

Chapter 17. Pivot Tables and the MyDataMart tool

Excel Pivot Table (see screenshot below) is a powerful and dynamic data analysis tool that can be automatically linked to the DHIS 2 data. While most reporting tools in DHIS 2 are limited in how much data they can present at the same time, the pivot tables are designed to give nice overviews with multiple data elements or indicators, and organisation units and periods (see example below). Furthermore, the dynamic features in pivoting and drill-down are very different from static spreadsheets or many web reports, and this makes it a useful tool for information users that want to do more in-depth analysis and to manipulate the views on the data more dynamically. This combined with the well-known charting capabilities of Excel, the Pivot Table tool has made it a popular analysis tool among the more advanced DHIS users for a long time.

Microsoft Excel - dhis2ke_pivots.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
2	Province	Coast														
3	County	(All)														
4	District	(All)														
5	annualized	(All)														
6	year	2010														
7																
8	Sum of IndValue		month													
9	main indicator groups	indicator	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Grand Total	
10	ART services	Enrolled and eligible but not started on ART	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.4	4.5	3.1	26.2	25.2	23.0	7.2
11		HIV+ patients starting ART	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12		New patients enrolled in HIV care	0.7	0.0	0.0	0.0	0.0	0.0	0.0	92.2	11.7	144.2	64.1	168.2	285.9	63.9
13		Patients currently on ARVs	0.7	0.0	0.0	0.0	0.0	0.0	0.0	91.5	11.9	143.2	61.6	164.5	283.4	63.1
14		Patients currently on prophylaxis	0.0	0.0	0.0	0.0	0.0	0.0	0.0	114.2	286.4	47.2	1994.0	1942.8	2195.9	548.4
15		Patients started on ARVs	0.0	0.0	0.0	0.0	0.0	0.0	0.0	26.5	5.2	3.5	70.5	29.2	29.0	13.7
16	Family Planning	All Other Family planning Methods CYP	13.6	8.0	10.0	4.7	27.2	8.4	35.3	36.2	21.5	53.4	40.2	110.5	30.8	
17		BTL Couple year protection	2.7	4.0	5.4	5.6	1.8	2.8	7.2	5.4	5.6	56.2	44.0	41.7	15.2	
18		Condom Couple year protection	0.3	0.1	0.1	0.8	1.3	0.9	1.5	1.1	1.8	3.7	4.6	9.2	2.1	
19		Family Planning New Cases	13.4	11.3	8.8	19.5	21.3	16.6	53.4	48.8	43.0	221.9	236.5	323.3	84.8	
20		Family Planning Revisits	34.0	12.5	14.0	14.8	22.5	15.3	111.6	109.5	68.9	604.7	609.5	752.8	197.5	
21		Ijectables couple year protection	44.9	26.5	26.0	31.0	30.2	25.0	261.8	263.3	133.7	1543.5	1573.0	1891.7	487.5	
22		Implants couple year protection	0.9	17.0	9.1	8.4	10.0	6.6	54.3	26.3	63.6	197.4	304.2	311.6	84.1	
23		IUCD Couple year protection	3.6	4.9	2.6	5.3	6.2	4.3	18.6	12.9	15.8	67.0	55.4	65.2	21.8	
24		Pills Couple year protection	62.2	27.0	23.5	27.5	26.8	25.4	84.1	88.0	78.0	375.2	432.6	463.3	142.8	
25		Vasectomy couple year protection	0.0	1.0	0.9	0.9	0.9	0.0	0.9	0.0	0.0	0.0	0.0	0.0	1.8	0.5
26		WRA receiving FP commodities	1.0	0.6	0.5	0.6	0.6	0.5	3.4	3.3	2.2	18.0	18.9	20.0	5.9	
27	Immunisation	BCG Coverage	119.9	127.8	143.6	133.4	32.0	1.4	4.3	3.2	3.3	95.3	101.6	73.6	70.0	
28		DPT 1 Coverage	108.8	102.4	110.4	103.1	28.3	0.8	2.5	2.9	3.2	82.4	91.5	58.4	57.9	
29		DPT 2 Coverage	111.5	102.7	94.9	85.0	22.5	1.1	2.7	2.9	4.1	72.5	86.7	53.5	53.3	
30		DPT 3 Coverage	111.6	117.3	102.3	83.8	21.7	0.8	2.5	2.4	3.8	73.6	81.1	50.8	54.3	
31		DPT1 dropout rate	-29.8	-189.9	86.7	227.9	274.0	93.6	5.9	203.0	-245.3	126.3	137.9	152.6	75.4	
32		Fully immunized Child Coverage	123.6	123.6	120.4	95.9	22.9	0.8	1.3	1.0	3.0	69.5	71.0	64.5	58.2	
33		Measles coverage	133.5	132.1	126.4	98.6	24.8	0.8	1.6	0.8	3.0	71.6	72.2	68.8	61.2	
34		Measles dropout rate	-266.8	-378.1	-169.9	53.3	147.6	-18.7	418.0	837.4	58.9	154.6	256.7	-210.4	-69.7	
35		OPV 1 Coverage	52.4	138.7	104.8	107.8	23.2	0.8	2.6	2.8	3.1	81.7	93.2	75.8	57.4	
36		OPV 2 Coverage	54.9	109.4	104.9	96.5	22.0	1.0	3.3	2.4	3.1	80.6	88.2	75.4	53.7	
37		OPV 3 Coverage	55.2	115.8	93.5	95.3	19.9	0.7	2.6	2.3	3.9	83.5	85.1	67.4	52.2	
38	Malaria	Malaria confirmed cases ratio	20.2	18.0	17.9	21.7	26.7			39.3	23.4	34.5	27.8	33.4	26.9	20.9
39		Malaria confirmed incidence rate	102.9	89.4	75.1	67.8	42.9	0.0	1.7	0.8	1.1	16.6	13.9	24.0	36.2	
40		Malaria incidence rate	406.0	406.9	344.5	244.5	117.5	0.0	2.7	2.7	2.0	43.1	27.7	65.0	137.9	
41		Malaria inpatient deaths	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	

