MER Structured Datasets User’s Guide and Data Dictionary

May 22nd, 2019 (Initial FY19 Q2 Release)

Table of Contents

[Overview 3](#_Toc8384016)

[Introduction 3](#_Toc8384017)

[Analytic Dataset Features 3](#_Toc8384018)

[Where to Find the MER Structured Datasets 4](#_Toc8384019)

[Data Sharing Policy 4](#_Toc8384020)

[Additional Information for Analyses 4](#_Toc8384021)

[Guidance for Data Extracts that contain Age Band Columns 4](#_Toc8384022)

[Additional Information about Fine, Semi-Fine, and Coarse Age Band Columns 5](#_Toc8384023)

[Most Complete Age-Sex Disaggregation 6](#_Toc8384024)

[Calculated Indicators 7](#_Toc8384025)

[Standardized Disaggregation 9](#_Toc8384026)

[How to Access Datasets 9](#_Toc8384027)

[Accessing and Downloading from PEPFAR Panorama 9](#_Toc8384028)

[Accessing and Downloading from the Data Genie 10](#_Toc8384029)

[Release notes 12](#_Toc8384030)

[How to Use Excel to Analyze Data from the MER Structured Datasets 13](#_Toc8384031)

[Reducing the File Size to Improve Performance in Excel 14](#_Toc8384032)

[Selecting Indicators, Disaggregates and Modality 14](#_Toc8384033)

[Known Nuances with Frozen Instances 16](#_Toc8384034)

[Special Notes about NAT and SUBNAT data 18](#_Toc8384035)

[Special Notes about IMPATT data 18](#_Toc8384036)

[Appendix A: Standardized Disaggregate Groupings 19](#_Toc8384037)

[Appendix B: MER Structured Data Dictionary 23](#_Toc8384038)

[Appendix C: Critical Site by IM Nuances 32](#_Toc8384039)

[Sensitive Site Names Masked & Key Population Indicators Removed from Site x IM Datasets 32](#_Toc8384040)

[Treatment of Department of Defense (DOD) Data and Implications for All Analyses 32](#_Toc8384041)

[Number of PSNUs Impacted by Rolling-Up DOD Funded Military Sites 33](#_Toc8384042)

# Overview

## Introduction to MER Structured Datasets

The PEPFAR Data Hub (PDH) team and the Interagency Collaborative for Program Improvement (ICPI) Data Access and Quality Subcommittee (DAQ) have developed six datasets known as the Monitoring, Evaluation, and Reporting (MER) Structured Datasets (MSDs). These datasets reduce data access burdens and provide standardized data for analysis across program areas, and they are released twice per quarter: once after the data entry period closes, and once after the data cleaning and deduplication period closes.

One of these datasets contains National (NAT), Subnational (SUBNAT), and Planning and Implementation Attribute (IMPATT) data. The remaining five datasets contain MER targets and results at five different levels:

1. Priority Sub-National Unit (PSNU),
2. Implementing Mechanism (IM),
3. PSNU by IM (both a global file and individual datasets for each OU),
4. PSNU by IM for Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS)-specific operating units, and
5. Individual datasets for each operating unit (OU) at the Site by IM level.

Each of these datasets contains results for FY17 and FY18 and targets for FY17, FY18, and FY19\*; results and targets for previous years (dating back to FY15) will be posted in a separate historical (or archive) file.[[1]](#footnote-1)

## MER Structured Dataset Features

The datasets include the following features:

* Different columns corresponding to the PEPFAR organizational hierarchy, including Region, Operating Unit, Country, Prioritization-level SNU (PSNU; as applicable), Community (as applicable), and Facility (as applicable).
* A column listing the specific indicator name (e.g., HTS\_TST) separate from other pieces of metadata to allow for easy display and filtering of data in analysis.
* PSNU Prioritizations as attributes which allow for easy filtering or grouping by prioritization level (not applicable to the IM dataset).
* Category Option Combo Names, which list all the categories from the DATIM data entry screen, broken out into separate columns for age, sex, human immunodeficiency virus (HIV) status, tuberculosis (TB) status, cervical cancer (CX) status, modality, and other disaggregates.
* [Additional age columns](#_Additional_Fine,_Semi-Fine,) to distinguish Fine, Semi-Fine, and Coarse age disaggregates for easy analysis.
* A filtering column to flag the Military sub-national unit (i.e., \_MIL SNU) for easy identification, removal, or analysis.
* A calculated [Most Complete Age-Sex Disaggregate (MCAD)](#_Most_Complete_Age-Sex_1), which utilizes Fine, Semi-Fine and Coarse age/sex bands for FY2017 and prior, for HTS\_TST, TX\_NEW, and TX\_CURR indicators
* Filter columns to easily distinguish coarse disaggregates from other, finer disaggregates.

## Where to Find the MER Structured Datasets

Five of the MER Structured datasets are posted on [PEPFAR Panorama](https://pepfar-panorama.org/pepfarlanding/#/login) (please see the [*How to Access Datasets*](#_How_to_Access_2) section for additional information). These datasets are a frozen instance of live data entered into DATIM, although daily versions of the datasets are available via the [Data Genie](#_Accessing_and_Downloading)). The PDH team works with the DATIM Database Administrators to update these datasets eight times annually on a schedule that is harmonized with the PEPFAR MER reporting calendar (that calendar, combined with others, is available [here](https://datim.zendesk.com/hc/en-us/articles/115001940503-PEPFAR-Data-Calendar)).[[2]](#footnote-2)

In an effort to provide simple, self-service datasets, many of the MER Structured Datasets summarize MER data in formats that are small enough to open in Excel. However, the global PSNU by IM dataset along with a small number of the OU-specific Site by IM and PSNU by IM datasets exceed the number of rows allowed within Excel. In these cases, datasets must be opened and filtered or transformed within a statistical package, such as R, STATA, or SAS, before being imported into Excel.

## Data Sharing Policy

The MER Structured datasets are only for use within the PEPFAR USG community. When sharing data with external partners and host country governments, we recommend using the publicly available data through [data.pepfar.net](https://data.pepfar.net/). *When sharing site-level data with Implementing Partners (IP), only the IP’s own site-level data should be shared*. If an IP needs additional data in order to have a better understanding of programmatic areas and context outside of that individual IP, use PSNU rather than site-level data.

Datasets may include sensitive key population or military data that should not be shared outside of the PEPFAR USG country teams. *Specifically, the Site by IM level datasets may include information about sites that serve sensitive key populations or the military.* For additional information on how site information for sensitive sites is handled in the MER Structured datasets, please see the [*Sensitive Site Names Masked & Key Population Indicators Removed from Site x IM Datasets*](#_Sensitive_Site_Names_1) section.

# Guidance for Analyses

## Introduction

The type and quantity of PEPFAR’s data has changed over time, and so has the way the program collects data. As a result, users will want to understand certain nuances of the data prior to running any complex analyses. At a minimum, users should be familiar with [different age bands](#_Working_with_Data) and [their relationship with certain indicators](#_Additional_Information_about), the [most-complete age-sex disaggregation](#_Most_Complete_Age-Sex_1), [standardized disaggregation](#_Standardized_Disaggregation_1), and [calculated indicators](#_Calculated_Indicators).

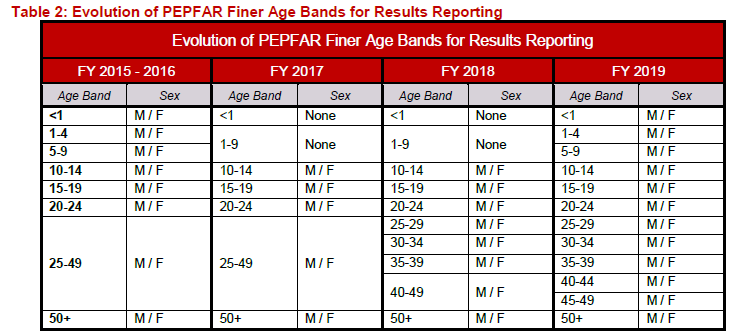
## Age Band Columns

The program’s age bands have become finer over time. To account for this, the MER Structured Datasets (and [*Data Genie*](#_Accessing_and_Downloading) report extracts) contain the columns “TrendsCoarse”, “TrendsSemiFine,” “TrendsFine,” and “AgeAsEntered.” Each column contains a different set age bands corresponding to a specific year or years.

To analyze Age Band data for FY2017, FY2018, and FY2019, use the following schema:

* **FY 2017**: use age bands in “TrendsSemiFine”, including 25-49.
* **FY 2018**: use age bands in “TrendsFine”, including 40-49.
* **FY 2019**: use age bands in “AgeAsEntered” which includes <1, 1-4, 5-9, 40-44, and 45-49.[[3]](#footnote-3)
* **TrendsCoarse** can be used for all coarse analysis for <15, 15+ and Unknown ages.

The table below shows which age bands were collected in each year for most indicators[[4]](#footnote-4):



***TIP #1:*** *When only analyzing data from FY18 at the level of ages <15 or 15+, the “AgeCoarse” column should be used for analysis.*

***TIP #2:*** *When analyzing data prior to FY17, the Most Complete Age Disaggregate (MCAD) is the preferred method of analyzing ages <15 or 15+. The MCAD can be used by filtering the “Disaggregate” column to contain only “MostCompleteAgeDisagg.” Please see the* [*Most Complete Age-Sex Disaggregation*](#_Most_Complete_Age-Sex_1) *section for more in-depth inctructions.*

For additional guidance related to analyses involving age bands, please see [Appendix D: Additional Guidance for Fine, Semi-Fine, and Coarse Age Band Columns](#_Appendix_D:_Additional).

## Age Band Columns for OVC and VMMC

For most indicators, the “AgeFine,” “AgeSemiFine,” and “AgeCoarse” columns contain age bands aligning with the clinical cascade (see “Standard Age Bands” in [Figure 1](#_Figure_1:_Age) below).

However, some indicators use a unique set of age bands for reporting. In these cases, for example for OVC\_SERV, that indicator’s unique age bands are used in the AgeFine, AgeSemiFine, and AgeCoarse columns (see “OVC Age Bands” and “VMMC Age Bands” in [Figure 1](#_Figure_1:_Age) below ).

#### Figure 1: Age Bands



## Most Complete Age-Sex Disaggregation

The Most Complete Age-Sex Disaggregate(MCAD) is a so-called calculated disaggregate used with HTS\_TST, HTS\_TST\_NEG, HTS\_TST\_POS, TX\_NEW, and TX\_CURR. ***While using the MCAD is the preferred method of analyzing data for these indicators when FY 2015-2017 is included in an analysis, MCAD is not necessary when analyzing only data from FY2018***.

