



**PEPFAR**

U.S. President's Emergency Plan for AIDS Relief

# DATIM Deduplication Training Review

# 1. Deduplication Overview

# 1. Deduplication Overview

## Reasons for Deduplication

- “The numbers don’t add up”
  - PEPFAR totals vs. National results
  - Same Things, Same Place, Same Time
- “I did this” vs. “We did this”
  - Partners report individual accomplishments
  - Programs report aggregate, unique results
- Double-counting vs. Service Overlaps
  - Sometimes it’s simple (2 partners, exact same clients)
  - Sometimes it’s complex (overlapping community interventions across populations, geographic areas)
- Program Accountability

# Deduplication Overview

## Why do we dedupe?

- PEPFAR is concerned about individual humans!
  - When two (or more) mechanisms reach the same beneficiaries at a site with the same services, this creates Duplication of reporting values in DATIM.
  - Deduplication allows PEPFAR to reconcile data systems to count *individuals reached by PEPFAR services*, even when the same client may have received services from different IPs and Agencies.
- Dedupe is incredibly important for meeting targets.
  - It can be the difference between meeting your targets and not meeting your targets.

# Deduplication Overview

## When do we dedupe? Who does it?

- Dedupe happens at the Interagency level.
- Dedupe starts once all partner data has been entered into the system.
- Dedupe takes place with all **agency** stakeholders IN THE SAME ROOM.



# Deduplication Overview

## Defining De-Duplication Terms

- Direct Service Delivery (DSD):
  - Direct interactions with the patient or beneficiary.
- Technical Assistance (TA):
  - Support to improve the quality of services through site visits
- “Pure/Normal/Simple” De-Duplication:
  - All DSD-DSD or TA-TA duplicate (or differing) values to be addressed within DATIM.
- DSD-TA Crosswalk/Complex De-Duplication:
  - When multiple partners report DSD and TA results at the same site, which double-counts patients/beneficiaries across DSD and TA results; revise TA value within DATIM.

## 2. DATIM Dedupe Application

# DATIM Dedupe App

## Considerations

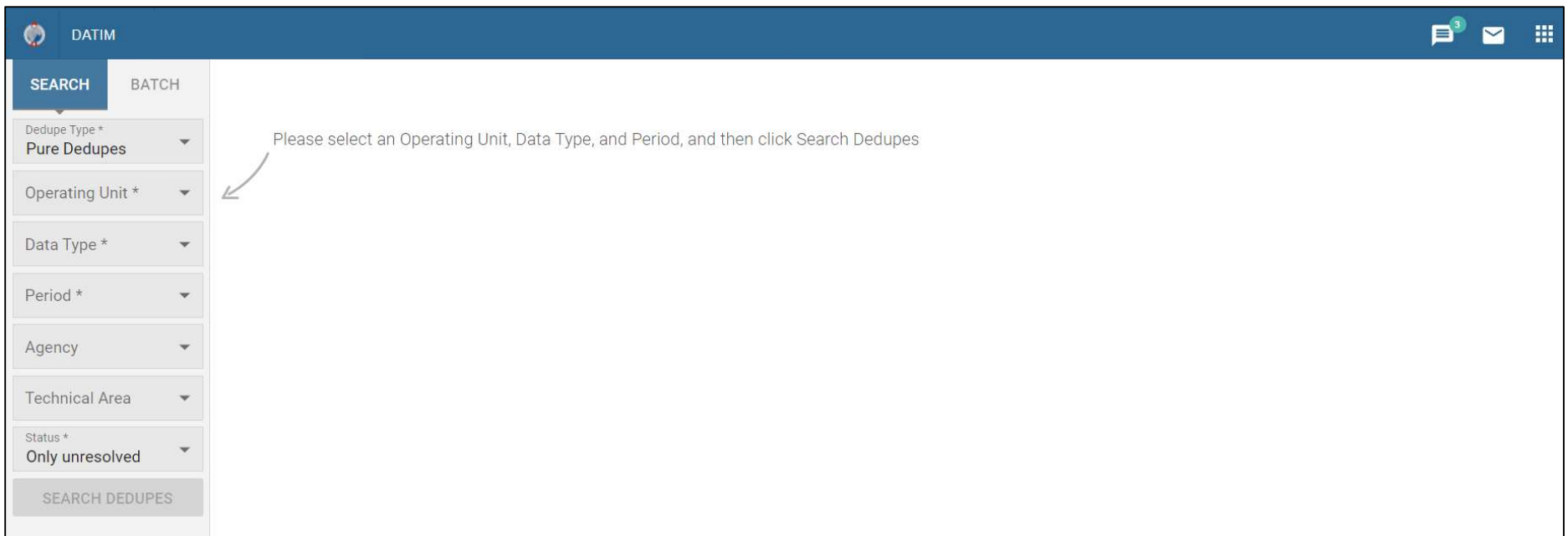
- Before performing deduplication in the DATIM Dedupe App, consider whether the duplicates being resolved are DSD or TA.
  - DSD/DSD and TA/TA duplicates are resolved using Pure Deduplication.
  - DSD/TA duplicates are resolved using Crosswalk Deduplication.
  - If there were any funding mechanisms that were DSD-TA, these are typically done *after deduplicating all DSD-DSD and TA-TA mechanisms*.



