-moyenner

Élément de données contre indicateur — En Python

```
district = vax[vax["report_submitted"] == True]
correct = (district["doses_administered"].sum()
           / district["target_population"].sum())
wrong = (district["doses_administered"] / district["target_population"]).mean()
print(f"ratio of sums {correct:.3f}, mean of ratios {wrong:.3f}")
```

```
vax |>
  filter(report_submitted) |>
  summarise(ratio_of_sums = sum(doses_administered) / sum(target_population),
            mean_of_ratios = mean(doses_administered / target_population))
```

- Un indicateur peut être demandé pour une période que les données ne soutiennent pas — Demandez la couverture annuelle. . .

Ensembles de données — le formulaire, et ce que « attendu » veut dire — En Python

```
reported = vax["report_submitted"] == True
print(f"{reported.sum()} of {len(vax)} facility-month-antigen values reported")
print(f"reporting rate {reported.mean():.1%}")
```

Ensembles de données — le formulaire, et ce que « attendu » veut dire — En R

```
mean(vax$report_submitted)
```

Ensembles de données — le formulaire, et ce que « attendu » veut dire — En Python

```
partial = (  
    vax[reported].groupby(["facility_id", "period"])["antigen"].nunique()  
)  
print(f"{(partial < 6).sum()} facility-months with only some antigens reported")
```

Ensembles de données — le formulaire, et ce que « attendu » veut dire — En R

```
vax |>  
  filter(report_submitted) |>  
  summarise(antigens = n_distinct(antigen), .by = c(facility_id, period)) |>  
  filter(antigens < 6) |>  
  nrow()
```

Demandez les métadonnées, pas seulement les données

- Les **définitions des éléments de données**, y compris leur mode d'agrégation entre périodes (somme, moyenne, ou...)
- La **combinaison de catégories** de chaque élément, pour savoir ce qui a été réduit.
- L'**affectation des ensembles de données**, pour savoir combien d'unités étaient attendues.
- Les **définitions d'indicateurs** de tout ce qui est calculé, avec les expressions de numérateur et de dénominateur.

- Vous savez ce qu'est une valeur et où elle se situe.

Lire la leçon complète, avec le code exécutable

```
https://data-analysis.cassion.dev/fr/cours/routine-data-and-dhis2/  
dhis2-01-data-elements
```