DHIS 2 WHO Digital Health Data Toolkit

Applicable to master version

DHIS 2 Documentation team
September 2020
Table of Contents

1 About the WHO Digital Health Data Toolkit
   1.1 Types of metadata packages

2 General design principles for WHO metadata packages
   2.1 Data analysis
   2.2 Aggregate reporting
   2.3 Cross-cutting principles
   2.4 Tracker

3 Naming conventions
   3.1 General naming rules
      3.1.1 Numbers as part of names
   3.2 Naming for specific object types
      3.2.1 Data Elements
      3.2.2 Indicators
      3.2.3 Favorites
      3.2.4 Option sets
   3.3 Codes
1 About the WHO Digital Health Data Toolkit

The WHO digital health data toolkit for DHIS2 consists of metadata and tools to facilitate the dissemination of health data standards at global scale. The toolkit is an extension of the WHO's toolkit for strengthening routine health information systems. The toolkit brings together WHO standards with UiO’s evidence-based system design principles and implementation lessons to help countries leap-frog into an era of digital information at-their-fingertips. The toolkit supports an ever-growing number of national health programs to digitize and integrate their health data according to WHO standards and best practices.

The digital health data toolkit contains the following resources:

- **Metadata packages**: Pre-configured template of DHIS2 metadata that can be installed on standalone instance or integrated into a country or organization’s existing DHIS2.
- **Installation guides**: Step-by-step guidance for metadata package installation and adaptation to local context.
- **System design guides**: Documentation about the configuration and design decisions to support implementers adapting the package to local contexts.
- **User training guides**: A collection of training templates for end users tailored to each metadata package. Countries can pick up the training guide off-the-shelf and easily customize for their local context.

1.1 Types of metadata packages

Several types of DHIS2 metadata packages exist to support the use of both aggregate and tracker data models in DHIS2:

- **Dashboard & analytics packages** (aggregate) include standard indicators and dashboards that countries can install into their own DHIS2 and map to their existing aggregate metadata.
- **Complete aggregate packages** includes the dashboard and indicators from the dashboard package, but also includes data sets and data elements for routine reporting and data validation rules for data quality. This package is recommended when a country is integrating routine aggregate data collection into their system for a given health area for the first time, or when a country has an opportunity to update the data collection forms to ensure standardized reporting.
- **Tracker packages** include one or more tracker programs for individual-level data collection (individual-level) and related metadata such as program indicators and dashboards. To the extent possible, tracker configurations include the necessary program indicators to populate the data sets of the related aggregate configuration packages.
2 General design principles for WHO metadata packages

2.1 Data analysis

The data analysis part of the configuration is centered around indicators and dashboards. Aggregate packages (both dashboard/analytics and complete versions) include one or more dashboards.

Dashboards in the aggregate packages are based exclusively on indicators. This means that even in cases where an equivalent data element exists for a concept (i.e. ANC visits), an indicator will be created and used in the dashboard item (pivot table, chart or map). This is a requirements for the dashboard-only packages because the indicators are used to map to existing data elements in a country's own DHIS2 instance. One downside to this approach is that we are creating “duplicate” metadata which is not strictly needed. However, there are significant advantages that outweigh this duplication. The most important advantage is that is allows the same dashboards to be used in the aggregate and dashboard-only configuration packages, without modifications. Since the standard configurations packages are aligned with a training curriculum, it is an advantage that any implementation of the dashboard is as similar as possible. Whether the aggregate or dashboard-only package has been installed, the same variables for analysis will be available. In addition, it is often easier for end users to access all of the data they want to analyze within an indicator group in the analytics apps, rather than toggling between data element groups and indicator groups. These analytics end users should not be expected to distinguish between what is a DHIS2 data element vs. a DHIS2 indicator.

Similar to how only indicators are used in the dashboards, only category option groups and category option group sets are used to apply disaggregations to the analytical outputs. That rationale for this is similar to the decision to use only indicators: the category option groups can be mapped to existing category options.

Finally, all analytical outputs (favourites: pivot tables, charts, maps) use relative organisation units and periods. This is necessary to make them portable across instances and over time. In some cases, some modifications are needed to set this according to the context. In those cases, this is described in the documentation.

2.2 Aggregate reporting

The aggregate reporting component of the metadata packages include:

- data sets
- data element
- data element categories, category option, category combinations
- validation rules

The data sets have all been based either on WHO recommendations and best-practice examples or published reporting frameworks (such as the “WHO Definitions and reporting framework for tuberculosis”). These data sets will in many instances have to be adjusted to fit with national reporting systems, to varying degree. On the one hand, there might be additional variables that are important in a national context which must be added. On the other hand, there might be information that is simply not available for reporting, for example if the data is not captured in the case-based registers at the clinical level. The implementation of these reference data sets will therefore often be a longer-term project. Even in contexts where they are not used directly, they
can be used as models for what and how data for different health areas/programs can be collected using DHIS2.

While the standard configuration packages are health area/program specific, the underlying metadata has been harmonized and used across health programs as much as possible. For example, if a data element or disaggregation applies in more than one data set, the DHIS2 metadata has been re-used.

A common gap seen in many country DHIS2 instances is that validation rules are not implemented consistently: they are either not used, or sometimes used to flag data quality issues that are unlikely, but possible. In the standard configuration packages, an effort has been made to add validation rules everywhere it is possible, but only for checks that are certain data quality issues (e.g. tests done vs positive tests).

