Procedure for Processing Education Research Center (ERC) Supplemental Data

Purpose

This procedure outlines the steps for ERC researchers submitting individual-level supplemental data after Advisory Board approval. It ensures proper data matching with the Texas Education Agency (TEA) and/or Texas Higher Education Coordinating Board (THECB), while maintaining FERPA compliance and adhering to relevant legal and ethical standards.

Submission Preparation and Required Documentation

## **Placement in Queue**

* If there are **no edits** to the data dictionary after Board approval and the researcher has a signed Supplemental Data Usage Approval Form or another agreement with the data owner to use the data for their research, the proposal will be immediately placed in the queue for processing.
* If there are **edits** to the data dictionary or an **outstanding agreement** between the data owner and the researcher, the ERC director will inform TEA/THECB when the proposal is ready to be placed in the queue.
* TEA/THECB will review the data for quality, format, compliance with the approved and completed data dictionary, and adherence to FERPA requirements.
  + All variable names and values must exactly match those listed in the approved data dictionary.
  + TEA/THECB will ensure that the submitted supplemental data aligns with what is described in the Proposal and Additional Data Request(s) and approved by the ERC Advisory Board.
  + Note: If any issues are identified with the data file, TEA/THECB may reject the file and will notify the ERC director.

## **Required Documentation**

* Researchers must submit a complete, approved data dictionary using the provided template.
* **Note**: If researchers submit variables that are listed in or similar to the personally identifiable information (PII) categories outlined in Appendices A and B, it will cause significant processing time, as these require extensive review and legal attention to ensure compliance with FERPA and mitigate reidentification risks.
* Do not submit variables that are redundant with existing CBM, TEA, NSC, or TWC data in ERCs.

## **Updates to Data Dictionary**

* If additional variables need to be added after the Board approval, researchers must submit an ‘Additional Data Request,’ which will be reviewed at a future Quarterly Board Meeting.
* The updated data dictionary must include both pre-approved and newly requested variables.
* Once the request is approved, the supplemental data will be placed at the end of the processing queue.

## **Researchers with Prior Board Approval**

* Researchers whose use of supplemental data was previously approved by the Board (without a data dictionary or with a non-THECB template) and whose supplemental data has not yet been processed by TEA/THECB may be required to work with their ERC directors to submit a data dictionary using THECB’s template.
* ERC directors will review the completed data dictionary and submit it to THECB for further review.
* If the researchers are requesting ID1 matching, THECB will forward the data dictionary to TEA for additional review.

Data File Submission and Transfer Guidelines

## **Data File Format**

* Acceptable formats: SAS datasets, Excel, or CSV.

## **Variable Naming Conventions**

* Variable names must not include spaces.
  + For example, change ‘First Name’ to ‘FirstName.’

## **Required Data Elements**

* Include as much personally identifiable information (SSN, first name, middle name, last name, date of birth) as possible for successful matching.
  + First, middle, and last names must be in separate fields.
  + SSN must be 9 characters (no hyphens, and leading zeros must be included).
  + Birthdates must be numeric in YYYYMMDD format.
  + Names should consist of characters only, with no special characters.

## **Data Transfer Process**

* TEA/THECB staff will coordinate with a single entity to ensure a secure data transfer method and confirm receipt.
* If submitting multiple supplemental data files, all files must be in the same format. TEA/THECB will not clean or reformat the files for researchers.
* Data files must be submitted to TEA or THECB only through a secure transfer portal provided by the agencies.
* Submission depends on the approved ERC data being matched to:

|  |  |  |
| --- | --- | --- |
| **TEA Data** | **THECB Data** | **Process** |
|  |  | Submit supplemental data to **TEA**. Once the data file is in queue, **TEA** will provide instructions to the researcher/data owner for data transfer. |
|  |  | Submit supplemental data to **TEA**. Once the data file is in queue, **TEA** will provide instructions to the researcher/data owner for data transfer. |
|  |  | Submit supplemental data to **THECB**. Once the data file is in queue, **THECB** will provide instructions to the researcher/data owner for data transfer. |

Data Matching and Processing

## **Matching Process**

* TEA matches the individual-level supplemental data to K-12 database, using SSNs or a combination of names and birthdates. Then TEA assigns ID1s and forwards the data to THECB.
* THECB matches data, using SSNs or a combination of names and birthdates, to the higher education database, assigns ID2s, and removes all PIIs before transferring the processed data to the ERC for researcher access.

## **Processing Time**

* TEA/THECB processes data based on the date of receipt. Processing time will vary depending on data quality, size, and scope. The following are estimated times only:
  + TEA (ID1 matching): Expect 3-6 months
  + THECB (ID2 matching): Expect additional 2-3 months

## **Important Notes**

* Fuzzy matching will not be conducted. Matches will not account for variations in names, such as alternate spellings or abbreviations. Only exact matches based on SSNs or precise combinations of names and birthdates will be performed.
* TEA/THECB will not clean or modify data to increase the matching rate, nor will they work with researcher-provided code.
* Any data obtained through the Public Information Request process **cannot** be processed as supplemental data, as it is masked and deidentified before release and cannot be reprocessed to add the ERC ID.