PEPFAR developed the MCAD because it was possible for an IM to enter both Semi-Fine and Coarse data at a single site prior to FY18. It was unclear whether such data could be appropriately summed together, or whether the Semi-Fine and Coarse data were duplicative (e.g., a single 18 year old beneficiary might accidently be recorded both in the 15-19 and the 15+ age bands).

PEPFAR applied the MCAD algorithm to its data in order to minimize the risk of over-counting age-disaggregated values.[[5]](#footnote-5) The algorithm takes data from across the different age disaggregates (Semi-Fine or Coarse) and then stores an additional calculated disaggregate value, which is displayed as <15/15+ (Male, Female, or Unknown).

***TIP #1: For analyses using data from FY2017 and earlier:*** *When analyzing age/sex disaggregated data from FY17 or earlier, the Most Complete Age Disaggregate (MCAD) is the preferred method of analyzing age <15 or 15+ data. To use MCAD, users should:*

1. *filter the “Disaggregate” column so that it contains only “MostCompleteAgeDisagg,”*
2. *filter the “indicator” column to select the testing or treatment indicator of interest, and*
3. *in a pivot table, use the “TrendsCoarse” column as a row, column, or filter to display age bands.*

***Note: when using the method above, you should not include the “standardizedDisaggregate” column in your pivot table.***

***TIP #2: For analyses using only FY17 age/sex disaggregated data at the level of <15 or 15+:***

1. *Use the indicator column to select an indicator,*
2. *use either the “disaggregate” or “standardizedDisaggregate” column to filter by “MostCompleteAgeDisagg”, and*
3. *use either the “AgeAsEntered” or “AgeCoarse” columns to display age bands.*

## Standardized Disaggregation

The MSDs (as well as the Data Genie’s reports) contain a “standardizedDisaggregate” column, the value of which is based on the “disaggregate” column. The “standardizedDisaggregate” can:

1. facilitate analysis of the MER 2.0 HTS\_TST indicator
2. indicate which disaggregates are comparable over different reporting periods, which is not necessarily intuitive, since some comparable disaggregates have different names.

The “standardizedDisaggregate” column’s values are based on the “disaggregate” column’s values. Essentially, it clarifies which disaggregates can be meaningfully compared. This allows users to analyze data using a single, standardized disaggregate rather than multiple disaggregates.

Please note, that results reported under the Malnutrition and Pediatric service delivery modalities are included under the Modality/Age Aggregated/Sex/Results standardized disaggregate and the Modality/MostCompleteAgeDisagg standardized disaggregates.

Standardized Disaggregate Examples

Some of the standardized disaggregates are described below. Please see Table 1 in [Appendix A: Standardized Disaggregate Groupings](#_Appendix_A:_Standardized_1) for additional details.

1. **What Periods**: All periods

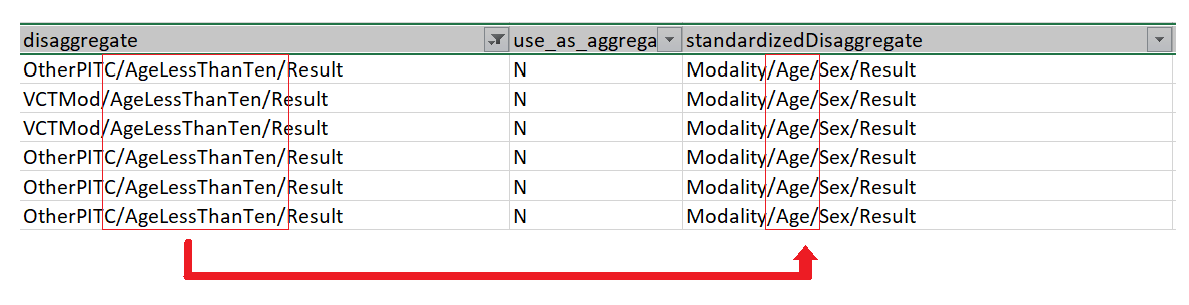
**Standardized Disaggregates**: “Total Numerator” or “Total Denominator”

**Why**: Replacing a blank or implied disaggregates with a Total label to facilitate analyses involving totals.

1. **What Periods:** FY2017

**Standardized Disaggregates**: “Age” replaces the phrase “AgeLessThanTen” or “AgeAboveTen”

**Why**: Facilitates analyses since FY2017 results and FY2018 targets are the only years with disaggregates that contained “AgeLessThanTen” or “AgeAboveTen” (screenshot below).



1. **What Periods:** All period

**Standardized Disaggregates:** The word “Modality” replaces each specific modality of HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG. For example, “Index/Age/Sex/Result” becomes “Modality/Age/Sex/Result”

**Why:** Gives the ability to filter for all data that has modality.Please note that results reported under the Malnutrition and Pediatric service delivery modalities are included under the Modality/Age Aggregated/Sex/Results standardized disaggregate and the Modality/MostCompleteAgeDisagg standardized disaggregates.

1. **What Periods:** starting in FY2019 Q1 and forward

**Standardized Disaggregates:** Adds numerals to HTS\_INDEX disaggregates

1:Age/Sex/IndexCasesOffered

2:Age/Sex/IndexCasesAccepted

3:Age Aggregated/Sex/Contacts

4:Age/Sex/Result

**Why:** Allows Panorama as well as analysts to sort the HTS\_INDEX data into the order needed to show the cascade.

## Calculated Indicators

The PEPFAR Data Warehouse (PDH) creates calculated indicators based on data from DATIM. Calculated indicators are created for select indicators in the PEPFAR Data Warehouse (PDH) for Panorama, Genie and other downstream systems. The calculations automatically create (a) data row(s) based on a specific grouping (e.g., positive or negative or for a specific set of disaggregate). These calculated indicators include:

* CXCA\_SCRN (denominators)
* CXCA\_SCRN\_POS
* CXCA\_TX (denominators)
* GEND\_GBV\_PhysicalEmotionalViolence
* GEND\_GBV\_SexualViolence
* HRH\_CURR (DSD+TA, ClinicalCadre)
* HRH\_CURR (DSD+TA, ClinicalSupportCadre)
* HRH\_CURR (DSD+TA, LayCadre)
* HRH\_CURR (DSD+TA, ManagementCadre)
* HRH\_CURR (DSD+TA, OtherCadre)
* HRH\_CURR (DSD+TA, SocialServiceCadre)
* HTS\_INDEX
* HTS\_INDEX\_KNOWNPOS
* HTS\_INDEX\_NEG
* HTS\_INDEX\_POS
* HTS\_TST\_NEG
* HTS\_TST\_POS
* LAB\_PTCQI CD4 CQI
* LAB\_PTCQI CD4 POTC
* LAB\_PTCQI CD4 POTC CQI
* LAB\_PTCQI CD4 PT
* LAB\_PTCQI HIV IVT/EID CQI
* LAB\_PTCQI HIV IVT/EID POTC
* LAB\_PTCQI HIV IVT/EID POTC CQI
* LAB\_PTCQI HIV IVT/EID PT
* LAB\_PTCQI HIV Serology/Diagnostic Testing CQI
* LAB\_PTCQI HIV Serology/Diagnostic Testing POTC
* LAB\_PTCQI HIV Serology/Diagnostic Testing POTC CQI
* LAB\_PTCQI HIV Serology/Diagnostic Testing PT
* LAB\_PTCQI HIV Viral Load CQI
* LAB\_PTCQI HIV Viral Load POTC
* LAB\_PTCQI HIV Viral Load POTC CQI
* LAB\_PTCQI HIV Viral Load PT
* LAB\_PTCQI Other Testing CQI
* LAB\_PTCQI Other Testing POTC
* LAB\_PTCQI Other Testing POTC CQI
* LAB\_PTCQI Other Testing PT
* LAB\_PTCQI TB AFB CQI
* LAB\_PTCQI TB AFB POTC
* LAB\_PTCQI TB AFB POTC CQI
* LAB\_PTCQI TB AFB PT
* LAB\_PTCQI TB Culture CQI
* LAB\_PTCQI TB Culture PT
* LAB\_PTCQI TB Xpert CQI
* LAB\_PTCQI TB Xpert POTC
* LAB\_PTCQI TB Xpert POTC CQI
* LAB\_PTCQI TB Xpert PT
* OVC\_HIVSTAT\_NEG
* OVC\_HIVSTAT\_POS
* OVC\_SERV\_ACTIVE
* OVC\_SERV\_GRADUATED
* OVC\_SERV\_OVER\_18
* OVC\_SERV\_UNDER\_18
* PMTCT\_ART (denominators)
* PMTCT\_EID\_LESS\_EQUAL\_TWO\_MONTHS
* PMTCT\_EID\_TWO\_TWELVE\_MONTHS
* PMTCT\_STAT\_KNOWNATENTRY\_POSITIVE
* PMTCT\_STAT\_NEWLYIDENTIFIED\_NEGATIVE
* PMTCT\_STAT\_NEWLYIDENTIFIED\_POSITIVE
* PMTCT\_STAT\_POS
* TB\_ART (denominators)
* TB\_STAT\_NEG\_NEWLYIDENTIFIED\_NEGATIVE
* TB\_STAT\_POS
* TB\_STAT\_POS\_KNOWNATENTRY\_POSITIVE
* TB\_STAT\_POS\_NEWLYIDENTIFIED\_POSITIVE
* VMMC\_CIRC\_FollowUp
* MostCompleteAgeDisagg *(MCAD; listed as a disaggregate, not an indicator)*

Users who are interested in analyzing data using calculated indicators listed above can find them in the “Indicator” column of the MSDs or Genie report. One example is calculating testing yield using a combination of calculated indicators in the formula: HTS\_TST\_POS / [HTS\_TST\_NEG + HTS\_TST\_POS].

#### Calculated Total Numerators and Total Denominators

Starting with FY2019 result data, all Totals are calculated rather than provided through DATIM data entry screens. Prior to FY2019, some Totals were entered into DATIM and some were provided through calculation. Unless the MER Indicator Reference Guide (such as [“*MER Indicator Reference Guide (Version 2.3 FY19).pdf*”](https://datim.zendesk.com/hc/en-us/articles/360000084446-MER-2-0-Indicator-Reference-Guide-)) specifies that an indicator should not have a total (i.e. SC\_STOCK), a Total Numerator is provided for each Indicator in PDH for Genie, Panorama and other downstream systems. Some indicators also have denominator data and each of these will have a Total Denominator provided in PDH.

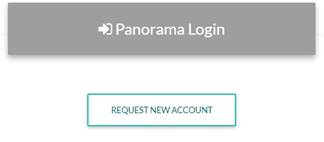
# How to Access Datasets

## Introduction

The MER Structured Datasets can be downloaded from [PEPFAR Panorama](https://pepfar-panorama.org/pepfarlanding/#login) using the instructions below. If users are only interested a certain implementing mechanism or indicator, they may access the same data using the instructions in the *Accessing and Downloading from the Data Genie* section.