# DATIM Dedupe App

## Deduplication App in DATIM

- In DATIM, after selecting Data Deduplication, your screen will appear as follows:



The screenshot displays the DATIM interface for the Dedupe App. The top navigation bar includes the DATIM logo and user profile, and notification icons for messages (3), email, and a grid menu. The main content area is divided into a left sidebar and a main panel. The sidebar contains two tabs: 'SEARCH' (active) and 'BATCH'. Below the tabs are several filter dropdowns: 'Dedupe Type \*' (set to 'Pure Dedupes'), 'Operating Unit \*', 'Data Type \*', 'Period \*', 'Agency', 'Technical Area', and 'Status \*' (set to 'Only unresolved'). A 'SEARCH DEDUPES' button is located at the bottom of the sidebar. The main panel contains a message: 'Please select an Operating Unit, Data Type, and Period, and then click Search Dedupes', with a grey arrow pointing to the 'Operating Unit \*' dropdown.

# DATIM Dedupe App

## Deduplication App in DATIM

- Here we highlight the dedupe entry screen into its components:

SEARCH	BATCH	<input type="checkbox"/> Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status
Dedupe Type * <b>Pure Dedupes</b> Operating Unit * Data Type * <b>MER Results</b> Period * <b>Oct - Dec 2020</b> Agency Technical Area Status * <b>Only unresolved</b> <b>SEARCH DEDUPES</b>	<input type="checkbox"/>	HTS_TST (N, DSD, KeyPop/Result): HTS received results	People in prisons and other enclosed settings, Positive	Alkaleri General Hospital	168	HHS/CDC	CATHOLIC CARITAS	30010	<input type="radio"/> Max (30020) <input type="radio"/> Sum (60030) <input type="radio"/> Custom	Unresolved
	<input type="checkbox"/>				16	HHS/CDC	PUBLIC HEALTH	30020		
	<input type="checkbox"/>	VMMC_CIRC (N, DSD, Age/Sex): Voluntary Circumcised	1-4, Male	Alkaleri General Hospital	180	USAID	ASSOCIATION	8	<input type="radio"/> Max (8) <input type="radio"/> Sum (10) <input type="radio"/> Custom	Unresolved
	<input type="checkbox"/>				180	USAID	WIDOWS AND ORPHANS	2		
	<input type="checkbox"/>	VMMC_CIRC (N, DSD, Age/Sex): Voluntary Circumcised	10-14, Male	Alkaleri General Hospital	180	USAID	ASSOCIATION	8	<input type="radio"/> Max (8) <input type="radio"/> Sum (10) <input type="radio"/> Custom	Unresolved
	<input type="checkbox"/>				180	USAID	WIDOWS AND ORPHANS	2		

Data Element

Mech ID

Partners

Resolution Type

Facility

Agency

Values

Dedupe Status



16 YEARS OF SAVING LIVES THROUGH AMERICAN GENEROSITY AND PARTNERSHIPS

# DATIM Dedupe App

## Standard Deduplication Resolutions

While using the DATIM dedupe app, you must decide which logic to apply to the data values:

- **SUM:** Add all IM results together at site level.
  - (No Duplication of clients by IMs.)
- **MAX:** Across all IMs, report the “highest reported value” of each indicator/disaggregation at the site level.
  - (Full or Broad Duplication of clients by IMs)
- **CUSTOM:** Across all IMs, report a “custom value” at the site level based on contextual or known programmatic factors .
  - (Partial Duplication of clients by IMs)

# DATIM Dedupe App

## Crosswalk Dedupes for DSD/TA

- Shown here is an example crosswalk dedupe (DSD/TA).
- Crosswalk dedupes are done *after* all pure DSD-DSD and TA-TA mechanisms have been deduped.

