

# Trainer's Guide to Data Set Design

---

## What is this guide?

This guide is a support document for DHIS2 Academy trainers for the session "Data Set Design Principles" This session follows the standard Academy training approach with

1. a live demo session where the trainer demonstrate and explain the features, and
2. a hands-on session with exercises where participants get to practice the same features.

This guide will help the trainer prepare for the live demo session. The "Live Demo step by step" section has a detailed walkthrough of all the steps to demonstrate with explanations and screenshots that should be easy to follow. Use that when preparing for the live demo session.

There is also a Quick Guide which lists the steps very briefly and this is meant as a lookup guide or "cheatsheet" WHILE doing the demo, to help the trainer remember all the steps and the flow of the demo.

## Learning objectives for this session

1. Identify when to use categories vs. a flat structure
2. Identify when to apply attributes to a dataset
3. Identify the different data set types
4. Describe the impact of outputs on data set design
5. Discuss translation principles
6. Create section based data sets

## Time Requirements

- Live Demo:
- Hands-on Exercises:
- Assignment:

## Background

## Preparations

## Best Practices

Before starting the demonstration, please keep in mind that the most important thing is that the audience is following, so make sure to ask questions to the audience to verify that they are following. If something is unclear, go back and go through it slowly. If you don't have time for all the steps, it is better to cut some steps, than to go fast while nobody understands.

In an online setting, you will be breaking regularly to allow them to perform various ungraded exercises in order to keep them engaged.

In an in-person setting, the participants may be doing the demo with you at the same time. In this scenario, it is ideal if there are other trainer's moving around the room to support participant's as it will be difficult for the

trainer leading the session to answer many individual questions during the demonstrations.

## Quick Guide

### Live Demo step by step

Review the presentation on Data Set Design Principles

[Link](#)

Review the presentation on mapping datasets

[Link](#)

Review the design of the RCH Dataset

Review the [RCH Dataset](#)

This form has many tables and each of them potentially represent a data element category combination (from now on referred to as a catcombo). As such there is no restriction on a dataset to only have one set of dimensions or catcombo, it can have many and as we see above this is necessary as the dimensions are very different from table to table. In the following paragraphs, we will analyse how to break down this form into its component pieces and suggest an implementation pathway in DHIS2.

Note: Take up the ANC table and Delivery table together in plenary, then ask them to review the remaining tables as a group exercise.

#### **ANC table**

This table in the top left corner is one the simpler ones in this form. It has two dimensions, the first column with the ANC activity or service (ANC visits, medication, etc) and the second and third column which represent the place where the service was given with the two options "Fixed" and "Outreach". Since the ANC service is the key phenomena to analyze here, and often there is a need for looking at the total of "ANC 1st visits" etc. no matter where they actually took place, it makes a lot of sense to use this dimension as the data element dimension.

Thus, all items on the first column from "1st ANC" visit to "Albendazole given" are represented as individual data elements. The where dimension is represented as a data element category (from now on referred to as category) with the name "fixed/outreach" with the two data element category options (from now on catoptions) "fixed" and "outreach". There is no other dimension here so we add a new catcombo with the name "Fixed/Outreach" with one category "Fixed/Outreach".

#### **CONCLUSION**

- Data Elements : ANC 1, ANC 2, ANC 3, ANC 4-8, ANC 8+, LLIN given at ANC, Iron Folate given, Albendazole given
- Category options : Fixed, Outreach
- Category : Fixed, Outreach
- Category Combination : Fixed, Outreach

#### **DELIVERY table**

This table is more tricky as it has a lot of information and you can see that not all the rows have the same columns (one field is greyed out/disabled.). If we start by looking at the first column "Deliveries assisted by" that seems to be one dimension, but only down to the "TBA" row, as the remaining three rows are not related to who assisted the delivery at all. Another dimension is the place of delivery, either In PHU or in Community as stated on the top column headings. These deliveries are further split into the outcome of the delivery, whether it is a live or still birth, which seems to be another dimension. So if we disregard the three bottom rows for a moment there seems to be 3 dimensions here, 1) assisted by, 2) place of delivery, and 3) delivery outcome. The key decision to make is what to use as the data element, the main dimension, the total that you will most often use and want easily available in reports and data analysis.

In this case, the outcome dimension as "Total live births" is a very commonly used value in many indicators (maternal mortality ratio, births attended by skilled health personnel etc.). In this case the "Assisted By" dimension could also have been used without any problem, but the added value of easily getting the total live births information was the decisive point for us. This means that from this table (or sub-table of row 1 to 6) there are only two data elements; "Live births" and "Still births".

Next, there are two more dimensions, the "PHU/Community" with its two options and a "Births attended by" with options ("MCH Aides", "Midwives", "CHW", "TBA"). These two categories make up the catcombo "Births" which is assigned to the two data elements "Live births" and "Still births".

Considering the final three rows of the delivery table we can see that "Complicated Deliveries" does not have the assisted by dimension, but has the place and the outcome. "Low birth weight" does not have the assisted by or outcome dimensions. The LLIN given after delivery does not have the place dimension, and is only provided during one possible outcome. Since these three rows do not share catcombos with any other row we decided to represent these fields as so called flat data elements, meaning data elements with no categories at all, and simply adding the additional information from the column headings to the data element name, and therefore ended up with the following data elements with the default (same as none) catcombo; "Complicated deliveries in PHU live birth", "Complicated deliveries in PHU still births", "Complicated deliveries in community live birth", "Complicated deliveries in community still births", "Low birth weight in PHU", "Low birth weight in community", and "LLIN given after delivery".

