ICPI Tool Style Guide April 7, 2016

OVERVIEW

Over the last year, PEPFAR has been setting targets and entering into DATIM since early last year. With a wealth of HIV/AIDS programmatic results down to the site level quarterly, DATIM is an extremely useful resource to support data-driven decision making and help build capacity necessary towards achieving epidemic control. DATIM has features, such as its built in pivot table and visualization that helps get a view of the data. This view, however, is just a peak into the data and can be rather limiting.

Agencies, TWGs, and ICPI workstreams need to get more than a glimpse at this data; they need to use this information to evaluate progress and make strategic decisions. As a result, we are seeing an influx in “tools” and “dashboards” to make sense of complex and ample data across numerous operating units and indicators. This is a great step towards data-driven decision making. One drawback is the lack of coordination and, as a result, a lack of uniformity in the creation of these tools.

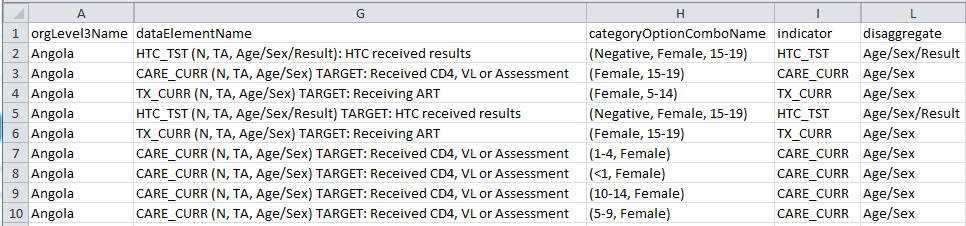
An optimal solution would be to coordinate efforts, designing visual tools and dashboards in a similar fashion. Adopting a uniform structure and overall style has a number of benefits including a similar SQL pull for each product to get large scale data out of DATIM and improved collaboration across ICPI analysts.

Given the widespread use of Excel given accessibility and skillsets, this document outlines style and structure guidance for developing dashboards and tools in Excel.

STRUCTURING RAW DATA

The best place to start at the onset of developing a tool is with the data itself. Whether you pull the data yourself out of the DATIM Data Genie or request the data through a SQL pull from someone in the Data Management Group, the data should be structured in a similar manner where you have your observations by row and indicators as your columns.

Figure 1: Raw Data Structure



When developing a dashboard, there are a number of key indicators you will likely make use of.

* orgLevel\*Name – Each level of the OU hierarchy is assigned a number starting with 3. The national/regional level starts at orgLevel3Name for all OUs. The organization level vary country to country, but as you increase on the numeric scale you identify smaller and smaller sub national units (SNU). It’s a good idea to know at what level each OU is setting target at, i.e. their priority level SNU. This information can be found on the [DATIM Support page](https://datim.zendesk.com/hc/en-us/articles/206855986-GIS-OU-Hierarchy-Guide).
* uidlevel\* - UID is the unique identifier used in spatial mapping, but also is good to use to ensure you are not aggregating two different SNUs in different regions , i.e. a country could have two separate districts called Centerville, one located in the North Province and one in the South Province.
* period – Knowing the time period you are drawing on is a key component of your dataset. It is important to realize that the output from DATIM is reported in calendar year unlike most of the rest of PEPFAR reporting. So, the second quarter of fiscal year 2016 (January to March 2016) would be reported as “2016Q1”. Most of the major indicators are reporting on a quarterly basis, but some are reported only semi-annually or annually. Check out the [MER Indicator Quick Reference Guide on DATIM Support](https://datim.zendesk.com/hc/en-us/articles/206630535-FY-2015-APR-Guidance-and-MER-Indicator-Quick-Guide) for more details. This reference guide will also provide you with guidance on how to report indicators at the end of the year, e.g. aggregating all four quarters or taking the fourth quarter’s results. Lastly, if you are working with targets, they are reported once a year. The period for 2016 targets would be recorded as “2015Oct”
* dataElementName – This field is a bundle that provides all data element pieces concatenated together: indicator, numeratorDenom, type, disaggregate, resultTarget, categoryOptionComboName. It is often much more useful to use each of the data elements pieces rather than this one that combines them all together.