## Accessing and Downloading from PEPFAR Panorama

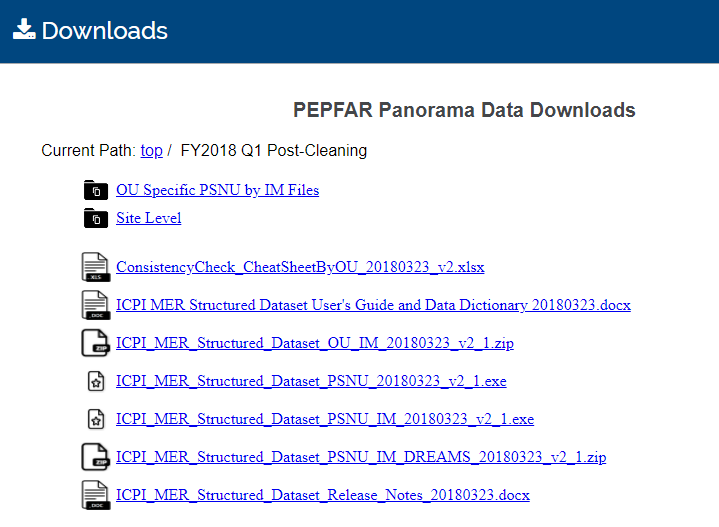
* You will need to log into [PEPFAR Panorama](https://pepfar-panorama.org/pepfarlanding/#login) with a registered e-mail and password.
* Your Panorama username is likely your e-mail address, but you may need to use the "Forgot User ID or Password" option for help if you have an old password.
* If you are a new Panorama user, please navigate to [Panorama](https://pepfar-panorama.org/pepfarlanding/#login) and click the Request New Account button:



To download the MER and SIMS Structured Datasets, log in to [Panorama](https://pepfar-panorama.org/pepfarlanding/#login) and select the green-arrow icon (i.e., “Downloads”) located in the upper-right of the home page (please see screenshot below).



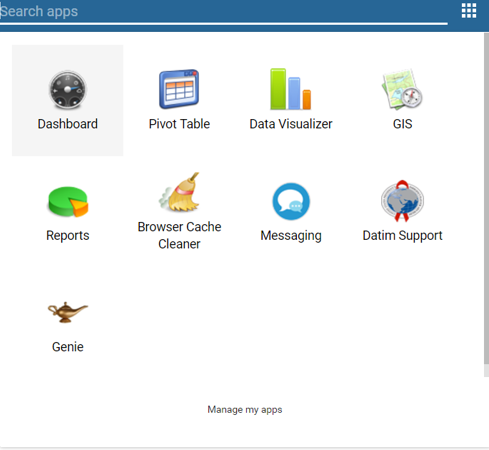
A new window showing PEPFAR Panorama Data Downloads will open. Select either the current MER / SIMS download such as "MER FY2018 Q3 Pre-Cleaning," where QX will have the current quarter. Alternatively, use the links for Previous MER/SIMS Releases to see any of the previous iterations of the Structured Datasets.

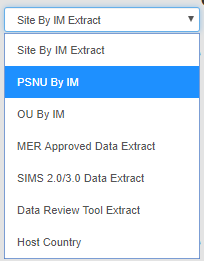


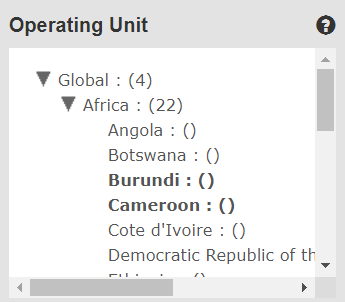
## Accessing and Downloading Data from the Data Genie

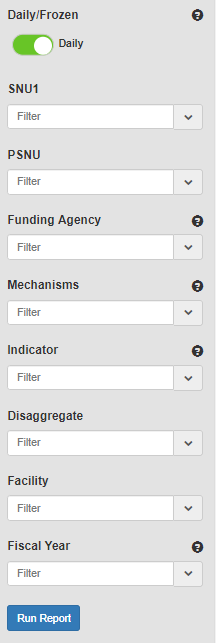
All of the data from the structured datasets may also be downloaded from the PEPFAR Data Genie, a web-based application useful for extracting subsets of data. To access the Data Genie, please refer to the instructions below. For more information, see the [Data Genie User’s Guide and Data Dictionary posted on DATIM support](https://datim.zendesk.com/hc/en-us/articles/360026196051-PEPFAR-Data-Genie-OU-PSNU-and-Site-by-IM-User-s-Guide-and-Data-Dictionary-January-21-2019).

1. Login to <https://datim.org>
2. Select ‘**Genie’** app from the list of available apps.

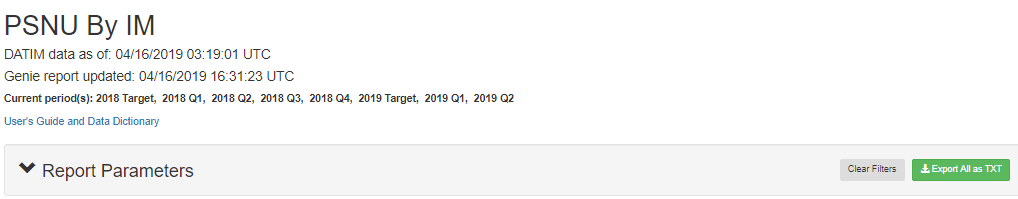
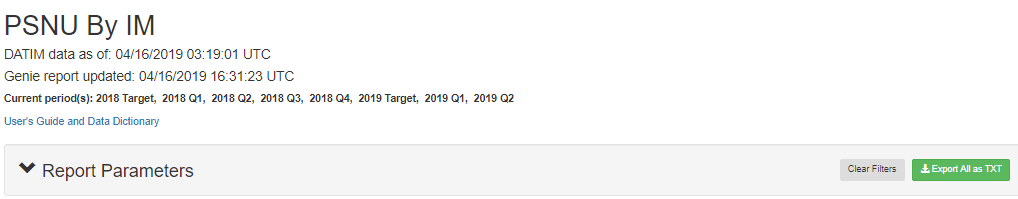




1. Select the desired report in the drop down. The ‘**PSNU by IM Extract**’ option is shown here.
2. Select the Operating Unit(s) of choice from the organizational hierarchy tree. For the ‘OUxIM’ and ‘PSNUxIM’ extracts, you may select multiple OUs. Click any OU to select it, and click again to de-select it. The selection shown would result in an export with data from both Burundi and Cameroon. Please exercise caution in selecting many OUs, as this may result in large datasets that take additional time to run.



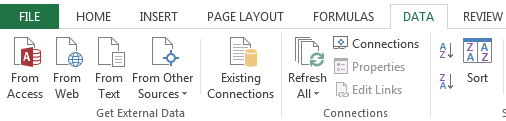
1. Click on the ‘**Run Report**’ option after refining your search results by the following (if applicable): SNU1, PSNU, Funding Agency, Mechanisms, Indicator, and Disaggregate filters. The ‘SitexIM’ extract will have an additional filter for facilities.
2. Users have the option to extract the data returned by selecting the ‘**Export All as TXT**’ option. This will create a tab-delimited output of all the rows returned, regardless of what is displayed on the Genie main table. For instructions on how to load this output into Excel, please see the *How to Use Excel to Analyze Data from the OU by IM, PSNU by IM or Site by IM Datasets section*.



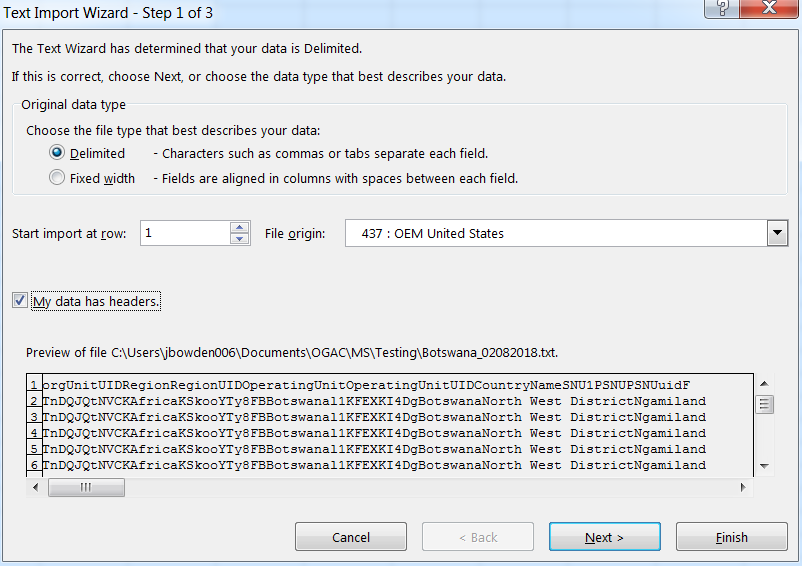
## How to Use Excel to Analyze Data from the MER Structured Datasets or Genie Reports

When using Excel, users must import the data using Excel’s built-in text import functionality to prevent data from importing incorrectly. This can be done by following these steps:

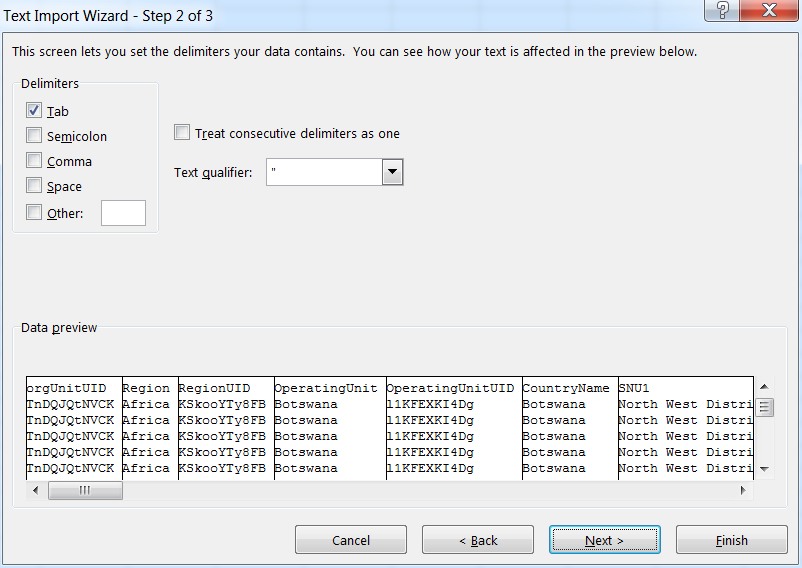
1. Open a blank Excel workbook
2. Select Data in the top ribbon
3. Select “From Text” in the Get External Data section of the “Data” tab