## **CONCLUSION**

- Data Elements : Live birth, Still birth, Complicated deliveries in PHU live birth, Complicated deliveries in PHU still births, Complicated deliveries in community live birth, Complicated deliveries in community still births, Low birth weight in PHU, Low birth weight in community, and LLIN given after delivery.
- Category Options : PHU, Community; MCH Aides, Midwives, CHW, TBA
- Categories : PHU/Community, Births attended by
- Category combination: Births

## **POST-NATAL CARE table**

This table is simple and we used the same approach as for the ANC table. 3 data elements listed in the first column and then link these to the catcombo called "fixed/outreach". Reusing the same category fixed/outreach for these data elements enables analysis on fixed/outreach together with ANC data and other data using the same category.

## **CONCLUSION**

- Data Elements : PNC 1st contact, PNC 2nd contact, PNC Vitamin A given

- Category options : Fixed, Outreach
- Category : Fixed, Outreach
- Category Combination : Fixed, Outreach

### ***TT table***

This table is somewhat more complex than the previous examples. We decided to use "TT1", "TT2" ... "TT5" as data elements which makes it easy to get the total of each one of these. There is fixed/outreach dimension here, but there is also the "In school place" that is only applied to the Non-Pregnant, or more correctly to any of the two as the school immunisation is done whether the girls are pregnant or not. You may need to consult program staff behind the form to found out how it works in practice; in this case it would be OK to register all school TT immunisations as non-pregnant, which simplifies the model a bit since we can reuse the "TT1" to "TT5" data elements.

So we ended up with a new category called "TT place" with the three options (Fixed, Outreach, In School), and another category called "Pregnant/Non-pregnant" with two options. The new catcombo "TT" is then a combination of these two and applied to the 5 TT data elements.

Since we agreed to put all In Schools immunisations under Non-pregnant in means that the combination of options (Pregnant+In School) will never be used in any data entry form, and hence become a possible unused optioncombo, which is OK. You can choose to grey out fields, and therefore it is not a problem to have such passive or unused catoptions.

Having school as one option in the TT place category simplifies the model and therefore we thought it was worth it. The alternative would be to create 5 more data elements for "TT1 in school" ... "TT5 in school", but then it would be a bit confusing to add these together with the "TT1" ... "TT5" plus TT catcombo. Having school as a place in the TT place category makes it a lot easier to get the total of TT1.. TT5 vaccines given, which are the most important numbers and most often used values for data analysis.

### ***CONCLUSION***

- Data Elements : TT1, TT2, TT3, TT4, TT5
- Category Options : Fixed, Outreach, In School; Pregnant, Non-Pregnant
- Category : TT place; Pregnant/Non-Pregnant
- Category Combination : TT

### ***Complications of early and late pregnancy and labour tables***

We treat these two tables as one, and will explain why. These two tables are a bit confusing and not the best design. The most important data coming out of these tables are the pregnancy complications and the maternal deaths. These data elements contain further detail on the cause of the complication or death (the first column in both tables), as well as a place of death (in PHU or community), and an outcome of the complication (when its not a death) that can be either "Managed at PHU" or " Referred".

We decided to create two data elements for these two tables; "Pregnancy complications", and "Maternal Deaths", and two category combinations, one for each of the data elements.

For the Pregnancy Complications data element there are two additional dimensions, the cause of the complication (the combined list of the first column in the two tables) and the outcome (managed at PHU or Referred), so these are the categories and options that make up that category combination.

For the "Maternal deaths" data element the same category with the different causes are used and then another category for the place of death (in PHU or In community).

This way the two data elements can share one category and it will be easy to derive the total number of pregnancy complications and maternal deaths. While the list of complications on the paper form is divided into two (early and late/labour) you can see that e.g. the malaria in 2nd and 3rd trimester are listed under early, but in fact are for a later phase of the pregnancy. There is no clear divide between early and late complications in the form, and therefore we gave up trying to make this distinction when considering the analysis outputs in the system.

### **CONCLUSION**

- Data Elements : Pregnancy complications, Maternal Deaths
- Category Options : Malaria 1st trimester, Malaria 2nd trimester, Anaemia, Hypertension, Abortion, Early other, Pre-eclampsia, Ante-partum hemorrhage, Obstructed labor, Post-partum sepsis, Retained placenta, Late other; Managed at PHU, Referred; PHU, Community (reused from before)
- Categories : Complication cause; Outcome; Place of death

### ***Family Planning Services table***

This table has 2 dimensions, the family planning method (contraceptive) and whether the client is new or continuing. We ended up with one data element only "Family planning clients" and then added two categories "FP method" with all the contraceptives as options, and another category "FP client type" with new or continuing as options. This way it will be easy to get the total number of family planning clients which is the major value to look at in data analysis, and from there you can easily get the details on method or how many new clients there are.

### **CONCLUSION**

- Data Elements : Family planning clients
- Category Options : Oral contraceptive, Male condoms, Female condoms, IUD, Injectables, Implants, Spermicide; New, Continuing
- Category : Family planning method; Family planning client type
- Category Combination : Family planning

### **Demonstrate how to create a dataset in DHIS2**

In order to create a new dataset, navigate to maintenance -> data set