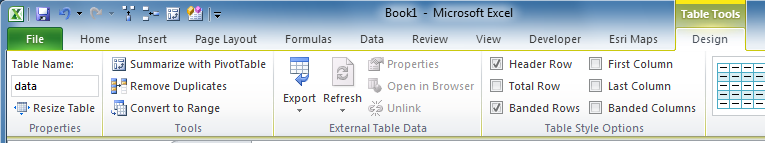
Figure 2: Pieces of the dataElementName



* indicator – This element provide which indicator is reported in the line, e.g. HTC\_TST, PMTCT\_ARV, TX\_CURR.
* numeratorDenom – This element is either “N” for numerator or “D” for demonimator. You will typically be using the numerator for most of your work.
* type – This element is not frequently used as you tend to combine DSD and TA results.
* disaggregate – This element comes in a number of shapes and sizes, varying indicator to indicator. Indicators can be reported as the full numerator or can be broken out into various parts such as by age, sex, result and combinations of those. It is also important to note that these disaggregates can also have aggregates such has “<15” and “15+”
* resultTarget – This element indicates whether the data element is result or target. “TARGET” is written into the data element name if it’s a target; the absence “TARGET” indicates the data element is a result.
* label – This element describes the indicator more clearly rather than just using the abbreviations. This element will be rarely used in typical analysis.
* categoryOptionComboName – This element is tied to the disaggregate. If there is no disaggregate, i.e. it’s just the numerator entered, the categoryOptionComboName would be “default”. If the disaggregate were Age/Sex, the categoryOptionComboName could look like “(Female, 5-14)”.It is important to note that the order of the categoryOptionComboName can vary across and within indicators, so you can see “(Female, 5-14)” or “(1-4, Female)”.
* value – This element provides the target or result value.

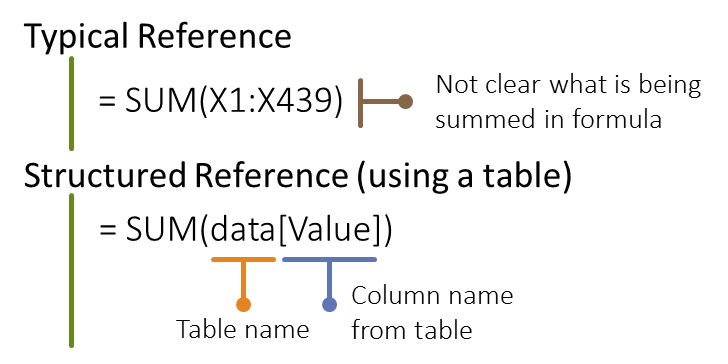
With the relevant indicators loaded into Excel, you should set up the data as a Table. To do so, you will need to highlight all the rows and columns containing data (Ctrl + \*) and navigate on the top ribbon to Insert > Tables > Table Structuring.[[1]](#footnote-1) After setting up the table, you should navigate in the ribbon to Design > Properties > Table Name, and change it to “data”.

Figure 3: Changing Table Name



The benefit of setting up your raw data as a table and renaming it comes from the ability to use structured references. This adjustment is incredibly important because it allows you to more easily reference your data in a formula and improves the readability of your formula for others.

Figure 4: Utilizing a structured reference



Now that you have your data structured using a table, you have the proper base for starting to build your tool.

EXCEL SETUP

The table that you have setup in Excel will form the basis for the rest of your file. You can call this tab “Raw Data.” In addition to the raw data tab, there are a few other useful tabs to include.

The first additional tab to include is the “Info” tab. This tab will include key information about the data being displayed, including any caveats or notes your audience should know about the data. Equally important, you should always include the date and location of your data pull so your audience will know how recent the data is and more easily identify if there are known issues with the dataset. This tab is also a good place to place the ICPI logo if it is not displayed elsewhere and to provide any relevant contact information for the tool developers.

Another good tab to include in the file is a reference tab, which can be called “rs” for Rosetta Stone. The purpose of this tab is to provide (a) a crosswalk between items, e.g. regions and countries, (b) the location of all lists to be used in dropdowns, or (c) any other relevant reference material called upon by the tool or necessary for the developer.

The last necessary tab or tabs to include provide your views of the data. The next section discusses how to use formulas and other key features to aggregate the data and tell a story via visuals or tables.

