

What exactly was counted?

Four questions, a provenance block that answers them, and a worked interrogation of one figure from raw value to published percentage. The question this course exists to make answerable.

Pierre Robentz Cassion

Lesson 8 of 8

Routine Data and DHIS2 — Reading it against something else

What this lesson covers

- The question
- Four parts
- The provenance block
- Interrogating a figure you did not produce
- What the system cannot tell you
- Where this course leaves you
- Where module 3 leaves you

The question

- Somebody presents a slide: penta3 coverage, 78%.

Four parts

- What was counted? — The data element, its aggregation operator, and whether any category disaggregation was collapsed. . .
- Where? — Which org units, at which level, on which version of the hierarchy
- When? — Which periods, by activity date or entry date, and how complete they were at extraction
- Out of what? — The denominator, its source, its year, and how many of the expected reporting units contributed

The provenance block — Example (cont.)

Figure	Penta3 coverage, district, 2024
Value	77.5%
What	Data element PENTA3_DOSES ("Penta 3rd dose administered"), aggregation operator SUM over periods and org units. No category disaggregation in the extract; the element has none configured.
Where	38 facilities at org unit level 4, hierarchy as pinned in extracts/2026-07-28/raw/organisationUnits.json. No facility was reassigned during 2024.
When	202401 to 202412, by activity period. Analytics tables last run 2026-07-27 23:14 UTC, so late entry up to that point is included.
Out of what	Surviving infants, MoH catchment estimates 2024, projected from the 2015 census at 2.4% annual growth. Denominator pro-rated to the months each facility reported: 9,238 of a

The provenance block — Example (cont.)

	full-year 11,774.
Completeness	349 of 456 facility-months reported (76.5%). August 29%, September 45%. The figure is coverage among reporting facility-months and is a lower bound on district coverage.
Extract	extracts/2026-07-28/, manifest.json, script extract.py 1.3

The provenance block

- Attach it to the figure, not to the report — A slide with the number and a footnote pointing at the block is what. . .

Interrogating a figure you did not produce

- 1. Is it a data element or an indicator? — If an indicator, ask for the numerator and denominator expressions
- 2. What is the denominator's year? — A 2024 numerator over a 2019 projection is common and understates coverage by. . .
- 3. What was the reporting rate? — Below 90%, the figure is a lower bound
- 4. Which org unit level? — A figure that mixes levels double-counts; a facility-level coverage figure is unreliable for. . .
- 5. Was the period complete when it was pulled? — A recent month is provisional, and a figure quoted from a pull three. . .
- 6. What is it being compared to? — Most misuse is comparison — against a different measure, a different denominator, or. . .

Interrogating a figure you did not produce — In Python

```
def interrogate(figure):  
    return {  
        "type": figure.get("element_or_indicator"),  
        "denominator_year": figure.get("denominator_year"),  
        "reporting_rate": figure.get("reporting_rate"),  
        "org_unit_level": figure.get("level"),  
        "period_complete_at_extraction": figure.get("period_closed"),  
        "comparable_to": figure.get("comparable_to"),  
    }
```

Interrogating a figure you did not produce — In R

```
interrogate <- function(figure) {  
  figure[c("element_or_indicator", "denominator_year", "reporting_rate",  
          "level", "period_closed", "comparable_to")]  
}
```

Interrogating a figure you did not produce

- A blank in any of those is the finding — You are not asking somebody to justify themselves; you are establishing. . .

What the system cannot tell you

- Why a facility stopped reporting — The system records that it did
- What the data entry clerk understood the field to mean — A column labelled "cases" collects whatever a person believes...
- What happened outside the catchment — Routine data sees what came to the service

What the system cannot tell you

The system is a record of what was reported, not of what happened. Every figure it produces inherits that, and stating it once in a report is worth more than any additional decimal place.

Where this course leaves you

- You can read a DHIS2 extract as the database it came from, aggregate up a hierarchy that moves, treat completeness as a denominator, pull the same extract every month with a manifest that says what you asked for, place a routine figure beside a survey estimate and explain the gap, and answer what was counted for any number the system produces.

Where module 3 leaves you

- Three courses.

Read the full lesson, with runnable code

[https://data-analysis.cassion.dev/courses/routine-data-and-dhis2/
dhis2-08-what-was-counted](https://data-analysis.cassion.dev/courses/routine-data-and-dhis2/dhis2-08-what-was-counted)