

## Matching Process

Within the Repository, there are two identifiers used for matching:

- **ID1**–Only appears with TEA Data. The ID1 is a unique replacement for the data posted in the PEIMS ID (PID) field.
- **ID2**– Appears in all data sets TEA and Non-TEA. This variable is a unique substitute for what is posted in the SSN field.

The sections below are intended as guides for matching data files using the identifiers. As mentioned in the *Policies & Procedures: Approved Project Handbook*, the Repository data is held in the format it is received by cooperating agencies. Depending on the research project, it may be necessary to match student level data across data files.

### Matching TEA Student Data

- **Condition**
  - If needing to match TEA Student Data file (master) to other Student TEA Data file (target); matching within PEIMS/TAAS/TAKS
- **Process**
  - **If matching within multiple school years**–Match by ID1 where the `invalid_id1_flag= '0'`. This scenario matches students with a valid TEA identity (permanent PID).
  - **If matching within a specific school year** – You can expand this type of match by including student records where the `invalid_id1_flag = '1'`. These are records where students have only been assigned a temporary TEA identity. Although these student records will not carry forth from year, the ID1 and `invalid_id1_flag = '1'` identifies a unique student record. This record can be matched with other TEA data files within the same school year.
- **Other Considerations**
  - A student may have attended school in multiple districts even within the same fiscal year. This may cause multiple student records (one for each district), where you may be expecting unique student records. Depending on the nature of the research, it may be appropriate to also match on the District variable

### Matching All Other Data

- **Condition**
  - If needing to match TEA to THECB, THECB to TEA, SBEC to TEA, or other combinations.
- **Process**– For the purposes below, we will call “valid ID2s” as ID2s that replace valid SSNs.

**Step 1**–Identify valid **ID2s** in **NEW** data set (Recently released file to THECB for the ERC Repository)

- **TEA** (Valid ID2 criteria (PEIMS/TAKS/TAAS)– State\_Assigned\_Flag='0', Invalid\_ID1\_Flag='0', ID1 field is not blank, and Invalid\_ID2\_Flag='0' (this last variable is in TAAS/TAKS only)
- **TWC** (data not yet posted) – ID2 is always valid unless the ID2 field is blank
- **All Other Sources** (THECB SBEC etc.) – Valid ID2 Criteria - ID2 field is not blank, and Invalid SSN Flag = '0'

**Step 2**–Match valid **ID2s** with **target** data

- No criteria are needed for target data. If the ID2 is valid in the **NEW** file, the ID2 that matches in the target data will also be a valid ID2

### **Additional Matching TEA data with THECB Data**

- **Condition**– Beginning in FY2009 the THECB Enrollment Report CBM001 includes the field TEA\_ID2. When this field is populated, it means the corresponding TEA data for this student included a State Assigned SSN/ID2 (state\_assigned\_flag = '1').
- **Process**
  - In these cases, the THECB SSN/ID2 variable will not match the TEA's State Assigned SSN/ID2. However, by matching the populated CBM001 TEA\_ID2 field to the TEA's ID2 field you will create a match. Although the TEA\_ID2 field only appears on the THECB CBM001, you can carry the TEA\_ID2 variable to other THECB data sets by matching on the ID2.

### **Other Approaches**

The Texas ERC recognizes some limitations in the data matching process. Below is something to consider with caution. The following description may assist the researcher in matching, but should be used with caution understanding the limitations and issues it presents.

