

Skid Row Collaborative Study

FEASIBILITY OF DATA LINKAGES USING DE-IDENTIFIED RECORDS

Data linkages will be made using unique, anonymous identifiers. This makes it possible to include de-identified records in the data merges.

The solution to the problem of how to include records without anyone seeing personal identifying data and yet being able to integrate additional data from county departments with those records is to use SAS DataFlux version 8 software. This rules-based matching software uses a combination of parsing and standardization rules, phonetic matching, and token-based weighting to standardize source information. After applying multiple data standardization rules to each field, the software outputs a “match key” referred to as the “statistical linkage key” (SLK). The SLK is an accurate representation of all versions of the same data, such as different spellings of a name or an address. By assigning an SLK to a record it is possible to link records without using any identifiable information about clients. The code itself is anonymous, for example, RX19E4, however the same code will be generated when any permutation of the same identifying information is passed through the SAS DataFlux match code generation algorithm. The Evaluation and Research Unit of the Los Angeles County Chief Administrative Office will carry out this record matching task using SAS DataFlux software.

Skid Row Housing Trust records with client information can be converted into de-identified records with SLK linkage by using the SAS DataFlux software to produce statistical linkage keys, or SLKs for each record. This will become the primary working file for this project. All fields containing individual identifiers such as name or taxpayer identification number will be deleted.