With the recent shift towards online deployments, the offline pivot tables in Excel also provide a useful alternative to the online reporting tools as they allow for local data analysis without Internet connectivity, which can be an advantage on unstable or expensive connections. Internet is only needed to download new data from the online server, and as soon as the data exists locally, working with the pivot tables require no connectivity. The MyDatamart tool, which is explained in detail further down, helps the users to maintain a local data mart file (small database) which is updated over the Internet against the online server, and then used as an offline data source that feeds the pivot tables with data.

17.1. Pivot table design

Typically an Excel pivot table file set up for DHIS 2 will contain multiple worksheets with one pivot table on each sheet. A table can consist of either raw data values (by data elements) or indicator values, and will usually be named based on which level of the organisation unit hierarchy the source data is aggregated by as well as the period type (frequency e.g. Monthly, Yearly) of the data. A standard DHIS 2 pivot table file includes the following pivot tables: District Indicators, District Data Monthly, District Data Yearly, Facility Indicators, Facility Data Monthly, Facility Data Yearly. In addition there might be more specialized tables that focus on specific programs and/or other period types.

One popular feature of pivot tables is to be able to drag-and-drop the various fields between the three positions page/filter, row, and columns, and thereby completely change the data view. These fields can be seen as dimensions to the data values and represent the dimensions in the DHIS data model; organisation unit (one field per level), data elements or indicators, periods, and then a dynamically extended lists of additional dimensions representing organisation unit/

indicator/data element group sets and data element categories (see other chapters of this guide for details). In fact a dynamic pivot table is an excellent tool to represent the many dimensions created in the DHIS 2, and makes it very easy to zoom in or out on each dimension, e.g. look at raw data values by individual age groups or just by its total, or in combination with other dimensions like gender. All the dimensions created in the DHIS 2 will be reflected in the available fields list of each pivot table, and then it is up to the user to select which ones to use.

It is important to understand that the values in the pivot tables are non-editable and all the names and numbers are fetched directly from the DHIS 2 database, which makes it different from a normal spreadsheet. In order to be edited, the contents of a pivot table must be copied to a normal spreadsheet, but this is rarely needed as all the names can be edited in DHIS 2 (and then be reflected in the pivot tables on the next update). The names (captions) on each field however are editable, but not their contents (values).

17.2. Connecting to the DHIS 2 database

Each pivot table has a connection to the DHIS 2 database and makes use of a pivot source view (SQL query) in the database to fetch the data. These queries pull all their data from the data mart tables, so it is important to keep the data mart updated at all times in order to get the most recent data into the pivot tables. A pivot table can connect to a database on the local computer or on a remote server. This makes it well suitable for use in a local network where there is only one shared database and multiple client computers using pivot tables. Excel can also connect to databases running on Linux. The database connection used in the pivot tables is specified in an ODBC data source on the Windows computers running pivot tables.