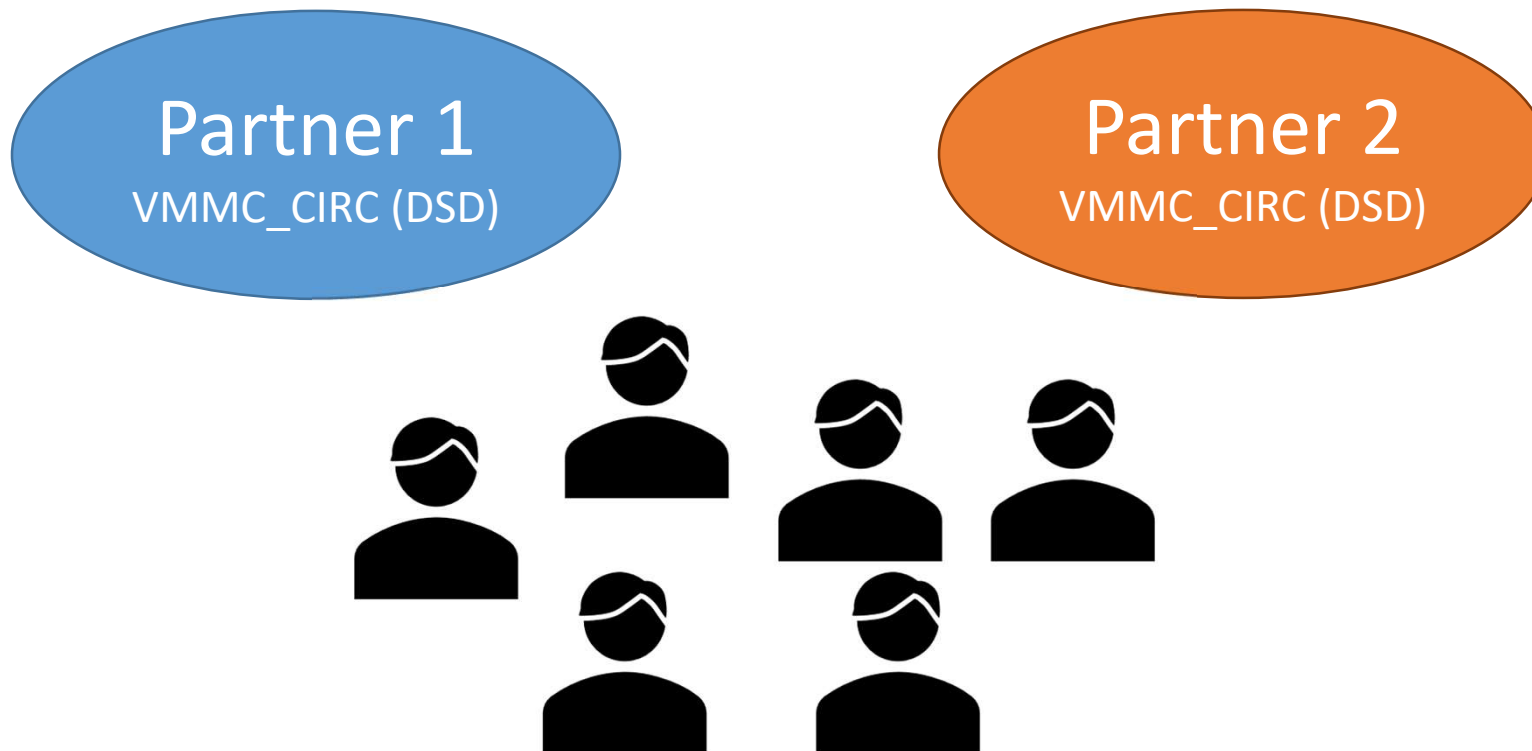
SEARCH	BATCH	<input type="checkbox"/> Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status		
Dedupe Type * <b>Crosswalk Dedup...</b> Operating Unit * Data Type * <b>MER Results</b> Period * <b>Oct - Dec 2020</b> Agency Technical Area Status * <b>Only unresolved</b> <b>SEARCH DEDUPES</b>		<input type="checkbox"/>	HTS_TST (N, TA, Inpat/Age/Sex/Result): HTS received results	<1, Female, Negative	Teaching Hospital	18677	HHS/CDC	CENTRE	3	<input type="radio"/> Max (3) <input type="radio"/> Sum (4) <input type="radio"/> Custom	Unresolved	
		<input type="checkbox"/>				DSD Value	DSD Value	DSD Value	1			
			<input type="checkbox"/>	HTS_TST (N, TA, Inpat/Age/Sex/Result): HTS received results	<1, Female, Positive	Teaching Hospital	18677	HHS/CDC	CENTRE	3	<input type="radio"/> Max (3) <input type="radio"/> Sum (4) <input type="radio"/> Custom	Unresolved
		<input type="checkbox"/>				DSD Value	DSD Value	DSD Value	1			
			<input type="checkbox"/>	HTS_TST (N, TA, Inpat/Age/Sex/Result): HTS received results	<1, Male, Negative	Teaching Hospital	18677	HHS/CDC	CENTRE	3	<input type="radio"/> Max (3) <input type="radio"/> Sum (4) <input type="radio"/> Custom	Unresolved
		<input type="checkbox"/>				DSD Value	DSD Value	DSD Value	1			