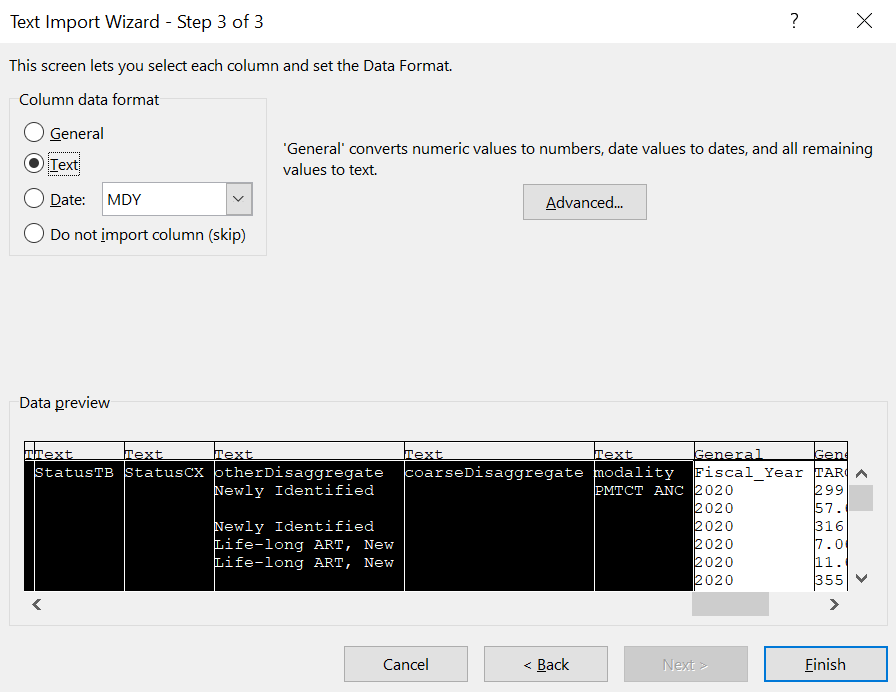
1. In the window that opens, select the file downloaded from Genie that you want to use.
2. In the Text Import Wizard window, select “Delimited” and “My data has headers,” then click Next.



1. In the next window (Step 2 of 3) select “Tab” in the list of available delimiter options and click next.



1. In the final window of the Text Import Wizard (Step 3 of 3), select the first column through the column titled “modality” (do not select any of the columns with result or target values) and click “Text” in the list of Column data format options. ***This will prevent age ranges from being read as dates***.

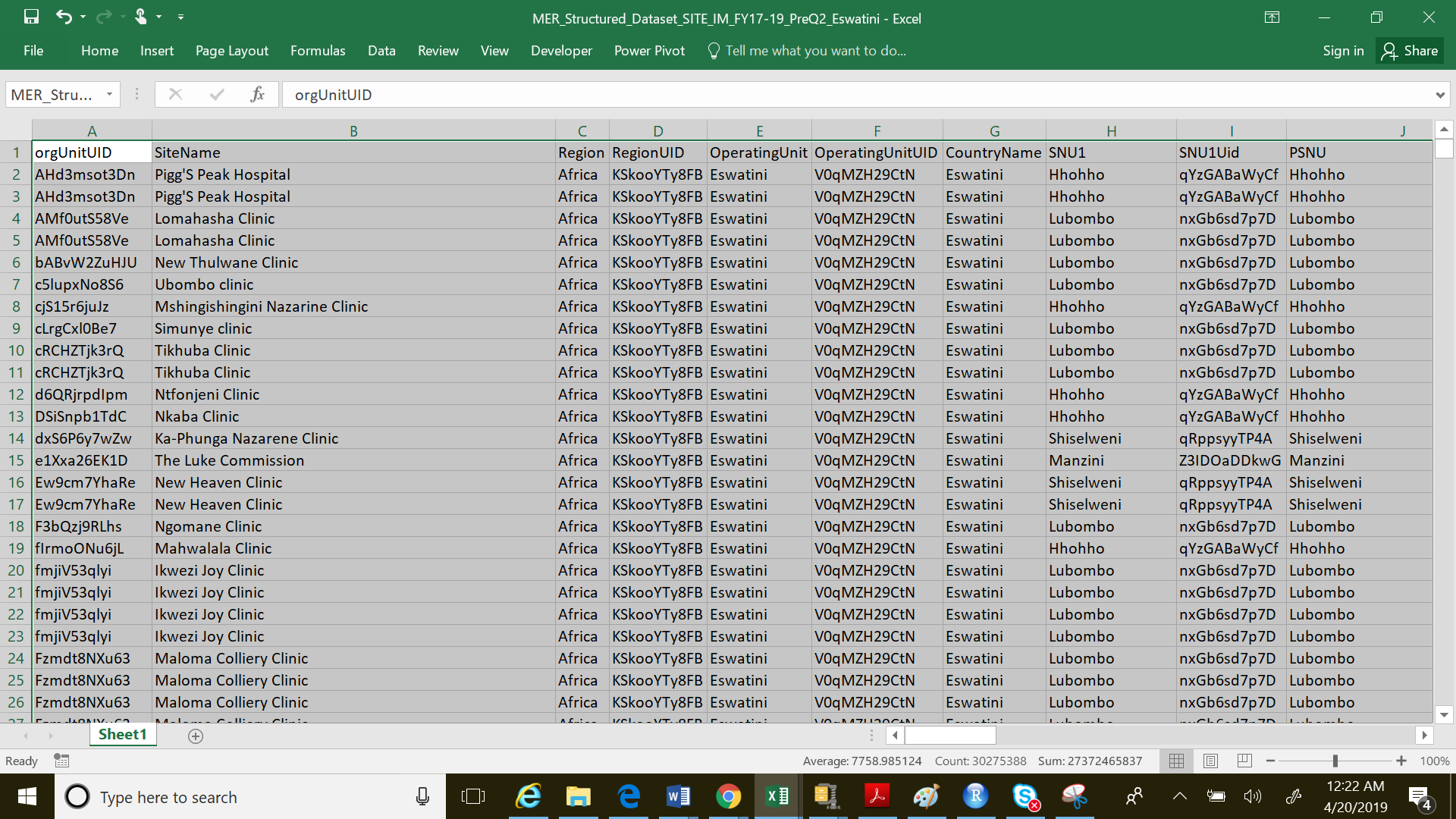


1. Click Finish and your data will import.

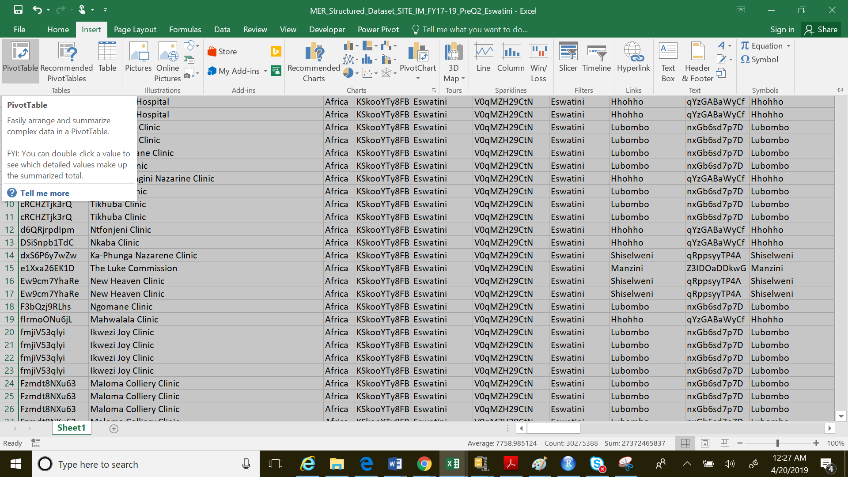
## How to Create Pivot Tables from Genie Extracts

Once users have imported their data, they can use it to build pivot tables, which are useful for summing or counting data across categories. There are many ways to build pivot tables. The process below outlines one possible approach.

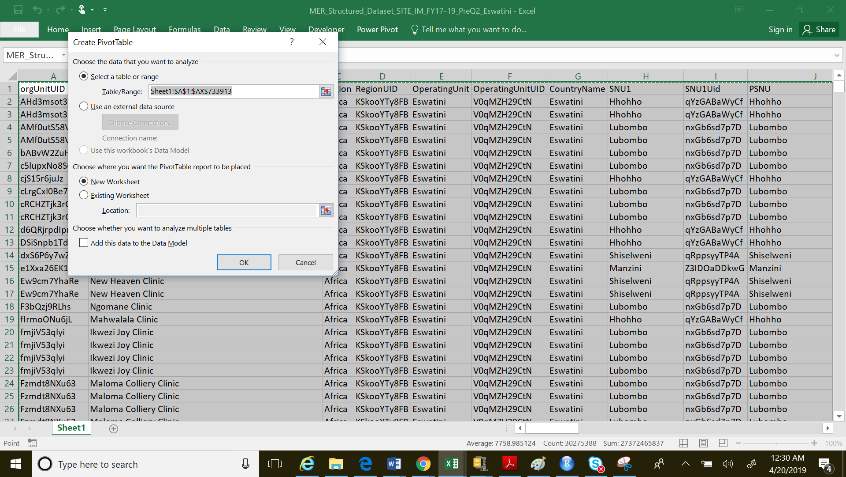
1. Select all of your imported Genie extract data by selecting the first cell with data in it, then holding the **Shift** key, and finally hitting the **Right Arrow** key and the **Down** **Arrow** key. Alternatively, you can click on the square in the upper-left corner of the spreadsheet, to the left of column “A,” and above row “1.”



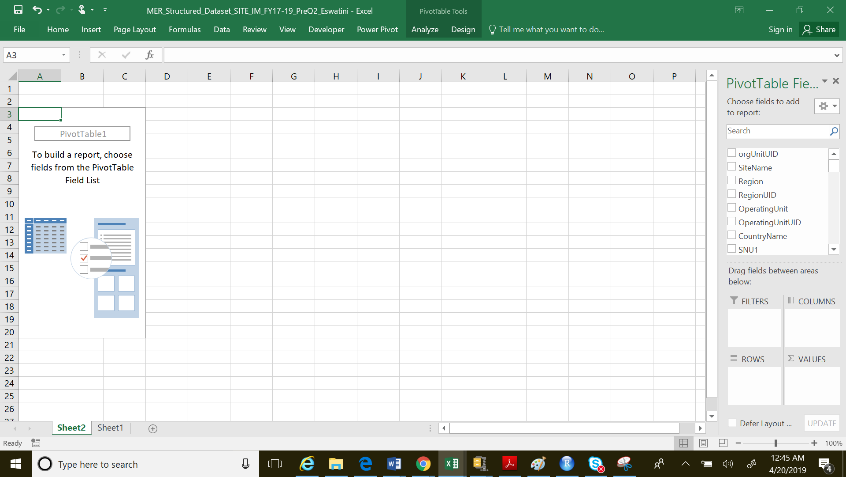
1. In the “Insert” tab, select “Pivot Table.”