CALLING AND AGGREGATING DATA

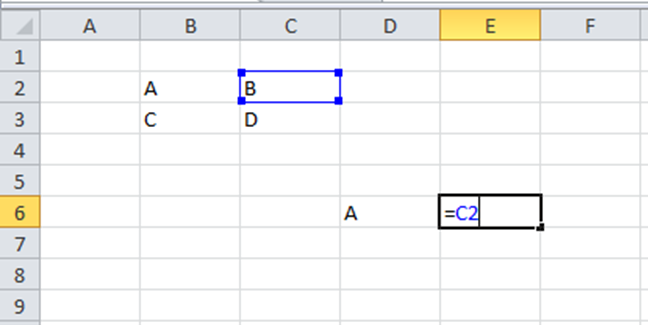
The bulk of the work comes when you have to start creating the tables and visuals for your tool. Working with Excel, there are often multiple ways to go about the process of “calling” or looking up data into a table to get the information you want to display. PivotTables are often one route and can be easy to setup, but can suffer from their bulky structure, their need to be refreshed, and random bugs. Although it can take a bit time more to setup and validate, another method would be to look up the data through formulas. These formulas can be a bit complex, but benefit from their ability to be dynamic and malleable. Setting up your raw data as table will make will decrease some of the complexity and improve readability of your formulas. Below are some key features and functions from Excel which will be the building blocks of your tool.

Features to be Familiar With

Before diving into the formulas, there are some key features to be aware of and use heavily in your tool.

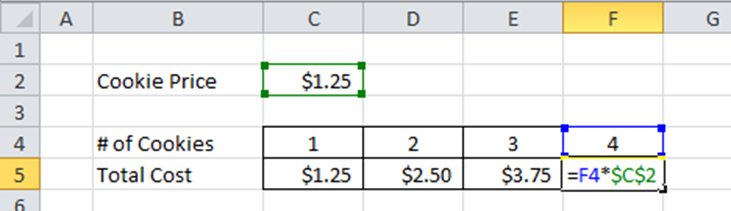
* Data Tables – setting up your data as a data table was referenced Structuring Raw Data section. As mention in that section, converting your data to a data table in excel will allow you to more easily reference your data in formulas and make your formulas more legible. If you are updating the tool on a frequent basis, you can just append new data to the bottom of the table, without the need to update any of your formula references elsewhere in the file. Additionally, setting up a table allows you to easily sort/filter, format, and carry formulas across all rows.
* References – cell references are critical for working with formulas in Excel. Below are four types of references.
  + Relative – relative references are the default in Excel. They are called relative because they move relative to the cell your formula is in. For example, in Figure 5, cell D6 references B2. If you copy this formula over one cell to the right, the reference will move one to the right, referencing C2.

Figure 5: Relative Reference



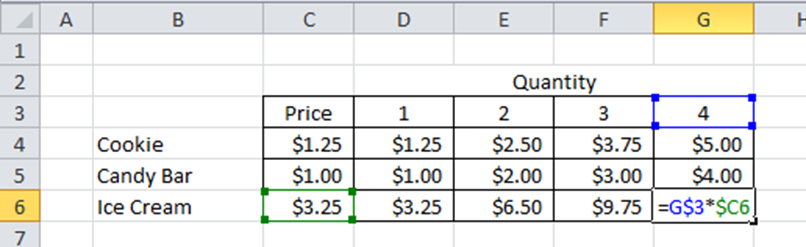
* + Absolute – when using an absolute reference, this essentially “locks” your reference cell regardless of where you move the formula. In Figure 6, the formula in row 5 uses a relative reference (the cell above) for quantity and an absolute reference (C2) for the price. To get an absolute reference, you just need to hit F4 once after you select a cell. The benefits of using an absolute reference are that if the price of a cookie changes, you only need to update it in one place without having to adjust all the formulas to reflect the new price of cookies.

Figure 6: Absolute Reference



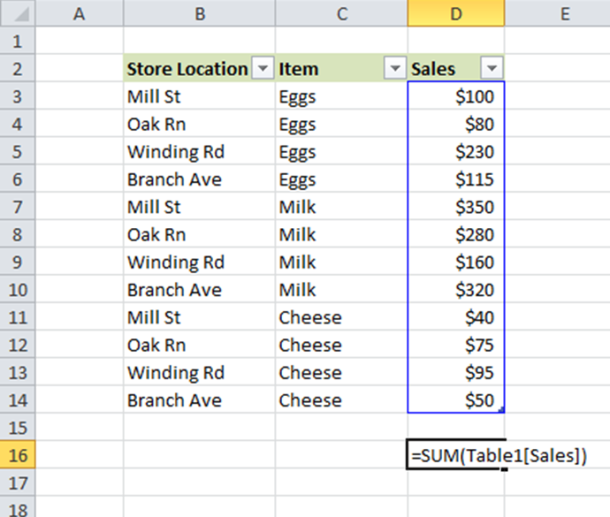
* + Mixed – Often times you want to have a mixed reference, which is a combination of the relative and absolute, allowing you to lock a column or row. To get a mixed reference, hit the F4 key two or three times to lock either the row or column, respectively. In Figure 7, to get the revenue (price x quantity), you can use one formula and copy it over to all the cells in the table. We need to lock in the column for the price (C) and lock the row (3) which contains the quantity. The formula in D4 would be =D$3\*$C4. As you copy the formula over to the other cells in the table, the formula holds the quantity fixed going down a row while adjusting the price and holds the price constant while moving horizontally as it adjusts the quantity.