## 3. Examples of Deduplication



# Examples of Deduplication

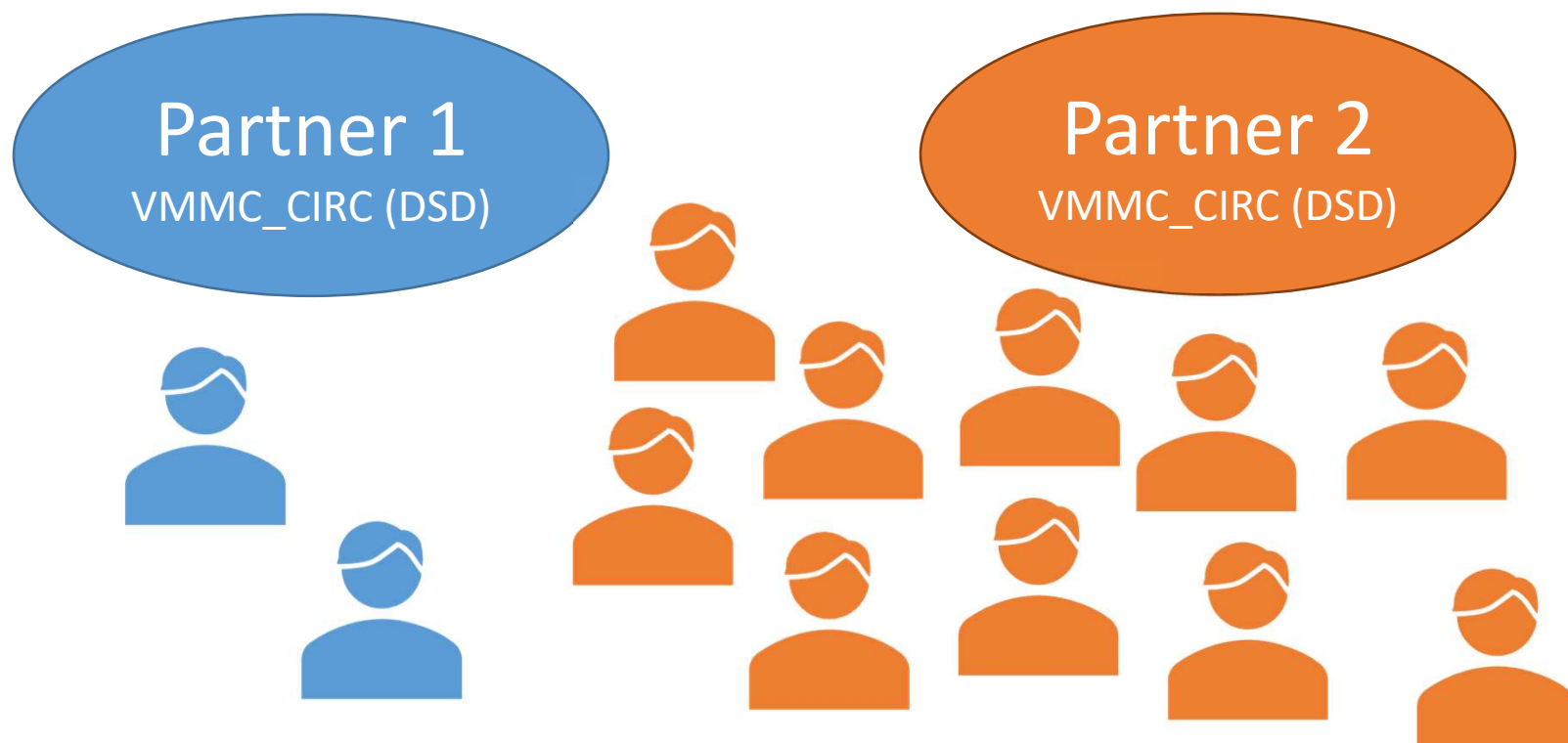
## Example 1: VMMC\_CIRC



In this scenario, two partners are performing VMMC (DSD) procedures at the same site. Is deduplication needed? If so, how do we decide how to deduplicate results?

