PEPFAR Data for Accountability Transparency Impact Monitoring (DATIM)

Data Deduplication User Guide (Version 5)

August 2021

Getting Started with PEPFAR DATIM

Please go to www.datim.org, and enter your username and password.

The Data Deduplication App allows DATIM Interagency users to view and reconcile any specific facility site with data entered by more than one Implementing Partner. The purpose of data deduplication is to ensure that results are not being double counted at the site level and to accurately reflect PEPFAR program results.

Note: *Data Deduplication can only take place during an open reporting period. Once the reporting period has been closed in DATIM, you can no longer perform Deduplication.*

Pure Data Deduplication

To access the Data Deduplication App, select the **Apps** feature on the navigation ribbon located at the top right of the screen. Once accessed, select the **arrow button** at the bottom of the App section to see more options (if necessary).
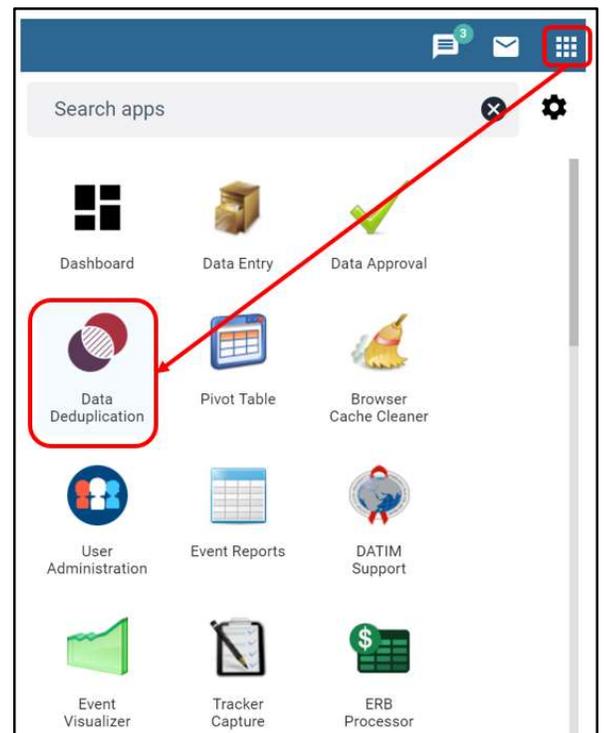
1.) Once you have selected Data Deduplication your screen will appear as follows:

2.) From the OU dropdown menu, you are required to select the **Operating Unit** and **Period**.
   a. When the app first loads, a few required fields are set to a default option:
      i. **Dedupe Type** – **Pure Dedupes** are typically DSD-DSD or TA-TA dedupes, sometimes referred to as "simple" dedupes. These are typically resolved before Crosswalk Dedupes (page 6)
      ii. **Data Type** - **MER Results** are the default option, but users can select Targets if necessary.
      iii. **Status** – The Dedupe app defaults to **Only unresolved**. This means that any potential duplications flagged by the system will display.
   b. Optional filters can also be selected for **Agency** and **Technical Area** (MER Indicator).

3.) Click the **SEARCH DEDUPES** button once all required and optional filters have been selected.

4.) Once the desired OU and Reporting Period are selected, the simple duplicates requiring resolution will display:

5.) The duplicated indicator will display in a table format with the following columns:
   a. **Data Element** – The MER data element identified by DATIM as a potential duplication.
   b. **Disaggregation** – The MER data element's disaggregation.
   c. **Org Unit** – The site/facility name where potentially duplicated records have been entered appears in this column.

d. **Mechanism** – The Mechanism ID numbers for each Partner with a potential duplication.
e. **Agency** - The funding agency of the mechanism with potentially duplicated data.
f. **Partner** – The implementing partner names that submitted the MER data element values.
g. **Value** – The indicator data element value entered by each of the implementing partners.
h. **Resolution**– The 3 possible deduplication resolution actions that can be made to reconcile the values. See the next step (#6) for their definitions.
i. **Status** – The resolution status of the identified duplicated values in the DATIM database.