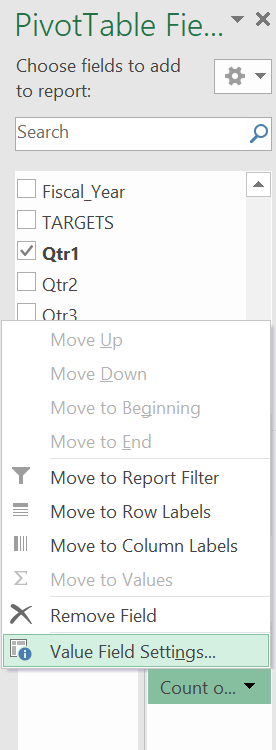
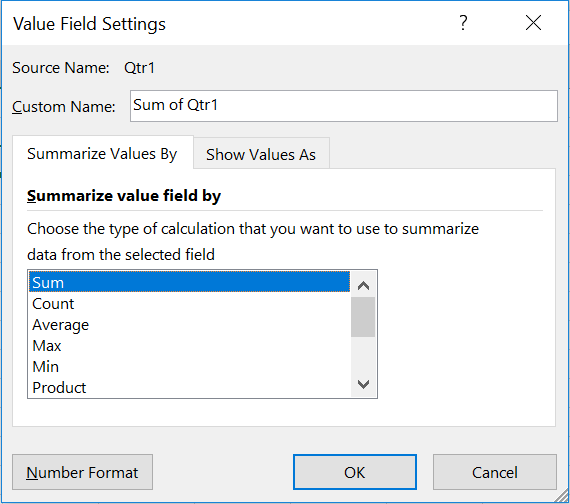
1. Select “Ok” from the window that appears.



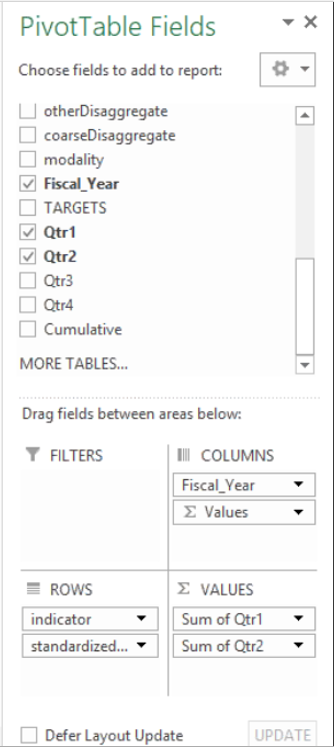
1. You will need to select which PivotTable Fields you are interested in from the list on the right side of the screen (not that they PivotTable Fields are the same as column headers).



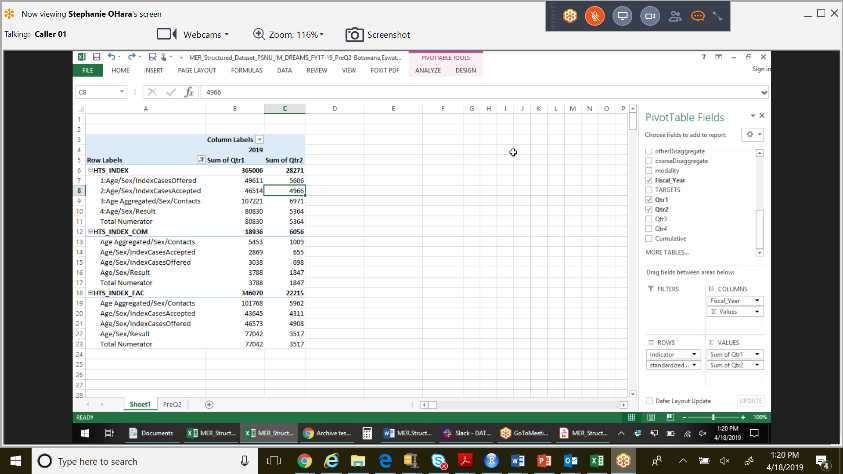
1. One potential pivot table “recipe” is the following:
   1. Drag the “Fiscal\_Year” field from the list to the box labeled “Columns”
   2. Drag the “Sum of Qtr1” and “Sum of Qtr2” columns from the list to box labeled “Values”
   3. Click on “Sum of Qtr1”, select “Value Field Settings,”
   4. Change the value field setting from “Count” (usually the default) to “Sum,” then repeat steps “c” and “d” for “Sum of Qtr2”



* 1. Drag “indicator” and “standardizedDisaggregate” from the list to the box labeled “Rows”



* 1. Your final pivot table will include the summed values of quarters 1 and 2 for whichever fiscal years and indicators you choose.



For additional guidance on how to correctly filter data within the MER Structured Datasets, refer to DAQ webinars/PowerPoints posted to the [training folder](https://www.pepfar.net/OGAC-HQ/icpi/Products/ICPI%20Data%20Store/MER/Training) on pepfar.net or submit a [DATIM Help Desk ticket](https://datim.zendesk.com/hc/en-us/articles/360026196051-PEPFAR-Data-Genie-OU-PSNU-and-Site-by-IM-User-s-Guide-and-Data-Dictionary-January-21-2019).

## Reducing the File Size to Improve Performance in Excel

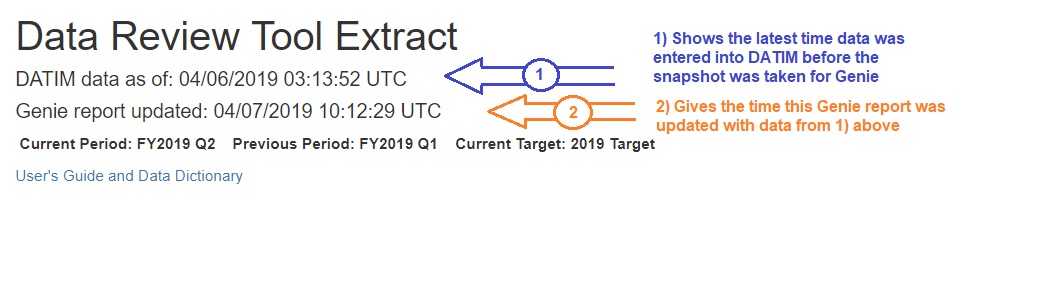
Excel may run slowly on some computers when using large files. If this occurs, it is recommended to reduce the size of your dataset by refining your filter criteria in Genie (i.e. selecting only the specific SNU1, PSNU, indicators, disaggregates, or mechanism that are needed for your analyses).

Filtering can also be done within Excel or an external statistical package (e.g., R, SAS, STATA, SPSS). For those interested in using statistical packages to filter or manipulate large datasets, the ICPI has provided sample code which can be found in the ICPI Data Store on pepfar.net ([link](https://www.pepfar.net/OGAC-HQ/icpi/Products/ICPI%20Data%20Store/MER/Code%20for%20Manipulating%20ICPI%20Fact%20View%20Datasets%20in%20Statistical%20Packages.docx)).

For additional guidance on how to correctly filter data within the analytical documents, refer to the Fact View webinars/PowerPoint slides posted to the [training folder on pepfar.net](https://www.pepfar.net/OGAC-HQ/icpi/Shared%20Documents/ICPI%20Data%20Store/MER/Training). Please contact [DATIM Support](https://datim.zendesk.com/hc/en-us) for any questions or inquiries.

## Data Availability in Genie

The Genie application synchronizes daily to reflect data entered in the DATIM system (www.datim.org). The landing page’s “DATIM data as of” message includes the most recent time stamp for the DATIM data. The “Genie report updated” message includes the time the Genie in question report finished updating; after this time, the selected Genie report will include all data from the most recent DATIM snapshot.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sync from DATIM to Genie | | | | |
|  | Time Zone | DATIM backup | ETL Start | End |
| Daily | US Eastern Time | 22:00PM | 4:00AM | 3:00PM Next Day |
| Bangkok, Thailand | 10:00AM \*11:00AM | 3:00PM | 2:00AM \* 3:00AM Next Day |
| Johannesburg, South Africa | 4:00AM \*5:00 AM | 10:00AM | 9:00PM \*9:59PM Same Day |

The following table details when the synchronization process from DATIM to the Genie begins and ends in several time zones:

*\*Notates local time when the United States is not observing Daylight Savings Time*

## Role-Based Access to Data (Security Trimming)

The data included in different Genie reports depends on 1.) user roles and 2.) implementing mechanism approval levels. Limiting report data in this way is known as security trimming. Note, for Partner Users, only four of the Genie Reports are available for use. These include: The Data Review Tool (DRT), OU by IM, PSNU by IM and Site by IM.

Following an update in FY2018 Q4, data with Approval Levels 1, 2 and 3 is considered “Approved,” while data with Approval Levels 4, 5 and 99 is considered “Unapproved.” The table below outlines the different approval levels that each user role/type of account is able to view. **Yes** means the data is viewable to users with that user role, while **No** means the data is not viewable.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Approval Level** | **Approval Level Description** | **Approved?** | **User Type** | | | |
| **Global User** | **Inter-Agency User** | **Agency User** | **Partner User** |
| 1 | Submitted by Global | Approved | Yes | Yes | Yes\* | Yes\* |
| 2 | Submitted by HQ Agency | Approved  (now used for Expenditure Reporting (ER)) | Yes | Yes | Yes\* | Yes\* |
| 3 | Submitted by Inter-Agency | Approved | Yes | Yes | Yes\* | Yes\* |
| 4 | Submitted by Funding Agency | Unapproved | No | Yes | Yes\* | Yes\* |
| 5 | Submitted by Implementing Partner | Unapproved | No | No | Yes\* | Yes\* |
| 99 | Not submitted to workflow | Unapproved | No | No | Yes\* | Yes\* |

*\*Agency and Partner level users can only view their own data. Regardless of approval level, users with partner level accounts cannot view data from other implementing partners.*

Role-Based Access to Data (Security Trimming) Example

For example, if data is submitted by the Partner for the Agency’s approval (labeled with level 5), then interagency and global users will not be able to view and extract that data for the desired period. The approval level security trimming is also applicable to previous closed periods. Further guidance on the data approval process can be found [here](https://datim.zendesk.com/hc/en-us/articles/204512385-PEPFAR-Data-Approval-Application-User-Guide) on the DATIM support page.