Figure 7: Mixed Reference



* + Structured – Structured references can only be used when working with tables. They act like mixed references but have the added benefit of using clear labeling in the formulas. For instance, to get the total sales from the table in Figure 8, instead of writing “=SUM(D3:D14)” you can use the structured reference, identifying the table and column heading (in brackets), “=SUM(Table1[Sales])”.

Figure 8: Structured Reference



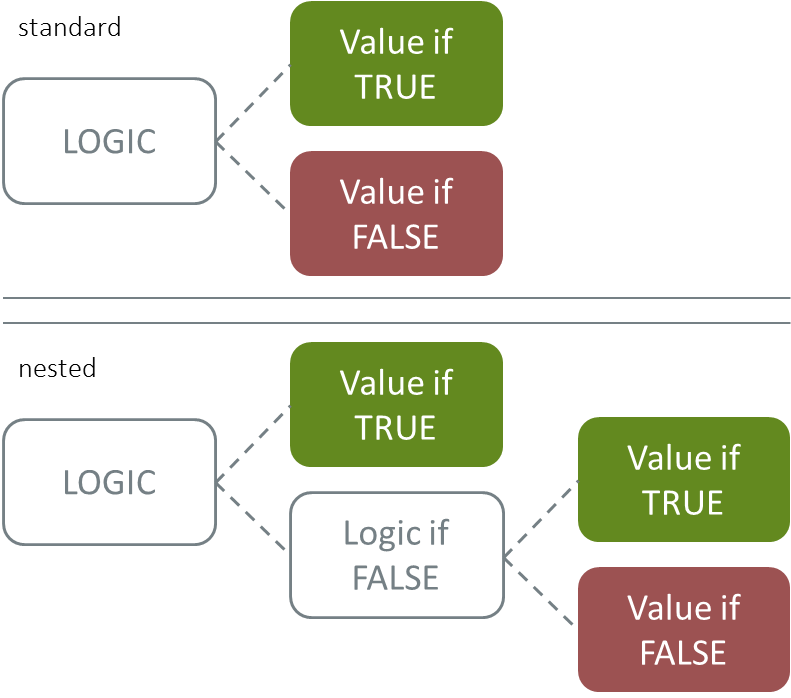
* Named Ranges – A very useful feature in Excel allows you to name ranges of cells. Like with the structured reference, using named ranges makes it more clear what information is being used within a formula. For instance, you have one cell named “ctry\_select” which is the Operating Unit of focus in the table or visual. All formulas can use the “ctry\_select” name and if the OU changes, none of the formulas need to be updated. To create a named range, you start by selecting the cell or group of cells you want to reference. You can add the name by navigating on the ribbon to Formulas > Defined Names > Define Name.
* Text Validation/Dropdown Menus – When designing a tool for use across multiple, it is often useful to have one set of visuals/tables that is dynamic and changes as you adjust the OU of focus. One way to do this is to develop a dropdown selection of OUs for the viewer to choose from. This list of OUs should be stored in the “rs” tab. To create a dropdown, click on the cell where you want to have the dropdown and then navigate on the ribbon to Data > Data Validation. Once there, you need to change the “Allow” box to “List” and add in the named range for your Source.
* Remove Duplicates – When creating named range, like a list of OUs, you often start with all the observations in your dataset and whittle it down to unique observations. One way to do this in Excel is to the use the Remove Duplicated feature. You can pull the full list of countries into your “rs” tab and then, with the list selected, navigate on the ribbon to Data > Data Tools > Remove Duplicates. You will be left with a unique list.

Logic Formulas

Using the following formulas will assist you create more complex formulas by allowing you to subset your dataset or adjust to errors.