For online deployments the recommended way to connect to the DHIS 2 data is to make use of the MyDatamart tool, which creates and updates a local data mart file (database) that Excel can connect to. The MyDatamart tool will be described in detail further down.

17.3. Dealing with large amounts of data

The amount of data in a DHIS 2 database can easily go beyond the capabilities of Excel. A table with around 1 million values (rows of data) tend to become less responsive to updates (refresh) and pivoting operations, and on some computers Excel will give out of memory errors when dealing with tables of this size. Typically, the more powerful the computer, the more data can be handled, but the top limit seems to be around 1 million rows even on the high-end computers.

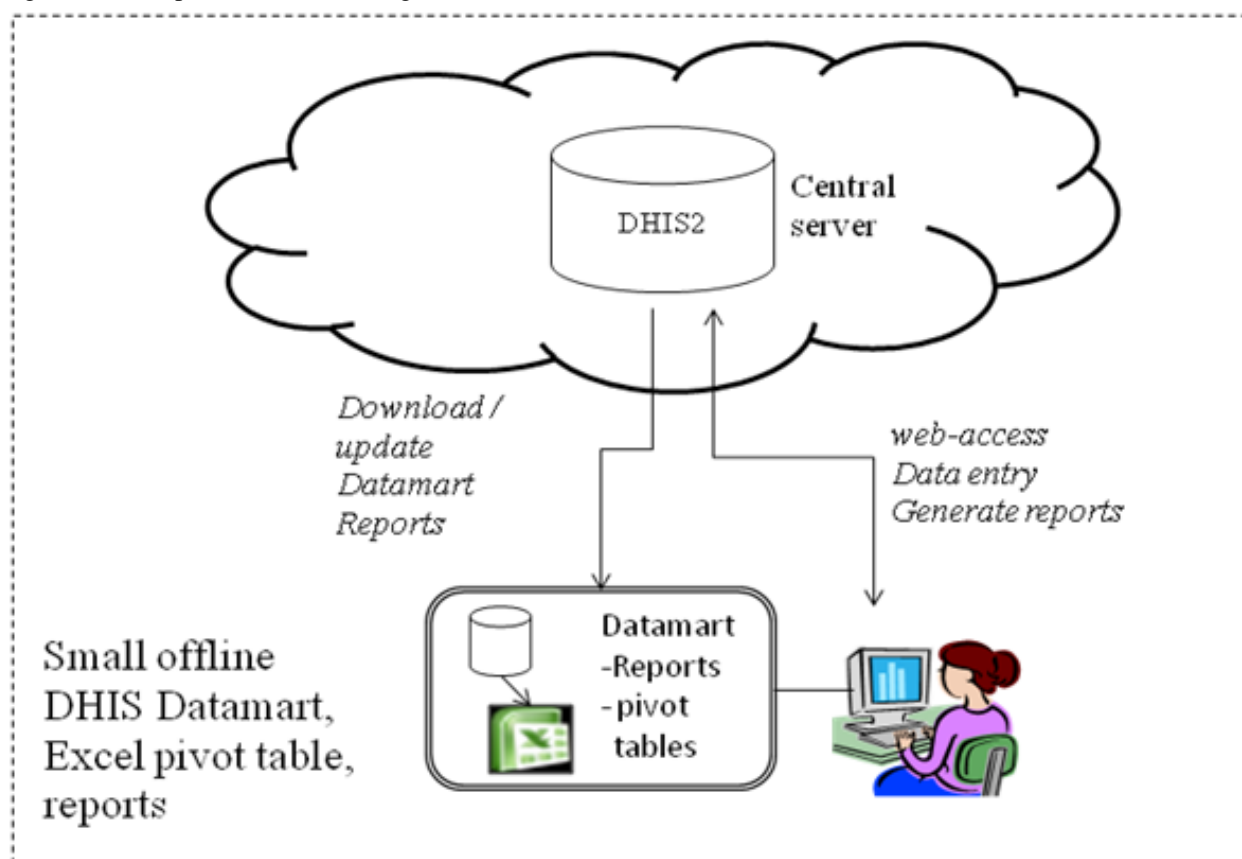
To deal with this problem the standard DHIS pivot table setup is to split the data over several pivot tables. There are different ways of splitting the data; by organisation unit aggregation level (how deep), by organisation unit coverage/boundary area (how wide), by period (e.g. one year of data at a time), or by data element or indicator groups (e.g. by health programs or themes). Aggregating away the lowest level in the organisation unit hierarchy is the most effective approach as it reduces the amount of data by a factor of the number of health facilities in a country. Typically there is no need to look at all the health facilities in a country at the same time, but instead only for a limited area (e.g. district or province). And when there is a need for data for the whole country that can be provided with district level aggregates or similar. At a district or province office the users will typically have facility level data only for their own area, and then for the neighboring areas the data will be aggregated up one or two levels to reduce the size of data, but still allow for comparison, split into e.g. the two tables Facility Data and Data District Data, and similar for indicator values. Splitting data by period or by data element/indicator groups work more or less in the same way, and can be done either in combination with the organisation unit splitting or instead of it. E.g. if a health program wants to analyse a few data elements at facility level for the whole country that can be possible. The splitting is controlled by the pivot views in the database where one specifies which data values to fetch.