# Examples of Deduplication

## Example 1: VMMC\_CIRC



In this scenario, **Partner 1** circumcises two individuals and reported a VMMC\_CIRC total of two at this site. **Partner 2** circumcises ten individuals and reported a VMMC\_CIRC total of ten at the same site. In this example, there is no duplication because **you cannot circumcise an individual more than once!**



# Examples of Deduplication

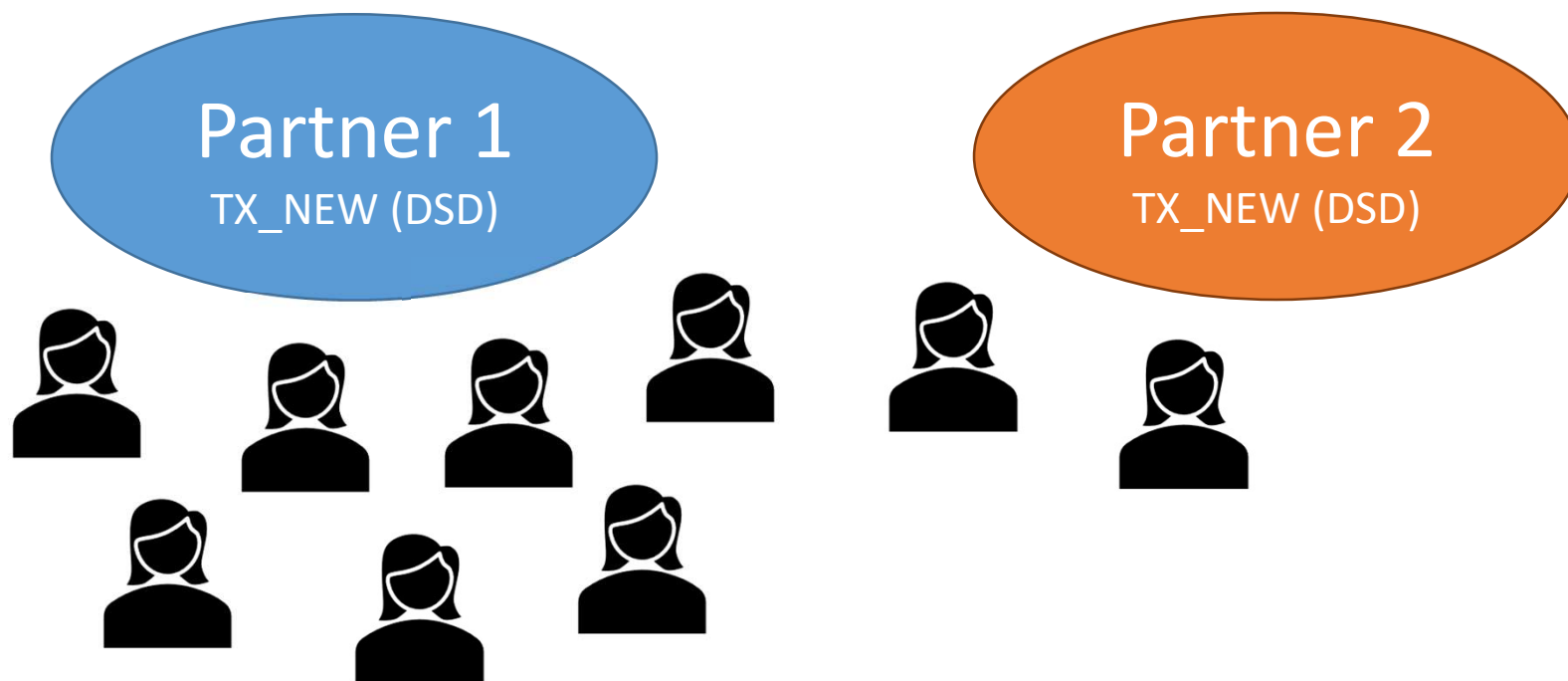
## Example 1: VMMC\_CIRC

<input type="checkbox"/>	Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution
<input type="checkbox"/>	VMMC_CIRC (N, DSD, Age/Sex): Voluntary Circumcised	20-24, Male	Hospital	180	USAID	Partner 2	8	<input type="radio"/> Max (8)
				18	USAID	Partner 1	2	<input checked="" type="radio"/> Sum (10)
								<input type="radio"/> Custom

However, DATIM flags this as a possible duplication because both partners are reporting values for the same data element at the same site. So, in this example, it is best to use **SUM** because both partners cannot circumcise the same individual more than once!

# Examples of Deduplication

## Example 2: TX\_NEW



For this example, the facility provides new clients with antiretroviral therapy (ART). **Partner 1** works with all pregnant women at this site. However, **Partner 2** is dedicated to working with Female Sex Workers (FSW) at this site.

