**Description of the folders**

1. **From IPs**: Contains the reports submitted by IPs. There is one indicator (or HTS\_TST modality) per tab in each file. There is a folder for each month.
2. **Compiled Reports**: Contains the Excel file I use to process the reports.
3. **Datasets**: Contains the datasets to be imported into the Amazon Web Services cloud. One dataset per IP, and one folder per month.
4. **Blank Templates**: Contains the blank templates to send to IPs every month.

**Step-by-step instructions**

Here is how we proceed to build the dataset using the latest Import File contained in the “Compiled Reports” folder:

1. Deleting the data from the previous month, only in the green tabs, except TX\_CURR. For TX\_CURR, we delete the data from the month before the previous one and keep the data from the previous month. It will be used with the data from the current month to generate the TX\_NET\_NEW.
2. Copying the data from the IPs’ reports into the Import File, by making sure the data related to every indicator is put in their respective tab: the green ones.
3. Using the Excel Power Query to transform the data contained in the green tabs from the “large” format to the “long” format. The transformed data goes to the blue tabs for each indicator. We simply refresh the data to run the saved queries.
4. The yellow tabs are used to generate the TX\_NET\_NEW. We refresh the pivot table and make sure the number of rows containing the formulas in the “TX\_NET\_NEW” tab is the same as the number of rows in the pivot table.
5. Combining the contents of all the blue tabs into the red tab and splitting the “Indicator” column to have the disaggregation. Here also, we use the “Refresh all” button to run the saved queries.
6. Copying the contents of the red tab into a blank Excel sheet, by making sure not to accidentally take the TX\_CURR data from the previous month. Then saving the dataset in .CSV-UTF-8 format.
7. We usually split the dataset to have a separate dataset per IP, so that we are not obliged to re-process the entire file when one of the IPs re-submits their report to update or correct the one the submitted previously.
8. We need to make sure the “01-04”, “05-09” and “10-14” age groups are not converted into dates in the datasets.
9. By the 15th of the following month, we update the blank templates by updating the period in all the tabs, and then send the templates to the IPs, for them to submit their reports no later than the 22nd.

So, in a few words, we want the data from the “From IPs” folder to be combined to generate either a combined dataset or one dataset per IP.

The files I am sharing with you contain the datasets for May already created and all the necessary files to create the datasets for June, for the four USAID IPs.