17.4. The MyDatamart tool

With online deployments and the use of one single central server (and database) the local use of pivot tables becomes more difficult as Excel connects to the database directly to fetch the data. This means that Excel (and every local

computer using DHIS2) would need connection details and access to the database on the server, which is not always wanted. Furthermore, the refresh (update the pivot table) operation in Excel completely empties the table before reloading all the data again (new and old), which leads to big and duplicated data downloads over the Internet when connecting to an online server. The solution to these problems has been to build up and maintain an updated "copy" of the central database in each local office where they use Excel pivot tables. These local databases are called data marts and are built specifically for serving as data sources for data analysis tools like Excel. The MyDatamart tool is a newly developed (May 2011) tool that creates a datamart file on a local computer and helps the users to update this against a central server. The pivot tables in Excel connect only to the local datamart and do not need to know about the central server at all.

The use of MyDatamart dramatically reduces the download size when routinely updating the local Excel files against the central server compared to a direct connection from Excel. It also brings comfort to the local level users to have a copy of their data on their local computer and not to rely on a Internet connection or server up-time to access it. The figure below explains how the linking between the central online server (in the clouds) and the local offices works.



17.5. Using Excel pivot tables and MyDatamart - a work-flow example

The details of using the MyDatamart tool are explained in a separate user manual and this section only tries to explain the typical work-flow involved in using the tool together with the pivot tables.

17.5.1. Download and run the MyDatamart tool for the first time

MyDatamart is a small tool that is easy to download and run immediately. Download mydatamart.exe to the Desktop and run it by double-clicking on the file. The first thing you need to do is to create a new datamart file, and then you type in the login details needed to access the central server (url, username, password). The tool will connect to the server (Internet connection needed at this point) and verify your credentials. The next step is to download all the meta-data from the server meaning all the organisation units, data elements, indicators, groups etc. This might take some time depending on your computer's specifications and the speed of the connection, but is a step that is rarely needed after this first download. Once the tool knows the organisation unit hierarchy you can specify which organisation unit you "belong" to and the analysis level you are interested in. These are settings that limit which organisation units you

will download data for. The next thing is to download the data from the server, and then you must specify which periods to download.

17.5.2. Setup and distribute the pivot tables

The first thing needed is to download and install an ODBC driver for SQLite, which is the database server running the local datamart. The database connections in the pivot tables depend on this driver and will fail without installed.

The next thing is to set up the pivot tables themselves. This is a one-off job since the file can be reused as a template in all other locations connecting to the same central database. The MyDatamart tool can produce a skeleton Excel file for you with all the necessary database connections already defined. This will help the process considerably and most of the work is to select which fields to use in each table and give them proper names. The user manual has all the detailed instructions on how to set up a pivot table using the MyDatamart connections.

Once the template Excel file is available it is a matter of distributing it to all local offices that will use pivot tables and to make sure the connections are still valid on the local computers. The connection details in Excel depend on the odbc driver being available and on the name and location of the datamart file. Either you can streamline all local datamart files (by name and location, e.g. "C:\dhis2\dhis2.dmart", or you can use the MyDatamart tool to update the connection details in an existing Excel file to match the location of the local datamart file.

17.5.3. Update MyDatamart

Whenever there is new data available on the central server, e.g. every month, the users will have to open the MyDatamart tool, log on to the server, and then pick the months to download. Once the download has finished the data is available locally in the datamart file.

17.5.4. Update the Pivot tables

Once the local datamart file has been update the users can update the pivot tables by using the Refresh function, once per table. It is important to remember to save the Excel file after refreshing all the tables.

17.5.5. Repeat step 3 and 4 when new data is available on the central server

Whenever there is new data on the server repeat step number 3 and 4 (the two previous steps) in order to update the pivot tables and get access to the latest data.