# Examples of Deduplication

## Example 2: TX\_NEW



In this scenario, **Partner 1** enrolled eight pregnant women on ART and reported a TX\_NEW total of 8 into DATIM. **Partner 2** enrolled two FSW and reported a TX\_NEW total of two into DATIM. However, since one of the FSW was pregnant, she was reached by both partners (shown in purple).



# Examples of Deduplication

## Example 2: TX\_NEW

<input type="checkbox"/>	Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status
<input type="checkbox"/>									Unresolved
<input type="checkbox"/>									Unresolved
<input type="checkbox"/>	TX_NEW (N, DSD, Age/Sex/HIVStatus): New on ART 15-19, Female, Positive	15-19, Female	Hospital	1801	USAID	ASSOCIATION FOR REPRODUCTIVE HEALTH	8	<input type="radio"/> Max (8) <input type="radio"/> Sum (10) <input checked="" type="radio"/> Custom 9	Ready to resolve <b>RESOLVE</b>
				1807	USAID	EMPOWERMENT ORGANISATION	2		

It is best to use CUSTOM in this case, since one of the FSW being served was pregnant and received services from both partners. You need to remove the duplicates that overlap by entering the custom value of 9. Ultimately, PEPFAR must have the total number of individuals reached, which in this case, are nine women newly enrolled in ART for this reporting period.

# Examples of Deduplication

## Example 3: HTS\_TST DSD: TB & PMTCT modalities

### *Deduplication among three partners*

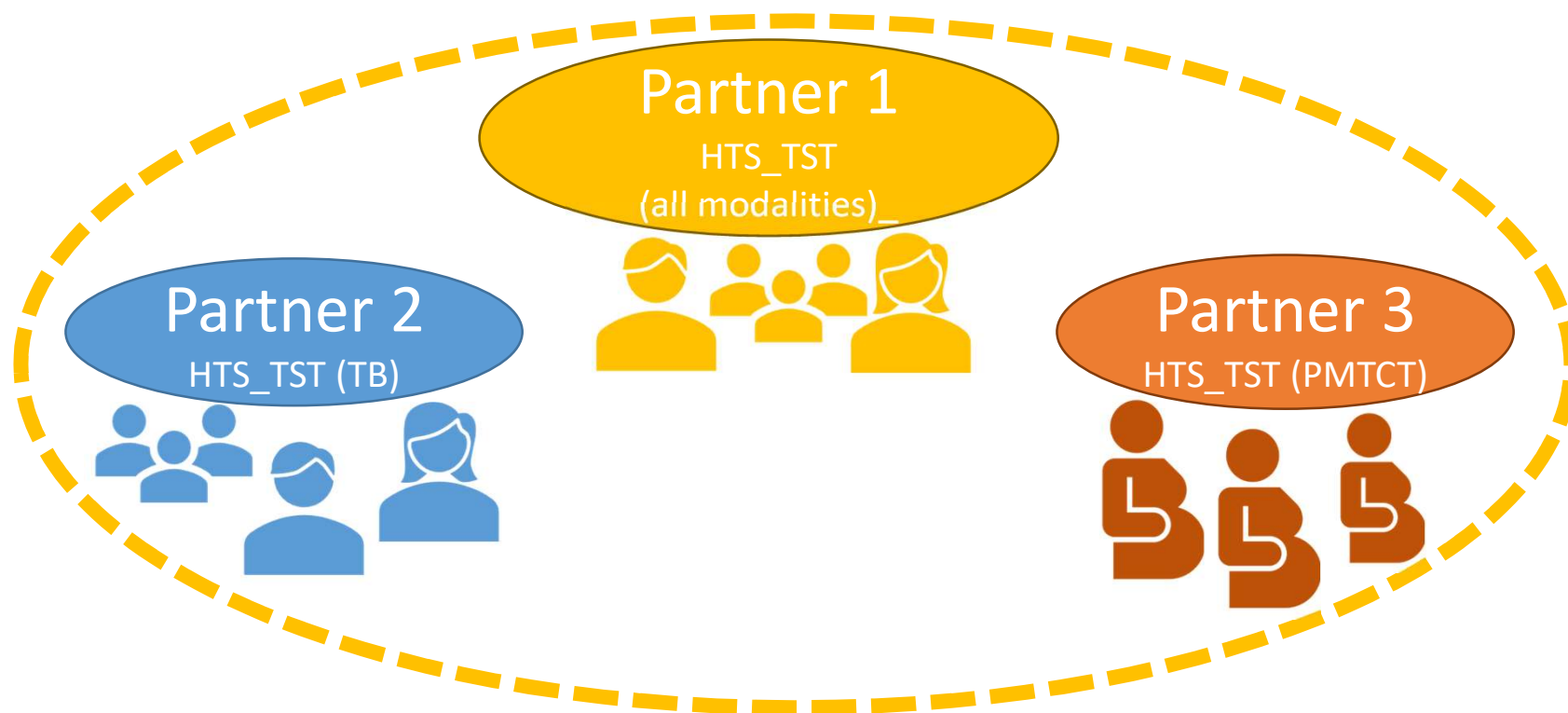


In this scenario, **three partners** are performing HIV testing services at this site:

- **Partner 1** strictly provides HIV testing services at this site.
- **Partner 2** provides a Tuberculosis (TB) clinic at the site, including HIV testing services.
- **Partner 3** provides pregnant women (PMTCT) services at the site, including HIV testing services.

# Examples of Deduplication

## Example 3: HTS\_TST DSD: TB & PMTCT modalities



Having three partners creates two dedupe records because they are all reporting testing services at the same facility – a potential duplication between **Partner 1** and **Partner 2** at the TB clinic, and another potential duplication between **Partner 1** and **Partner 3** at the PMTCT clinic.

# Examples of Deduplication

## Example 3: HTS\_TST DSD: TB & PMTCT modalities

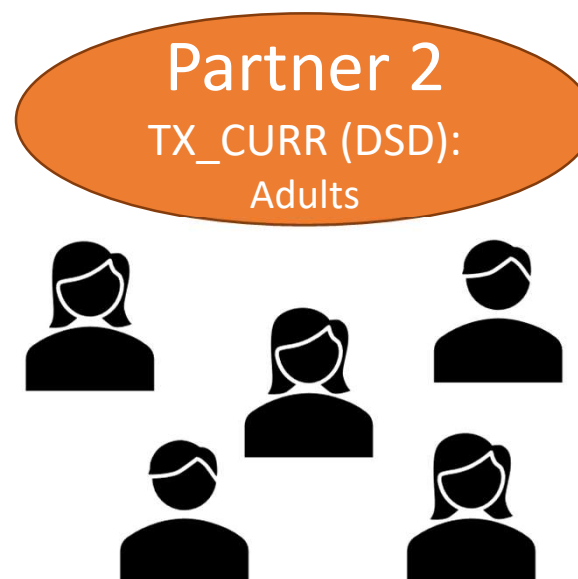
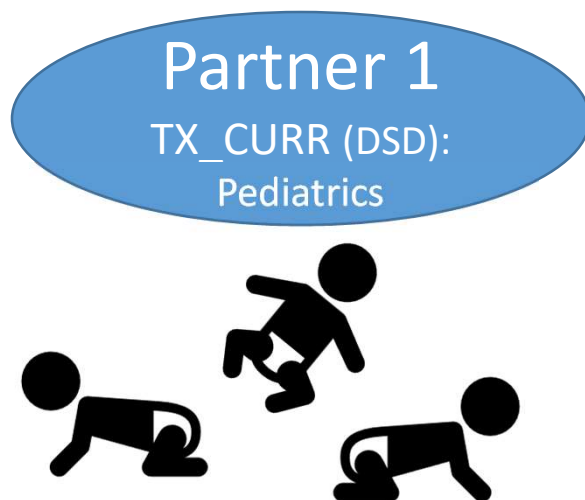
<input type="checkbox"/>	Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status
<input type="checkbox"/>	TB_STAT (N, DSD, Age/Sex/KnownNewPosNeg): New/Relapsed TB with HIV	20-24, Known Positives, Female	Hospital A	321	HHS/CDC	Partner 1	3	<input checked="" type="radio"/> Max (3) <input type="radio"/> Sum (4) <input type="radio"/> Custom	Ready to resolve RESOLVE
1234				HHS/CDC	Partner 2	1			
<input type="checkbox"/>	PMTCT_STAT (N, DSD, Age/Sex/KnownNewResult): Known Results	20-24, Known Positives, Female	Hospital A	321	HHS/CDC	Partner 1	3	<input checked="" type="radio"/> Max (3) <input type="radio"/> Sum (4) <input type="radio"/> Custom	Ready to resolve RESOLVE
180				DSD Value	Partner 3	1			



For both mechanisms, it is best to use “MAX,” because **Partner 1** is reporting the same individuals as **Partner 2** and **Partner 3**. This way, the correct number of individuals tested for HIV are accurately being reported to PEPFAR.

# Examples of Deduplication

## Example 4: TX\_CURR



In this scenario, **Partner 1** and **Partner 2** work at the same site. **Partner 1** provides treatment only to HIV positive children. **Partner 2** provides treatment only to HIV positive adults. Is there duplication between these two partners' work at this site?