### 2.3 Cross-cutting principles

Where possible, all metadata objects from data sets and data elements to charts favourites should have a meaningful description. In order to facilitate a harmonized HMIS, aggregate metadata such as data elements and category combinations are shared wherever possible. For example, if the malaria package includes a data element on ‘Number of ANC visits’ as a denominator for an indicator, this data element would be shared with the RMNCAH package in the complete aggregate packages.

### 2.4 Tracker

*Coming soon!*
3 Naming conventions

This document describes overall naming conventions that should be applied, as far as possible, to the all content of the standard configuration packages.

3.1 General naming rules

- The general principle is to design data elements and indicators that can be effectively used for analysis. In general, the amount of end-users interacting with these items will be much higher than system administrators. As a result, making items easier to recognize and find is the general backbone of the design principles suggested.
- Shorten and simplify names/short names where possible, for example, there is in general no need to include “Number of…” and similar phrases, in particular for aggregate data elements where practically everything is numbers.
- Put the key information early in the name, as long as this can be done without making it hard to understand the meaning of the name.
- DHIS2 is an integrated system with information for many different diseases. It is therefore important to be explicit of what health area/program is being referred to. A name like “Case incidence rate per 100 000 population” could refer to malaria, TB, cholera, etc.

3.1.1 Numbers as part of names

- Less than (or equal): use 0–4, not 0 - 4, < 5, <= 4, ≤ 5
- More than (or equal): use 15+, not >14, >= 15, ≥ 15
- Intervals: use 5–14, not 5 - 14

Example

Example of bad naming practice for a data element: *Number of pregnant women who make first ANC visit.*

- *Number of* is generally not necessary. It is understood that it is a number being reported in a monthly report, and this increases the name length.
- Pregnant women is redundant. Only pregnant women make ANC visits.
- First ANC visit is the key information and should not be at the end of the name.
- Ideally, the first part of the name should allow you to group similar data elements in an ordered list.

If we review ANC visits 1 - 4, we could use:

- First ANC visit
- Second ANC visit
- Third ANC visit
- Fourth ANC visit

As plain language this makes sense, however when combined in an ordered list (such as what would appear in DHIS2 analysis applications) with other data elements they will not be either in order or grouped together.
We can change the naming of these data elements to ensure the most important part of the name is what is determining the order of the data elements in an ordered list:

- ANC 1st visit
- ANC 2nd visit
- ANC 3rd visit
- ANC 4th+ visit

This will allow the items to appear together within a thematic area; they are also in order from first to last as is intended.
If it is decided to use a known nomenclature for short forms common in a program (typically for data elements or indicators), select a uniform method across all data metadata (i.e. P., Pl. or Plasmodium).
3.2 Naming for specific object types

3.2.1 Data Elements

- Data element names should be prefixed with the acronym/code of the health program/area, e.g. TB, MAL, HIV, EPI etc.
- Data Elements typically represent a raw count of values therefore it is assumed that, if a data element has no postfix indicating what value type it is, than it is a raw number (i.e. ANC first visit - no postfix is present therefore we can assume that you are reporting the number of ANC visits).
- If you are collecting some type of rate, proportion, etc directly as a non-calculated data value (i.e. as a raw value through a data collection process) than this should be indicated by adding a short postfix in brackets at the end of the data element name.
- We can eliminate plain text when possible from a data element if it is not overtly meaningful when reviewed as an output to shorten the name (i.e. “Number of malaria cases positive with RDT” can be changed to “Malaria cases positive (RDT)” to make a more readable output in analysis.

3.2.2 Indicators

- Data element names should be prefixed with the acronym/code of the health area/program, e.g. TB, MAL, HIV, EPI, etc.
Indicators are calculations that use a combination of data elements to create an output for analysis.

- Naming is often differentiated from data elements by the name of the indicator itself (i.e. ANC 1st visit vs. ANC 1st visit coverage (%))

- These may be slightly longer than data element names due to the addition of a postfix indicating the type of indicator it is.

- Indicators should not start with % at the beginning of their name. In a list, they will be disorganized.

- Long names such as proportion or percentage can be used in descriptions, but should not be included in the indicator name. We can see this adds additional text at the start of the name. It also groups items together by “proportion of/percentage of” rather than the identifying characteristic (i.e. foci, malaria cases, malaria species type).

- If we add % as a postfix items are organized in a more meaningful manner, the name is shortened and unnecessary spaces at the beginning of the name are removed.
3.2.3 Favorites

- Favorite names: Prefix favorite names with health area/program name and colon, e.g. TB: Case notifications..., MAL: Incidence rate...
- Favourite titles: Titles do not need prefixes, as they are not used for searching.

3.2.4 Option sets

Generic/reusable option sets should have generic names as far as possible. For example, if an option set is needed for HIV test results with the options “Positive” and “Negative” it should be named “Positive/Negative” rather than “HIV test result” so that if can be reused. In cases where there is a chance that several objects could be confused or are similar but not the same, it should be made explicit, for example “TB treatment outcome”, “Malaria treatment result”.

3.3 Codes

- All codes should be prefixed with the acronym/code of the disease programme/area and an underscore, e.g. TB_, MAL_, HIV_, EPI_ etc.
- Codes should be capitalized.
- Where codes exists for a particular programme/disease area, those codes should be used. However, they should always prefixed with acronym/code for that area and capitalized.
- If new codes are created, they should follow these guidelines:
  - Only alphanumeric characters and underscores should be used.
As far as possible, codes should be meaningful (e.g. “HIV_TEST_POS” rather than “HIV_T01_”).