6.) To resolve the duplicated values, select from the "**Resolution**" radio buttons:

a. **Max** – This option will use the largest value submitted by an implementing partner for the indicator**.** This reports a full or broad duplication of clients served by the Partners.
b. **Sum** – This option will sum all values entered by implementing partners for the indicator and submits the grand total. This reports that no values are duplicated.
c. **Custom** – This option allows you to manually enter a value for the indicator. This reports a partial duplication, or overlap, of clients served by the Partners at this site, for this MER indicator.

Users can also reference the "Information Bubble" feature within the Dedupe App. Use your cursor to hover over a Resolution Type to see a pop-up with a definition and dedupe example of each type.

☐	Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status
☐	PrEP_NEW (N, DSD, Age/Sex) v2: Newly Enrolled PrEP	20-24, Male	Hospital A	321	USAID	Partner 1	3	○ Max (	
				999	HHS/CDC	Partner 5			
☐	PrEP_NEW (N, DSD, Age/Sex) v2: Newly Enrolled PrEP	25-29, Male	Hospital K	123	USAID	Partner 2			
				404	HHS/CDC	Partner 9			

This resolution uses the largest data value submitted between the implementing partners for the indicator. By selecting **Max**, you are indicating that all values are duplicated between both partners.

Example: Partner A tested **5** clients for HIV. Partner B tested **2** clients for HIV at the same site. Using **Max** means there was a reporting duplication (2 of the clients were served by both partners) and only a net total of **5** clients were tested at this site.

NOTE - The definition and example will be different between a Pure Deduplication and Crosswalk Deduplication since they accomplish different actions in DATIM's database

7.) Once the resolution has been selected, click **Resolve** to adjust the Indicator value in the submitted data set.

Reconciling Multiple Duplicated Indicators

1.) Once the desired OU and Reporting Period are selected, and you have applied any desired filters, then clicked **Search Dupes**, the duplicated indicators by mechanism will display.

2.) On the left side of the Data Deduplication app screen, you can perform multiple deduplication actions by clicking **Batch** or the "check" box. These are circled in red to the right:

3.) The **Batch** tab has 3 important sections to aid users when reconciling multiple dupes:

    a. **Status** – This box will indicate to the user the status of every selected duplication.
    b. **Selection** – Use this section to select every filtered dedupe (SELECT EVERYTHING) or only the dedupes that appear on the page (SELECT THIS PAGE).
    c. **Method** – Use this section to quickly specify the mass (batch) deduplication action to resolve the values in DATIM, either by their Max value or Summed value.
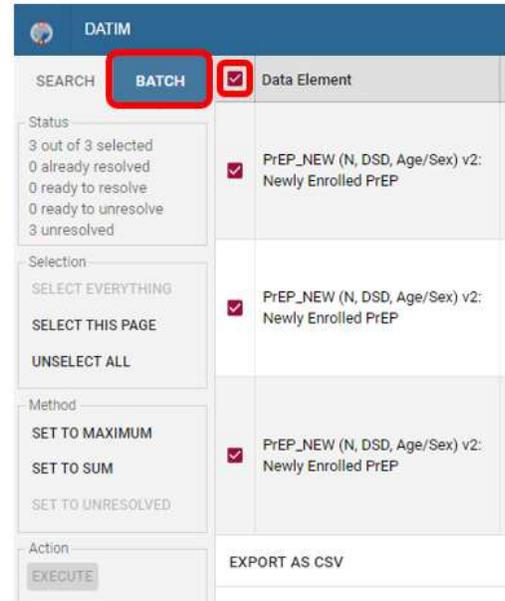
4.) Once a desired dedupe resolution is made, click the **Execute** button.