# Examples of Deduplication

## Example 4: TX\_CURR

Partner 1  
TX\_CURR (DSD):  
Pediatrics



Partner 2  
TX\_CURR (DSD):  
Adults

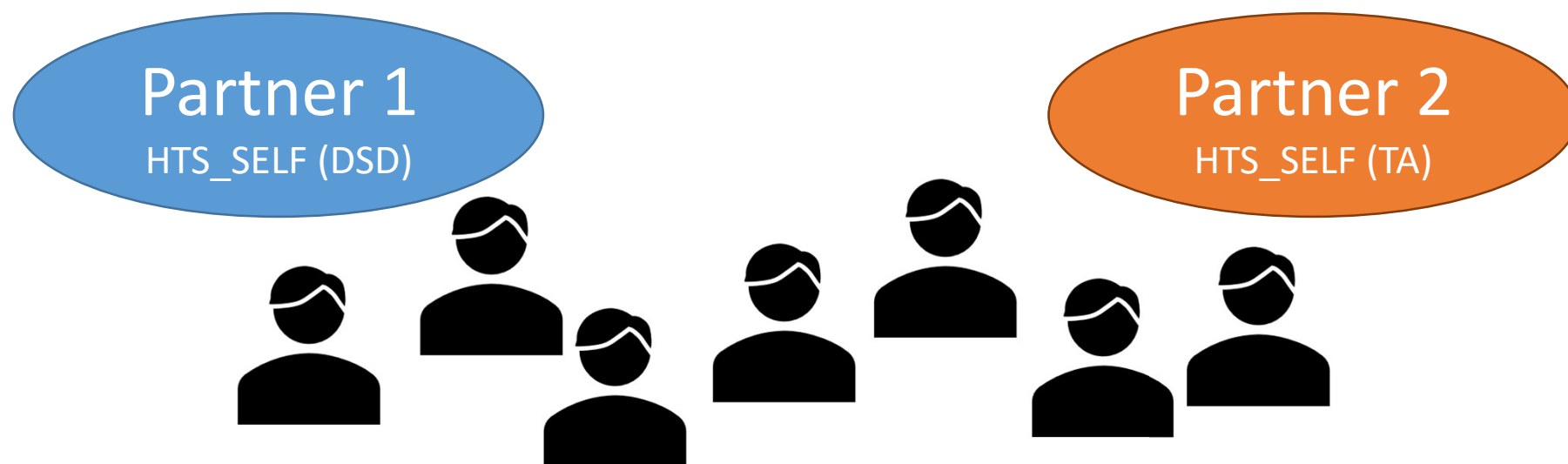


No. Even though these partners provide the same service at the same site, they work with two different populations. Since there is no overlap in individuals served, the system **will not** create a dedupe record.

# Examples of Deduplication

## Example 5: HTS\_SELF

### *Crosswalk (DSD\_TA) Deduplication*



In this scenario, **Partner 1** is distributing HIV self test kits at this site. **Partner 2** is providing training on how to use self testing kits at the same site. Is deduplication needed? If so, how do we decide how to deduplicate results?

# Examples of Deduplication

## Example 5: HTS\_SELF

### *Crosswalk (DSD\_TA) Deduplication*



In this scenario, **Partner 1** reported seven kits distributed to clients at the site as DSD. **Partner 2** also reported seven clients as TA because they provided 7 clients with brief self testing kit training.

This DSD/TA duplication triggers a DSD-TA Crosswalk duplicate that must be resolved in DATIM.

# Examples of Deduplication

## Example 5: HTS\_SELF

### *Crosswalk (DSD\_TA) Deduplication*

<input type="checkbox"/>	Data Element	Disaggregation	Org Unit	Mech	Agency	Partner	Value	Resolution	Status
<input type="checkbox"/>	HTS_SELF (N, TA, Age/Sex/HIVSelfTest): HIV self test kits distributed	25-29, Female, Directly-Assisted	Hospital A	321	HHS/CDC	Partner 1	7	<input checked="" type="radio"/> Max (0) <input type="radio"/> Sum (7) <input type="radio"/> Custom	Ready to resolve <input type="button" value="RESOLVE"/>
						Corresponding DSD Value	7		



All seven individuals were reached by both partners – with Direct Service Delivery (DSD) by **Partner 1**, and Technical Assistance (TA) by **Partner 2**. Any duplicated data in the DSD-TA Crosswalk is assumed duplicate data, as DSD always takes precedence over TA. Specifying a MAX value would “cancel” out the duplicated TA value – this reports the people who were reached by the DSD mechanism, but not by the TA mechanism.

NOTE – using a MAX value is the most common resolution for Crosswalk Dedupes