For additional guidance on how to correctly filter data within the MER Structured Datasets, refer to DAQ webinars/PowerPoints posted to the [training folder](https://www.pepfar.net/OGAC-HQ/icpi/Products/ICPI%20Data%20Store/MER/Training) on pepfar.net or submit a [DATIM Help Desk ticket](https://datim.zendesk.com/hc/en-us/articles/360026196051-PEPFAR-Data-Genie-OU-PSNU-and-Site-by-IM-User-s-Guide-and-Data-Dictionary-January-21-2019).

# Appendix A: Standardized Disaggregate Groupings

## Table 1: Standardized Disaggregate Groupings

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | WHEN Conditions are met | | | |
| The Standardized  Disaggregate is | **Indicator is one of** | **Numerator Denominator** | **Disaggregate contains** | **Disaggregate cannot contain** |
| 1:Age/Sex/IndexCasesOffered | HTS\_INDEX | N | Age/Sex/IndexCasesOffered |  |
| 2:Age/Sex/IndexCasesAccepted | HTS\_INDEX | N | Age/Sex/IndexCasesAccepted |  |
| 3:Age Aggregated/Sex/Contacts | HTS\_INDEX | N | Age Aggregated/Sex/Contacts |  |
| 4:Age/Sex/Result | HTS\_INDEX | N | Age/Sex/Result |  |
| Modality/MostCompleteAgeDisagg | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | Result/<5,  MostCompleteAgeDisagg |  |
| Modality/Age/Sex/Result | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | Result/<5 | MostCompleteAgeDisagg |
| Modality/MostCompleteAgeDisagg | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | MostCompleteAgeDisagg |  |
| Modality/Age/Sex/Result | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | AgeAboveTen/Sex/Result  AgeLessThanTen/Result |  |
| Modality/Age Aggregated/Sex/Result | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | Age Aggregated/Sex/Result Age Aggregated/Result |  |
| Modality/Age/Sex/Result | HTS\_TST, HTS\_TST\_POS, HTS\_TST\_NEG | N | Age/Sex/Result Age/Result |  |
| Total Numerator | TX\_CURR, TX\_NEW, TB\_ART,TB\_PREV,TX\_TB | N | HIVStatus |  |
| Total Denominator | TX\_CURR, TX\_NEW, TB\_ART,TB\_PREV,TX\_TB | D | HIVStatus |  |
| Total Numerator | PMTCT\_STAT,VMMC\_CIRC | N | Sex |  |
| Age/Sex | OVC\_SERV,TB\_ART, TX\_RET | N | AgeAboveTen/Sex  AgeLessThanTen |  |
| NewExistingArt/Sex/HIVStatus | PMTCT\_ART | N | NewExistingArt/HIVStatus |  |
| Age/Sex | PMTCT\_STAT,VMMC\_CIRC | N | Age |  |
| Age/Sex/KnownNewResult | PMTCT\_STAT, PMTCT\_STAT\_POS | N | Age/KnownNewResult |  |
| Age/Sex/HIVStatus | TX\_CURR, TX\_NEW | N | Age/Sex,AgeAboveTen/Sex AgeLessThanTen |  |
| Age Aggregated/Sex/HIVStatus | TX\_CURR, TX\_NEW | N | Age Aggregated/Sex |  |
| KeyPop/HIVStatus | TX\_NEW | N | KeyPop |  |
| Age/Sex/Indication/HIVStatus | TX\_PVLS | N | AgeAboveTen/Sex/Indication  AgeLessThanTen/Indication |  |