Access Restrictions for Supplemental Data

* Researchers should not access their supplemental data externally when they are using the processed deidentified supplemental data at the ERCs.

Data Destruction

## **Timeline**

* TEA/THECB will destroy all supplemental data within 60 days of the Advisory Board-approved project end date.

For questions, researchers should reach out to the ERC directors.

**Appendix A**

**List of Possible Personally Identifiable Information (PII)**

The table below provides a categorized breakdown of various types of information that can be used to identify individuals. While the table outlines a wide range of PII examples, it is not an exhaustive list. Additional types of information may also be considered PII depending on the context in which they are used or combined with other data.

|  |  |  |
| --- | --- | --- |
| **Category** | **Types of PII** | **Examples** |
| Names | Individual Names | Researcher, Parents, Guardians, Legal Names, Aliases, Agencies |
| ID Numbers | Identification Numbers | SSN, Driver's License Number, Client Number, Personal ID, Student ID, Financial Account Numbers (debit/credit), License Plate, IP Address, Vehicle ID/Serial Number, Certificate/License Numbers, Record IDs, Employer ID, Badge ID, Researcher-created IDs |
| Location Information | Location or Address Data | Home Address, Place of Birth/Marriage/Death/Divorce, Geolocation Data (GPS coordinates), Postal Code, Voting/Polling Location, Census Geographic Areas |
| Dates | Time-Based Identifiers | Birthdate, Admission Date, Discharge Date, Date of Death, Service Dates, Marriage/Divorce Date, Conviction/Arrest/Disposition Dates, Event Participation Dates, Voting Date |
| Other Identifiers | Contact or personal details | Phone Numbers, Fax Numbers, Email Address, Medical Record Numbers, Health Plan Numbers, Device Identifiers, Facial Photographs/Images, Job Title, Political Affiliation, Sexual Orientation, Voting Party |
| Biometric Identifiers | Physical or Biological Data | Voiceprints, Fingerprints, Retina/Iris Scans, Hand Geometry, DNA, Facial Recognition Data |
| Sensitive Personal Information | Additional Sensitive Data | Medical Conditions, Genetic Data, Criminal History (including offense/charge description), Financial Data, Insurance Policy Numbers, Employment History, Religious Affiliation, DPS Offense Codes, Driver Records, License Status, Traffic Violation History |

**Appendix B**

**Examples of Potential Personally Identifiable Information (PII) When Combined**

The table below provides examples of how combining secondary identifiers with other data points can increase the risk of identifying individuals, highlighting specific combinations and their potential privacy implications.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Secondary Identifiers** | **General Example of Potential PII** | **Specific Example** |
| **Personal Characteristics** | Race | Zip code + race + school district | Pacific Islander student in a small school district |
| Language Proficiency/Primary Language | Zip code + language +school | Mandarin-speaking student at a bilingual elementary school in a specific district |
| Educational Attainment | Zip code + educational attainment | M.D. in a rural county |
| Financial Information - Salary | County + salary | $300k+ salary in a rural district |
| Gender | Gender + major | Female biology major at a small university |
| Religion | Religion + zip code | Seventh Day Adventist in a small town |
| Hobbies or Interests | Hobby + gender or location | Male ballet dancer; Female student participating in robotics club at a rural high school |
| Health Information (Prescriptions, Genetics) | Genetic condition + zip code | Congenital heart condition in a small town |
| Employment Information | County + job title | Veterinarian in a rural town |
| Voting Preference | Voting preference + student club or political group + school | Student from a political science department voting Republican |
| Weight | Weight percentile + club affiliation | 60kg powerlifting competitor |
| Court Records | Type of conviction + school | Student with a criminal record attending a community college |
| **Location/Address Info** | Geographical Indicators | Zip code + religion, birth county, or marriage county | Seventh Day Adventist in a small town |
| Place of Birth | Place of birth + educational attainment | M.D. born in a small town |
| Neighborhood Characteristics | Community infrastructure + school district | Bike path, park, and pool in AISD |
| Homeownership Status | Neighborhood characteristics + homeownership status | Owning a home in a predominantly rented neighborhood |
| **Group Characteristics** | Languages Spoken in the Home | Language + neighborhood | Mandarin speaker in low-income housing |
| Household/Sibling ID, Family Size | Household ID + marital status or job title | Divorced sheriff with a family of 5 |
| Family Structure | Family structure + family income | Family of 12 with an income of $300k+ |
| Utilities Usage/Provider | Wi-Fi provider + county | Usage of less common utility provider in rural area |
| Activities (Event/Occasion) | Event + job title | Sheriff attending a school board meeting |
| Housing Type | Neighborhood characteristics + house type | Duplex in a primarily townhouse neighborhood |
| Organization/Clubs | County + affiliated club + gender | Female kicker for a football team |
| Pets | Pets + neighborhood characteristics | Horse owner in a small town |
| Transportation | Vehicle ownership + location | No car in a small town |
| Social Assistance Program | Participation in assistance program + location | WIC participant + primary language spoken |