5.) The Status column will specify if the dedupe resolution was processed.
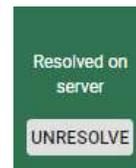
Reviewing and Correcting Deduplicated Records

Once data deduplication has been applied, a review of the duplicated records can be performed.

1.) On the Search Tab, select **Include resolved** for the Status drop down, in order to view already resolved dedupes:
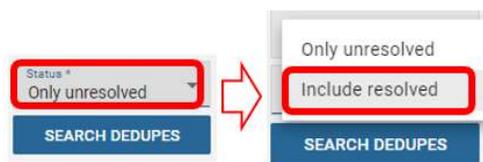
2.) To undo or change an already resolved dedupe, click the **UNRESOLVE** button, which will set it back to "Unresolved":
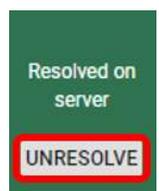
3.) To **UNRESOLVE** multiple dedupes, navigate to the **Batch** tab, select the already resolved dedupes you would like to undo, then click **SET TO UNRESOLVE**, then the **EXECUTE** button:
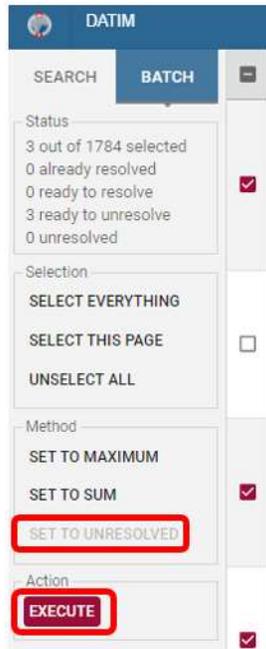
"DSD-TA Crosswalk" Data Deduplication

The DSD-TA Crosswalk allows DATIM Interagency users to deduplicate indicator data across different service types (e.g. DSD vs. TA) to ensure "double counting" is not present. As part of PEPFAR guidance, DSD (Direct Service Deliver) always takes precedence over TA (Technical Assistance) in terms of reporting submissions. Performing a dedupe crosswalk revises the TA value reported in DATIM in order to report the final total of clients served for that indicator, at that site.

Please Note – Most Dedupe Crosswalks cannot be completed until Pure Deduplication has been completed.

1.) Change the left side "Dedupe Type" to **Crosswalk Dedupes** in order to see any potential duplications between DSD and TA.

2.) Then be sure the desired OU and Reporting Period are selected, and click **Search Dedupes** to display the DSD-TA Crosswalk duplicates.
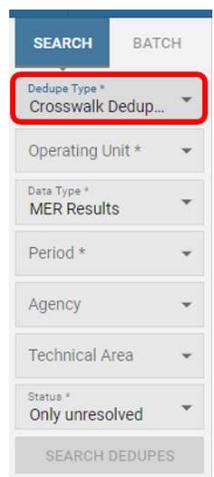
3.) Crosswalk dedupes flagged by DATIM typically display one mechanism along with a corresponding DSD Value:

4.) Similar to Pure Deduplication, there will be 3 resolution actions that can be performed:

    a. **Max** – This option will change the TA value in DATIM's database so that it overlaps as much as possible with its corresponding DSD value. Using Max for a crosswalk dedupe means that all of the TA values are duplications of the DSD value, that the same clients served are being reported between DSD and TA.

        i. **Note –** DATIM will automatically default to the Max resolution radio button for Crosswalk deduplications since it is the most common type of resolution. Users will still need to review and resolve them appropriately.

    b. **Sum** – This will sum all TA values and DSD values for the indicator and submits the grand total. This reports that no values are duplicated, signifying that the TA clients served are distinct from the DSD clients.

    c. **Custom** – This option allows you to manually enter an adjusted TA value for the indicator in order to report a partial duplication of clients served through DSD and TA at this site, for this MER indicator.

5.) Once the resolution has been selected, click **Resolve** to adjust the Indicator value in the submitted data set.