# Appendix B: Data Dictionary[[6]](#footnote-6)

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Column Description | Data Type (Characters) | Comments |
| orgUnitUID | 11 digit alphanumeric Organization Unit (mixed case string uniquely identifying each Organization Unit) pertaining to the location for which data is reported. | String (12) |  |
| SiteName | Name of the organizational unit pertaining to the location for which data is reported. | String (200) | * The SiteName column uses the following logic:   If SiteType is a facility then Facility Name  If SiteType is a community then Community Name  If SiteType is military PSNU then PSNU Name   * For above-site indicators (like HRH\_CURR), this field is blank |
| Region | World region within which an OU sits. For example: for Uganda, Region is “Africa”; for Vietnam, Region is “Asia”. Also known as *orgLevel2Name.* | String (8) |  |
| RegionUID | 11-digit, alphanumeric, mixed case string uniquely identifying each Region. | String (12) |  |
| OperatingUnit | Name of the Operating Unit pertaining to the location for which data is reported. Also known as *orgLevel3Name.* | String (32) | * For some indicators (e.g., LAB\_PT, LAB\_ACC), the Operating Unit will be the lowest level at which data is reported. |
| OperatingUnitUID | 11-digit, alphanumeric, mixed case string uniquely identifying each Operating Unit. | String (12) |  |
| CountryName | Name of the country pertaining to the location for which data is reported. For Regional Programs, this will differ from Operating Unit (e.g., for Asia Regional Program, this column would include Thailand, Laos, China, and \_Military Asia Regional Program). In all other cases, this column will list the OU. | String (50) | * Helpful in combining data with historic MER/NGI results obtained from FACTS Info. * Caution is advised when using this column in isolation from Operating Unit, especially for \_Military data. Because \_Military data in Regional Programs is collected at a subnational level, selecting any Regional Program country using this column (e.g., China, Barbados, Bahamas) will inherently exclude all \_Military data pertaining to that country and yield an underestimate of actual results or targets for that country. |
| SNU1 | The name of the sub-national unit immediately below the OU pertaining to the location for which data is reported. For \_Military data, the SNU1 will be listed as \_Military [*Operating Unit Name*]. Also known as *orgLevel4Name.* | String (50) | * For Operating Units that prioritize at the SNU1 level, and for all \_Military locations, SNU1 will be identical to PSNU. |
| PSNU | Name of the Prioritization-level SNU pertaining to the location for which data is reported. The level at which prioritization occurs differs from country to country, varying from SNU1 to SNU3. This column correctly lists the name of the Prioritization-level SNU regardless of which level prioritization occurred at. | String (200) | * If data is reported at the \_Military level, the PSNU will be listed as \_Military [*Operating Unit Name*]. * If data is reported at the OU level, the PSNU will be left blank. |
| PSNUuid | 11-digit, alphanumeric, mixed case string uniquely identifying each PSNU. | String (12) | * Important for use in mapping outside of DATIM. |
| SNUPrioritization | The prioritization assigned to the Prioritization-level SNU (PSNU). Options include:   * 1 - Scale-Up: Saturation * 2 - Scale-Up: Aggressive * 3 - NA * 4 - Sustained * 5 - Centrally Supported * 6 - Sustained: Commodities * 7 - Attained * 8 - Not PEPFAR Supported | String (50) | * Prioritizations may change each fiscal year depending on Country and PEPFAR priorities. This column reflects only the prioritization level from the fiscal year of the current reporting period (e.g. during the FY18 Q4 reporting period, this will show FY18 prioritization levels. During the FY19 Q1 reporting period, this will begin to show FY19 prioritization levels). * Many Countries have not reported prioritizations for all SNUs at the level at which prioritization occurs. These have been left blank in this dataset. |
| MechanismUID | 11-digit, alphanumeric, mixed case string uniquely identifying each Implementing Mechanism. | String (12) |  |
| PrimePartner | Name of the organization that is the prime partner for a given mechanism | String(200) | * Caution is advised to use Implementing Mechanism Name instead of Prime Partner as the two are easily confused. A Prime Partner can be responsible for multiple implementing mechanisms across multiple organization units. |
| FundingAgency | Name of the PEPFAR funding agency for the mechanism | String(50) | |  | | --- | | * Dedup * DOD * HHS/CDC * HHS/HRSA * HHS/NIH * PC * State/AF * State/PRM * USAID | |
| MechanismID | Four or five digit, numeric value uniquely identifying each Mechanism. These are common across both FACTS Info and DATIM – as DATIM receives these directly from FACTS Info. | String (50) | * The two de-duplication mechanisms will be denoted as either ‘00000’ (cross-mechanism de-duplication) or ‘00001’ (DSD-TA de-duplication). |
| ImplementingMechanismName | Name of the Implementing Mechanism which collected the data. | String (200) | * While MechanismID and MechanismUID often do not change for a single mechanism across its lifetime, the name of the mechanism may change slightly over time. |
| CommunityUID | 11-digit, alphanumeric, mixed case string uniquely identifying each Community. | String (12) | * Important for use in mapping outside of DATIM. |
| Community | Name of the Community site. | String (200) |  |
| CommunityPrioritization | The prioritization assigned to the Community level. Options include:   * 3 - Scale-Up * 4 - Sustained * 5 - Centrally Supported * 8 - Not PEPFAR Supported | String (50) | * Prioritizations may change each fiscal year depending on Country and PEPFAR priorities. This column provides the prioritization value for the current fiscal year only. |
| FacilityUID | 11-digit, alphanumeric, mixed case string uniquely identifying each Facility. | String (12) | * Important for use in mapping outside of DATIM. |
| Facility | Name of the Facility site. | String (200) |  |
| FacilityPrioritization | The prioritization assigned to the Facility level. Options include:   * 3 - Scale-Up * 4 - Centrally Supported * 5 - Sustained * 8 - Not PEPFAR Supported | String (50) | * Prioritizations may change each fiscal year depending on Country and PEPFAR priorities. This column provides the prioritization value for the current fiscal year only. |
| SiteType | Indicates whether the organizational unit at which the data was reported is a facility, community, or \_Military PSNU. | String(9) | * For above-site indicators (like HRH\_CURR), this field is blank |
| dataElementUID | 11-digit, alphanumeric, mixed case string uniquely identifying each DATIM data element name. | String (11) |  |
| Indicator | The most commonly used method of referring to the specific measurements outlined in PEPFAR MER guidance (e.g., HTC\_TST, TX\_CURR, PMTCT\_STAT) | String (50) | * In the case of calculated indicators where a data element is structured as a disaggregate of a disaggregate (e.g., for KP\_PREV, MSM/TG who are sex workers, which is a subset of MSM/TG) this column will list the full disaggregate (KP\_PREV\_MSMTGSW) as opposed to the actual indicator (KP\_PREV). Other examples of this are: PMTCT\_EID\_POS\_12MO, PMTCT\_EID\_POS\_2MO, GEND\_GBV\_PEP, and all disaggregates for LAB\_PT. These can easily be identified anywhere there is more than 1 underscore (‘\_’) in the indicator name. |
| numeratorDenom | Lists whether a data point pertains to the numerator (‘N’) or denominator (‘D’) of a MER indicator. | String (4) |  |
| indicatorType | Lists whether a data point pertains to Direct Service Delivery (‘DSD’) or Targeted Assistance (‘TA’). | String (4) | * Due to the unique nature of some indicators this column may also list “Not Applicable” in addition to “DSD” or “TA” |
| disaggregate | Lists the type of disaggregate the data point relates to, such as ‘Age/Sex/Result’, ‘Known/New’, ‘ServiceDeliveryPoint’, etc. | String (30) | * This is largely similar to the column by the same name available via DATIM Genie MER Approved Data Extract export, with the exception that when a data point pertains to a total numerator or total denominator value, this column has been enhanced to list ‘Total Numerator’ or ‘Total Denominator’, respectively. This was previously available only by using numeratorDenom where categoryOptionComboName is ‘default’. |
| standardizedDisaggregate | Standardizes disaggregate names – particularly across testing modalities (for HTS\_TST) and across time (when implied disaggregations were applied to some indicators in FY18Q1 within DATIM) | String (30) | * Standardizes the names of comparable disaggregate across time. This column is particularly useful for indicators for which implied disaggregations were added to the DATIM disaggregate name in FY18Q1. * See [Standardized Disaggregation](#_Most_Complete_Age-Sex) and [Appendix A: Standardized Disaggregate Groupings](#_Appendix_A:_) for additional details. |
| categoryOptionComobUID | 11-digit, alphanumeric, mixed case string uniquely identifying each DATIM categoryOptionComboName. | String(11) |  |
| categoryOptionComboName | Lists the specific disaggregation the data point relates to, such as ‘Male, <1, Positive’, ‘Known at Entry Positive’, etc. | String (500) |  |
| AgeAsEntered | Lists the age range of service recipients described in the associated data point (where this is provided). This corresponds with the age group from the data entry screen. It is created as a byproduct of the ‘categoryOptionComboName’ column. | String (25) | * In previous versions of the Site x IM extracts, this column was called “Age”; it was renamed in FY18 Q2 * To allow for easier ordering in visualizations and tables, this column structures ages as two digits, including a leading zero for all ages less than 10. * For age ranges in terms of months rather than years, ages are preceded by ‘[months]’ to distinguish them from year ranges, as well as to allow most ordering algorithms to list these before any year ranges. * VMMC\_CIRC also includes an age disaggregation called “02 months - 09 years” * This column (or the TrendsFine, TrendsSemiFine, or TrendsCoarse columns) should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns, to prevent inadvertent duplication of data across disaggregate groups. |
| TrendsFine | Displays age ranges that correspond to the Fine age bands that were introduced in FY18Q2 | String (25) | * See the [Additional Fine, Semi-Fine, and Coarse Age Band Columns](#_Additional_Fine,_Semi-Fine,) for additional information. * This column (or the AgeAsEntered, TrendsFine, TrendsSemiFine, or TrendsCoarse columns) should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns, to prevent inadvertent duplication of data across disaggregate groups. |
| TrendsSemiFine | Contains age ranges from both the Semi-Fine age bands and the Fine age bands (which are aggregated up to the corresponding Semi-Fine band) | String (25) | * See the [Additional Fine, Semi-Fine, and Coarse Age Band Columns](#_Additional_Fine,_Semi-Fine,) for additional information. * This column (or the AgeAsEntered, AgeFine, or AgeCoarse columns) should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns, to prevent inadvertent duplication of data across disaggregate groups. |
| TrendsCoarse | Contains age ranges from the Coarse age bands as well as from the Semi-Fine and Fine age bands (which are aggregated up to the corresponding Coarse band) | String (25) | * See the [Additional Fine, Semi-Fine, and Coarse Age Band Columns](#_Additional_Fine,_Semi-Fine,) for additional information. * This column (or the AgeAsEntered, AgeFine, or AgeSemiFine, columns) should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns, to prevent inadvertent duplication of data across disaggregate groups. |
| Sex | Lists the sex of service recipients described by the associated data point (where this is provided). This is created as a byproduct of the ‘categoryOptionComboName’ column. Options include ‘Female’, ‘Male’, ‘Unknown Sex’ and blanks. | String (6) | * This column should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns to prevent inadvertent duplication of data across disaggregate groups where sex may be used in both (e.g., Age/Sex Aggregated Result and Age/Sex/Result). |
| StatusHIV | Lists the HIV result status (positive or negative) of an indicator where provided or can be inferred – as is the case for indicators such as TX\_NEW or TX\_CURR. This is created as a byproduct of the ‘categoryOptionCombo Name’ column. Options include ‘Positive’, ‘Negative’, ‘Unknown’, or blanks. | String (8) | * This column should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns to prevent inadvertent duplication of data across disaggregate groups. |
| StatusTB | Lists the TB result status for Tuberculosis Indicators | String(25) |  |
| StatusCX | Lists the Cervical Cancer status for Cervical Cancer (CXCA) Indicators | String(25) |  |
| otherDisaggregate | Any piece of disaggregation not already captured in the ‘Age’, ‘Sex’, ‘StatusHIV’, or ‘modality’ columns is listed here. | String (500) | * This column should usually be used in combination with the ‘disaggregate’ or ‘standardizedDisaggregate’ columns to prevent inadvertent duplication of data across disaggregate groups. |
| coarseDisaggregate | Where a data point pertains to a coarse disaggregate (i.e., data entry forms provided multiple options for entering the same data at varying levels of granularity and users entered data against the coarser of the options), this column will list ‘TRUE’. This is created as a byproduct of the ‘disaggregate’ column. | String (4) | * In some odd cases, coarser age groupings are mixed among finer age groupings (e.g., TX\_CURR Age/Sex) and this column will not denote these as coarse disaggregates. Use caution in using and interpreting results falling in this case. * Use this column in combination with other disaggregate elements to provide context and prevent duplication of values. * When the isMCAD column = Y, the coarseDisaggregate column = FALSE. |
| modality | Lists the HTS modalities by the associated data point (where this is provided). This is created as a byproduct of the ‘disaggregate’ column. Options include: Index, IndexMod, OtherMod, MobileMod, HomeMod, Inpat, TBClinic, Pediatric, Malnutrition, PMTCT ANC, VCT, VCTMod, OtherPITC, VMMC, Emergency Ward, and STI Clinic. KeyPop is also included in this column even though it is technically not a service delivery point/modality and should not be summed along with other modalities (doing so will lead to duplicative results). | String (60) | * Displays the modality for HTS indicators. Use this column to filter data by modality (Service Delivery Points). * KeyPop is not included in the calculation of HTS\_TST\_POS or HTS\_TST\_NEG |
| Fiscal\_Year | Gives the four-digit Fiscal Year for the data | String(4) |  |
| TARGETS | Values for Targets | Decimal (36,2) | Use this with the Fiscal\_Year column to see Targets for each specific year. |
| Qtr1 | Results for October – December periods, entered in DATIM from January-March each year. Only quarterly indicators will have values represented in this column. | Decimal (36,2) |  |
| Qtr2 | Results for January – March periods, entered in DATIM from April-June each year. Only quarterly indicators will have values represented in this column. | Decimal (36,2) |  |
| Qtr3 | Results for April – June periods, entered in DATIM from July-September each year. Only quarterly indicators will have values represented in this column. | Decimal (36,2) |  |
| Qtr4 | Results for July – September periods, entered in DATIM from October-December each year. Only quarterly indicators will have values represented in this column. | Decimal (36,2) |  |
| Cumulative | The Annual Program Review (APR) value according to the MER Indicator guidance for the annual total | Decimal (36,2) |  |
| ApprovalLevel | Numeric value representing the approval level | Decimal (36,2) | * 1, 2 and 3 are “Approved” and 4, 5 and 99 are in process and “Unapproved.” See [Security Trimming](#_Security_Trimming) |
| ApprovalLevelDescription | Description of the numeric approval level values | String (45) |  |

# Appendix C: Critical Site by IM Nuances

## Sensitive Site Names Masked & Key Population Indicators Removed from Site x IM Datasets

The Site by IM datasets include the facility name variable. However, in order to avoid the publication of sensitive site names that provide services to Key Populations and DOD-funded military sites the following steps have been taken:

1. The following indicators have been excluded from the Site x IM Datasets: KP\_PREV and KP\_MAT.

2. Key Population disaggregates for the following indicators have been excluded from the Site x IM Datasets: HTS\_TST, TX\_NEW, PrEP\_NEW.

3. Facility names that have been deemed sensitive because they contain information about military or key populations (e.g., facility names containing key words/phrases like “sex worker,” “narcology,” or “army”) have been masked and replaced with a name generated using a generic naming convention[[7]](#footnote-7) .

User’s should also take note of the data sharing policy outlined in the MER Structured Datasets User’s Guide and Data Dictionary which specifies that datasets that include sensitive key population or military data (especially sensitive site names) should not be shared outside of the PEPFAR USG country teams.

## Treatment of Department of Defense (DOD) Data and Implications for All Analyses

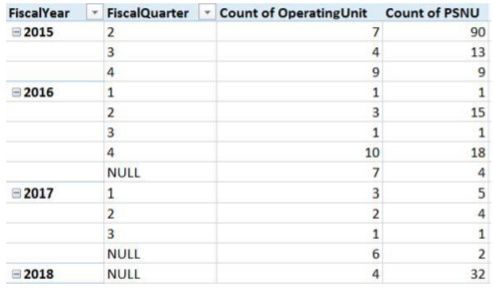
In multiple OUs, data from military facilities funded by the DOD have been reported on at the site-level within DATIM – this is contrary to Appendix G of PEPFAR’s APR Guidance which states that these sites should have been reported at the \_MIL priority subnational unit (PSNU) at each OU. DATIM and DOD POC are in the process of correcting data that has been entered at the site level that should not have been.

In the interim, in order to protect the names and locations of these sensitive military sites, the Site by IM MER Structured Datasets combines (or rolls up) all DOD-funded military sites[[8]](#footnote-8) into the SNU value \_Military (operating unit name).[[9]](#footnote-9) This is in follows DOD-funded military requirements on how these data should have been entered into DATIM by country teams. **Please be aware that this does NOT align with how data is displayed and aggregated in Final.DATIM.org, Panorama, or the other MER Structured Datasets.**

**What this means in practice is you should not roll-up or aggregate the site level data in the Site X IM dataset to their associated PSNU** – as these values will not match what is available in other sources (Panorama, Final.DATIM.org, and other MER Structured datasets). Because all DOD-military data is rolled up, attempting to sum the Site by IM datasets to PSNU levels would result in some PSNUs having lower levels of results and targets while the corresponding military PSNUs would have higher levels than recorded in all other data sources. *If you are interested in looking at data at the PSNU level, please use the PSNU by IM or PSNU MER Structured datasets or refer to Panorama.*

Number of PSNUs Impacted by Rolling-Up DOD Funded Military Sites

Less than 200 out of over 3,000 PSNUs are affected by rolling-up DOD-funded military sites across all time periods in the Site by IM MER Structured datasets. Below is a table showing the number of PSNUs (and OUs) with data from DOD-funded military sites improperly entered at the site-level (rather than the \_Military PSNU) in DATIM over time (please note that targets are displayed as NULL in the FiscalQuarter column).