- Higher Education student records that have an Invalid\_SSN\_Flag='1', generally have ID2's that cannot be used for matching. However, when a College/University cannot assign a SSN/ID2, the school is supposed to assign a SSN replacement unique to the school (fice). This same SSN/ID2 replacement number will follow the student through subsequent years at the same school. Hence, matching on the ID2 **and** the FICE will allow you expanded Higher Ed school tracking. The problem with this method is that Higher Ed data sets still have some SSN records entered as '00000000' and also as other invalid extraneous "gibberish". Some schools are better/worse than others. In addition, each Higher Ed schools '00000000' record has a different ID2. Where these bad records are obvious non-matches when you can see the actual SSN data, they are not obvious at all when you only have the ID2. The older the data, the worse the SSN/ID2 data you will encounter. **Under no circumstance should this be used on data prior to FY 2000. Use this search method at your own risk.**

## Teacher IDs

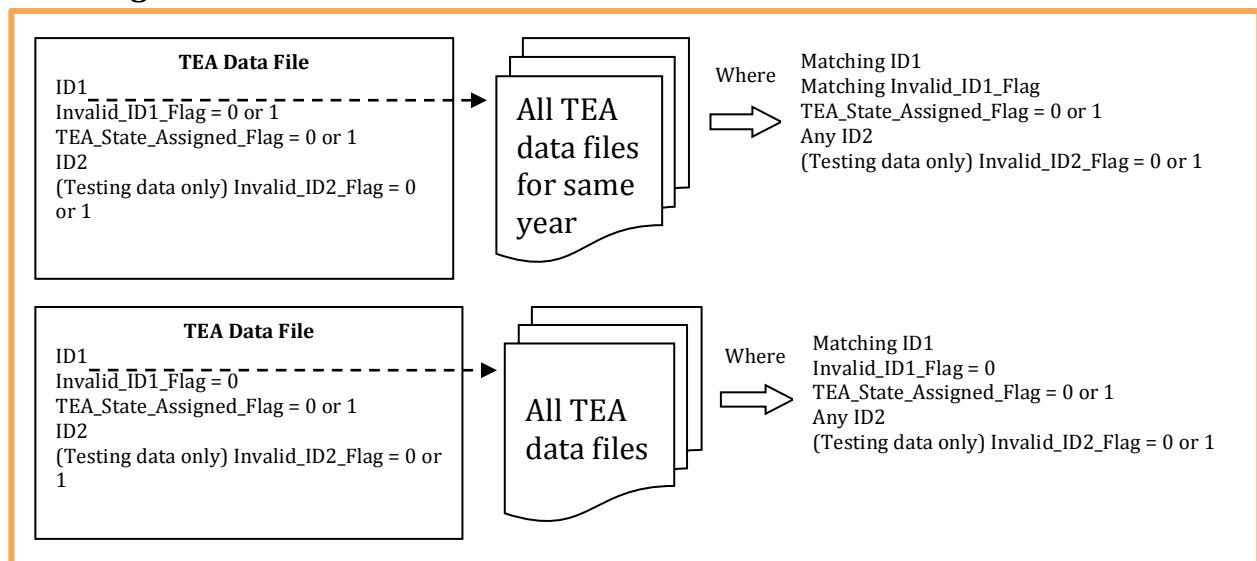
- There is not an invalid\_ID1 field on the SBEC certification data files because TEA only keeps the certification records that match TEAs crosswalk. So, there are no invalid ID1s.
- There are no invalid SSNs or PID numbers in P.employ, and there are no state-assigned numbers given to staff.

## Student IDs

- Generally, all students in TEA data sets have either a valid SSN or a State Assigned (state assigned flag='1'). Either way, there is a unique ID2 posted for both SSNs and State Assigned SSNs. Technically, the ID1s with a State Assigned Flag = 1 cannot be matched with non-TEA data. Since matching within TEA data uses the ID1, those ID2s where there is a State Assigned Flag='1' are problematic.
- Each year there can be a unique record for each ID1/ID2 combination. Where there may just be two records where the ID1 matches more than one ID2 on one file, each of those unique ID1/ID2 combinations may actually show up over the years, possibly even once for each year. So each record of two ID1/ID2 combinations could potentially have many mismatches over time.

## Matching Process Visuals

### Matching Student Records within the TEA Data Files



## Matching Student Records within the THECB Data Files

