# Linkage Data Preparation

Sites can extract, transfer, and load (ETL) clear-text identifiers used in the linkage process from a local data source into the structure and format of the *source\_data table (Table 1)* into a flat file on a local file server housed entirely within the site’s secure environment. Clear-text identifiers will not leave sites’ server at any point in the CURL record-linkage process. The clear-text flat file will be hashed locally using one-way hashing functions (i.e., SHA-256) by the CURL Site application to produce hash data, also in flat file format. Only hash data will be securely transported to the honest broker. Encrypted hash data can be transferred to the CURL Honest Broker using PopMedNet.

Conventions

* Export patient’s identifiers into flat file format using pipe (|) delimiter (use the attached source\_data.txt file as the template)
* For each data value, make sure of the following
  + No quotations
  + No pipe (|) in any of the data value
  + No new line character in the data value
* Include header as line 1. The delimiter of the header must match the delimiter of the data.
* Don’t use NULL or empty string with quotations (e.g. “”) for fields with missing value
* Use UTF-8 encoding

Table 1 - source\_data table

|  |  |  |  |
| --- | --- | --- | --- |
| Field name | Data type | Required | Note |
| id | Integer | Yes | A random unique identifier of each patient in the source dataset. Source unique identifier such as medical record number (MRN) should NOT be used. Maintain the linkage between the source identifier and the random identifier. |
| first\_name | Varchar (50) | Yes | First name |
| middle\_name | Varchar (50) | No | Middle name or middle initial |
| last\_name | Varchar (50) | Yes | Last name |
| dob | Date | Yes | Date of birth. Format = yyyy-mm-dd |
| sex | Varchar (1) | Yes | F: female  M: male  O: other |
| ssn4 | Varchar (4) | No | Last 4 digits of SSN |
| street\_address | Varchar (255) | No | Street address line 1 (from current address) |
| city | Varchar (50) | No | City name (from current address) |
| state | Varchar (2) | No | State abbreviation of street\_address (from current address) |
| Zip5 | Varchar (5) | No | 5-digit zip codes (from current address) |
| race | Varchar (50) | No | Use race\_concept\_id in the Person table in the OMOP data model. |
| ethnicity | Varchar (50) | No | Use ethnicity\_concept\_id in the Person table in the OMOP data model. |
| home\_phone | Varchar (10) | No | Current US home phone number |
| cell\_phone | Varchar (10) | No | Current US cell phone number |
| email | Varchar (50) | No | Current email address |
| pcp\_npi | Varchar (50) | No | Primary care physician NPI |