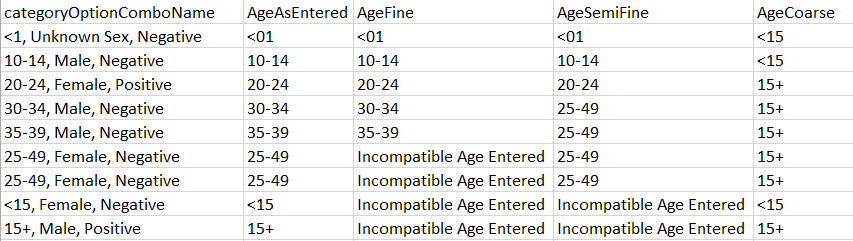


For more details about the full scope of DOD-funded military sites which were submitted at the site-level, please contact the country or HQs DOD representative.

# Appendix D: Additional Guidance: Fine, Semi-Fine, and Coarse Age Band Columns

* The “AgeAsEntered” column corresponds to the value entered in the DATIM data entry screens. This age band does not contain any calculations or aggregations.
* The “AgeFine” column contains only data entered into fine age bands in the DATIM data entry screens. These bands are useful for analyzing the new Fine age bands that were introduced in FY18Q2 (these are sometimes also referred to as 5-year age bands).
* The “AgeSemiFine” column includes data from both the Semi-Fine age bands and the Fine age bands (which are aggregated up to the Semi-Fine level; see below).
* The “AgeCoarse” column contains data entered directly into DATIM at the coarse age levels (<15/15+) or through the Most Complete Age-Sex Disaggregate (MCAD) calculation for FY15-FY17 (refer to the [*Most Complete Age-Sex Disaggregation*](#_Most_Complete_Age-Sex_1)section below). The AgeCoarse data is also derived from data entered at the AgeFine or AgeSemiFine age band levels for FY18, but double-counting is avoided through DATIM mutual exclusivity rules for age disaggregates.

***Please exercise caution when analyzing only the “ageFine” or “ageSemiFine” columns due to possible data completeness issues*.**



“AgeFine,” “AgeSemiFine” and “AgeCoarse” are first captured based on “AgeAsEntered.” If on the DATIM data entry screens data was entered into the fine age band options in FY18Q2 or later, this data will fall into AgeFine. In cases where the coarse age options or adult semi-fine band for age 25-49 are used, the data will also fall into these categories.

Data for adult fine age bands will be found only under AgeFine for FY18Q2 and later. For FY15 through FY18Q1, data entered in DATIM will fit the AgeSemiFine as the finest age options available for those time periods.

# Appendix E: Known Nuances with Frozen Instances

Upon the creation of a frozen instance of DATIM by DATIM Database Administrators, the PDH team works with the DATIM Database Administrators to create the five MER Structured datasets and perform an initial analysis to assess certain aspects of data quality. Given the nature and complexity of the systems involved, nuances are to be expected. Each time that MER Structured datasets are posted, *Release Notes* will also be posted which outline known nuances for that dataset. Users are strongly encouraged to read this document before conducting their analyses. These nuances may affect any data retrieved via any application that draws from the DATIM frozen instance, including Panorama or Final.DATIM.org.

These can include but are not limited to:

1. **Critical notes about Site by IM datasets:** there are several critical issues that are unique to the Site by IM MER Structured datasets and which users must know before attempting analyses. These issues are outlined in the *Release Notes* and in the training module entitled “Special Considerations for Site x IM Datasets”. The two issues outlined in the release notes that need to be highlighted are:
   1. The Site by IM datasets include site (facility/community) names (as of FY17Q1). Sensitive site names that provide services to Key Populations and military populations have been anonymized.
   2. KP\_PREV, KP\_MAT and KP disaggregates for other indicators (TX\_NEW, HTS, and PrEP\_NEW) will not be included in the Site x IM MER Structured Datasets.
   3. The Site by IM MER Structured Datasets rolls-up all DOD-funded military sites that were improperly entered into DATIM at the site-level (rather in the \_Military SNU for each OU per MER Reporting Guidance). This is in contrast to the ways in which that DOD-funded military data is aggregated to the SNU level in DATIM.org, Panorama and the other MER Structured Datasets. We highly recommend when doing SNU level analyses, users use either the PSNU or PSNU by IM MER Structured datasets. For further information please see the *Release Notes* or the“Special Considerations for Site x IM Datasets” training module*.*
2. **Mechanisms being dropped due to mechanisms being deactivated in FACTS Info:** Every night, DATIM receives an up-to-date list of active mechanisms from FACTS Info. Based on the details in this list, DATIM then marks data as either valid or invalid based on whether the associated mechanism is marked active or inactive for the related fiscal year. This may cause mechanisms that had at one point had data entered against them to drop from Genie, Panorama, or MER Structured outputs. NOTE that in many cases, these mechanisms will remain visible via DATIM.org, even when not visible in any other tools. However, the MER Structured datasets match Panorama unless otherwise noted in the release notes.
3. **Mechanisms being dropped due to other issues in the approval workflow:** Data that has been returned to a mechanism for edits or deletion but has not yet been submitted by the mechanism for approval by the related Funding Agency is being dropped from DATIM.org, DATIM Genie, Panorama, and the MER Structured dataset. In some cases, DATIM users with higher-level permissions, or users at the Agency or Interagency level may be able to see this data, while average users will not. This has primarily affected FY 2015 Q2 (SAPR) data. At the closure of this period’s reporting window, mechanisms were promoted one by one to “Globally Approved” status as opposed to promoting entire datasets to “Globally Approved” as has been done for subsequent reporting periods. This may have caused some mechanisms to be left stranded outside the approval workflow. It is also possible these mechanisms were purposefully dropped in response to Operating Unit requests.
4. **Data being dropped due to potential error in data exchange process:** In some Operating Units, potentially due to issues occurring during data exchange processes, data is entered without a necessary mechanism attribution and, when this is imported into DATIM, this data is read as invalid and dropped from being accessible via DATIM Genie, DATIM.org, Panorama, and the MER Structured datasets.
5. **\_Military data appearing in DATIM Genie, but not in DATIM.org**: In some rare cases, data for \_Military units is available in DATIM Genie, Panorama, and the MER Structured datasets, but is not available via DATIM.org. The cause of this error is yet unknown.
6. **\_Military data entered at Operating Unit level rather than at SNU1 level:** In these rare cases, data pertains to a \_Military location and an indicator collected at the Facility or Community level (as opposed to at the OU level), but data was entered such that SNU1 and all lower SNUs are blank (instead of \_Military [*Operating Unit Name*] and the site name is listed as the Operating Unit.

## Special Notes about NAT and SUBNAT data

* NAT and SUBNAT data are defined and collected as part of the MER indicator package.
* These data are entered by USG staff and so have no mechanism, partner, or agency attribution.
* These data represent national-level figures and are not constrained to only those areas supported by PEPFAR.

## Special Notes about IMPATT data

* Planning and Implementation Attribute (IMPATT) data is comprised of the following:
  + Prioritizations at the PSNU, Community, and Facility levels (included in relevant PSNU, PSNU x IM, and Site x IM datasets)
  + SIMS High Volume flag at the Community and Facility levels (not currently included in the NAT/SUBNAT dataset)
  + PLHIV, Prevalence, and Population figures at the PSNU level, disaggregated by age and sex (included in the NAT/SUBNAT dataset)
* PLHIV, Prevalence, and Population figures are reported by PEPFAR Funding Agencies and may represent only those parts of a country where PEPFAR provides support (as opposed to the entirety of the country).

1. The decision to only include two fiscal years’ worth of data was made in part to reduce file sizes. The historical datasets containing data for previous fiscal years will be periodically updated if changes occur that would impact data from previous time periods (e.g., certain types of organizational hierarchy shifts or implementing mechanism changes). [↑](#footnote-ref-1)
2. Occasionally, DATIM Database Administrators refresh the frozen instance of the database more than the planned two times per quarter in order to accommodate major issues with the data for one or more OUs. When this occurs, DAQ will also release a new MER Structured Datasets. [↑](#footnote-ref-2)
3. Please note that the “AgeAsEntered” column corresponds to the value entered in the DATIM data entry screens. For the Most Complete Age Disaggregation this is a calculated value. [↑](#footnote-ref-3)
4. Indicators such as OVC\_SERV and VMMC\_CIRC and a few others may have slightly altered age bands used for data entry in DATIM. More information can be found on [DATIM support](https://datim.zendesk.com/hc/en-us). [↑](#footnote-ref-4)
5. Prior to FY2018, the MCAD selected the most complete age-sex data, selected between Semi-fine and Coarse disaggregate choices for each site by an IM for DSD or TA. The value is then stored as an additional calculated disaggregate value which is displayed as <15/15+, Male/Female/Unknown. This is based on a ‘Proximal-Maximal’ algorithm that chooses only one disaggregate, either Semi-Fine or Coarse entered at the Site-IM-DSD/TA level (and modality in case of HTS\_TST), depending on which one is closer to the total numerator (proximal), or the in the absence of total numerator, the maximal value. In the case of FY17 Targets for HTS\_TST and HTS\_TST\_POS, disaggregates are chosen for the MCAD according to MER Guidance about which targets should have been entered for DREAMS vs non-DREAMS districts. [↑](#footnote-ref-5)
6. The Site by IM dataset is inclusive of each field in this table. The OU by IM and PSNU by IM datasets do not include organizational hierarchy levels below the OU or PSNU levels respectively. [↑](#footnote-ref-6)
7. It is worth noting that the facility unique ID (facilityuid) variable is still included the Site x IM dataset. If a user comes across an unfamiliar-looking site name, believes they’ve found an instance in which a facility with a sensitive site name has been masked, and has a need to know the original facility name, then they can use the facilityuid to lookup the original facility name. For step-by-step instructions on how to lookup the site names associated with each facility unique ID (UID), please refer to guidance written by USAID Office of HIV/AIDS which is available here ([link](https://www.pepfar.net/OGAC-HQ/icpi/Shared%20Documents/Clusters/Geospatial%20Data/Archived%20Workstream%20Documents/USAID_OHA_StructuringDataforUsewithGIS_20160902.pdf)). Although the guidance is specifically geared towards GIS users, steps 1-3 are applicable to anyone who needs to download a list which links all site names to their UIDs [↑](#footnote-ref-7)
8. DOD-funded civilian sites and military sites funded by other agencies were not rolled up to the \_MIL level. [↑](#footnote-ref-8)
9. Please contact your DOD Program Manager or DOD HQ for questions on site-level data for military programs. [↑](#footnote-ref-9)