Chapter 11. Data Elements and Custom Dimensions

This chapter first discusses an important building block of the system: the data element. Second it discusses the category model and how it can be used to achieve highly customized meta-data structure for storage of data.

11.1. Data elements

The data element is together with the organisation unit the most important building block of a DHIS 2 database. It represents the what dimension and explains what is being collected or analysed. In some contexts this is referred to an indicator, however in DHIS 2 this meta-data element of data collection and analysis is referred to as a data element. The data element often represents a count of some event and its name describes what is being counted, e.g. "BCG doses given" or "Malaria cases". When data is collected, validated, analysed or presented it is the data elements or expressions built with data elements that describe what phenomenon, event or case the data is registered for. Hence the data element become important for all aspects of the system and decide not only how data is collected, but more importantly how the data is represented in the database and how data can be analysed and presented.

An important principle behind designing data elements is to think of data elements as a self-contained description of an phenomenon or event and not as a field in a data entry form. Each data element lives on its own in the database, completely detached and independent from the collection form. It is important to consider that data elements are used directly in reports, charts and other tools for data analysis, in which the context in any given data entry form is not accessible nor relevant. In other words, it must be possible to clearly identify what event a data element represents by only looking at its name. Based on this one can derive a rule of thumb saying that the name of the data element must be able to stand on its own and describe the data value also outside the context of its collection form.

For instance, a data element called “Malaria” might be concise when seen in a data entry form capturing immunization data, in a form capturing vaccination stocks as well as in a form for out-patient data. When viewed in a report, however, outside the context of the data entry form, it is impossible to decide what event this data element represents. If the data element had been called “Malaria cases”, “Malaria stock doses received” or “Malaria doses given” it would have been clear from a user perspective what the report is trying to express. In this case we are dealing with tree different data elements with completely different semantics.

11.2. Categories and custom dimensions

Certain requirements for data capture necessitates a fine-grained breakdown of the dimension describing the event being counted. For instance one would want to collect the number of “Malaria cases” broken down on gender and age groups, such as “female”, “male” and “< 5 years” and “> 5 years”. What characterizes this is that the breakdown is typically repeated for a number of “base” data elements: For instance one would like to reuse this break-down for other data elements such as “TB” and “HIV”. In order to make the meta-data more dynamic, reusable and suitable for analysis it makes sense to define the mentioned diseases as data elements and create a separate model for the breakdown attributes. This can be achieved by using the category model, which is described in the following.

The category model has three main elements which is best described using the above example:

1. The category option, which corresponds to “female”, “male” and “< 5 years” and “> 5 years”.
2. The category, which corresponds to “gender” and “age group”.
3. The category combination, which should in the above example be named “gender and age group” and be assigned both categories mentioned above.

This category model is in fact self-standing but is in DHIS 2 loosely coupled to the data element. Loosely coupled in this regard means that there is an association between data element and category combination, but this association may be changed at any time without loosing any data. It is however not recommended to change this often since it makes the database less valuable in general since it reduces the continuity of the data. Note that there is no hard limit on the

number of category options in a category or number of categories in a category combination, however there is a natural limit to where the structure becomes messy and unwieldy.

A pair of data element and category combination can now be used to represent any level of breakdown. It is important to understand that what is actually happening is that a number of custom dimensions are assigned to the data. Just like the data element represents a mandatory dimension to the data values, the categories adds custom dimensions to it. In the above example we can now through the DHIS 2 output tools perform analysis based on both “gender” and “age group” for those data elements, in the same way as one can perform analysis based on data elements, organisation units and periods.

This category model can be utilized both in data entry form designs and in analysis and tabular reports. For analysis purposes, DHIS 2 will automatically produce sub-totals and totals for each data element associated with a category combination. The rule for this calculation is that all category options should sum up to a meaningful total. The above example shows such a meaningful total since when summarizing “Malaria cases” captured for “female < 5 years”, “male < 5 years”, “female > 5 years” and “male > 5 years” one will get the total number of “Malaria cases”.

For data capture purposes, DHIS 2 can automatically generate tabular data entry forms where the data elements are represented as rows and the category option combinations are represented as columns. This will in many situations lead to compelling forms with a minimal effort. It is necessary to note that this however represents a dilemma these two concerns are sometimes not compatible. For instance one might want to quickly create data entry forms by using categories which does not adhere to rule of a meaningful total. We do however consider this a better alternative than maintaining two independent and separate models for data entry and data analysis.

An important point about the category model is that data values are persisted and associated with a category option combination. This implies that adding or removing categories from a category combination renders these combinations invalid and a low-level database operation much be done to correct it. It is hence recommended to thoughtfully consider which breakdowns are required and to not change them too often.

11.3. Data element groups

Common properties of data elements can be modelled through what is called data element groups. The groups are completely flexible in the sense that they are defined by the user, both their names and their memberships. Groups are useful both for browsing and presenting related data, and can also be used to aggregate values captured for data elements in the group. Groups are loosely coupled to data elements and not tied directly to the data values which means they can be modified and added at any point in time without interfering with the low-level data.