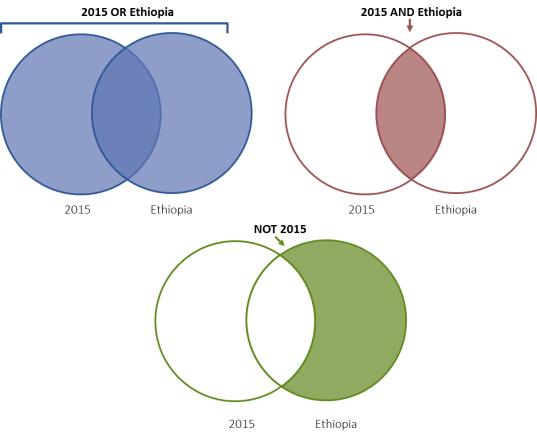
* AND – determines whether multiple inputs are true [AND(logical1, [logical2], ...)]
* OR – determines whether at least one of the inputs is true [OR(logical1, [logical2], ...)]
* NOT – reverses the logic, e.g. NOT(1+1=2) = TRUE [NOT(logical)]
* CONCATENATE/& - combines multiple items together into one cell as text; this can be done via the CONATENATE function or using an “&”. All text, i.e. non-cell references, need to be surrounded by quotes, e.g. = “Date: “ & TODAY() 🡪 Cell would read: “Date: April 7, 2016” [CONCATENATE(text1, [text2],…)]
* IF – These logical formulas become more useful when combined with IF statements. Essentially, you are giving Excel orders based on the output of logic. For example, if the value of a cell is greater than 100%, write “>100%”, otherwise use the value. You can also “nest” IF statements within other IF statements (up to 64 times!). It is often useful to combine the other logic operators (AND, OR, NOT) in the IF statements. [IF(logical\_test, value\_if\_true, [value\_if\_false])]

Figure 9: IF Logic, Standard and Nesting



* IFERROR – tells Excel what the cell should display or value it have if the function produces an error; this is useful where you may not want Excel to display “#N/A” when it experiences an error in the calculation [IFERROR(value, value\_if\_error)]

Figure 5: Visualizing logical expressions – OR, AND, and NOT



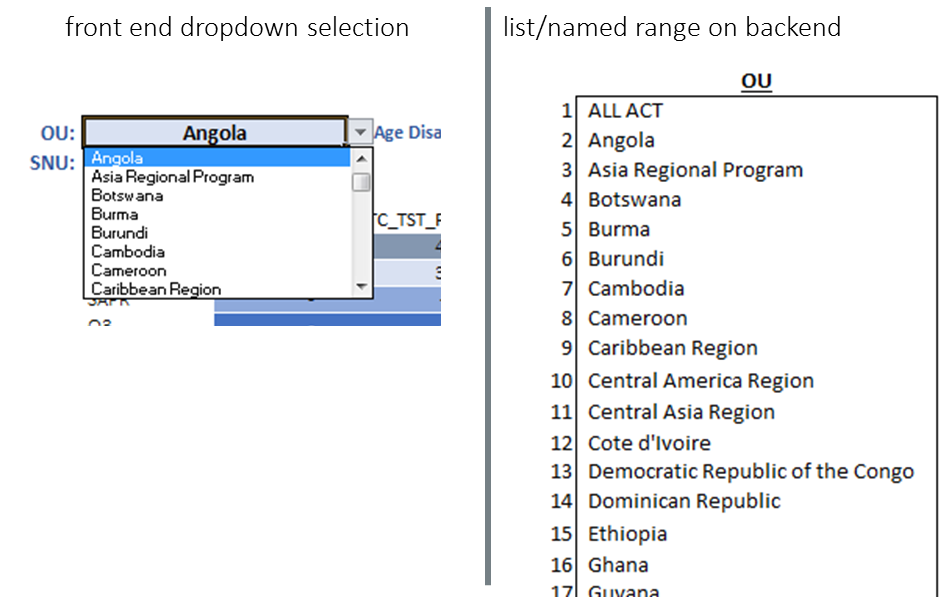
Lookup Formulas

Armed with a better understanding of some of Excels features and logic operators, it is now time to dive into the meat of creating dynamic tools: lookup functions. Below are a number of functions that prove to be quite useful for pulling data from the raw data to aggregate or reference in the table.

* SUMIFS – This
* SUMPRODUCT
* VLOOKUP
* INDEX MATCH

APPENDIX

Dropdown



Dynamic Named Ranges

1. To remove the table formatting, make sure you have selected a cell within the table and navigate on the ribbon to Design > Table Styles > Clear. [↑](#footnote-ref-1)