PiiScrubber

An application to obfuscate personally identifiable information in one or more ISAM or RELATIVE files.

# Objectives

* Hand craft a Synergy function that knows how to replace the values of PII fields with new randomly generated values.
  + Randomly generated data must still look “correct” (i.e. a persons name must still look like a persons name, not just be randomly generated data).
* Identify certain fields as containing PII. This will be done by “tagging” those fields in the Synergy repository.
  + These tags will identify the TYPE of PII data in a way that will make sense to the randomization function.
* Process one or more ISAM and/or RELATIVE files. For each file:
  + For ISAM files, create an XDL file.
  + For relative files, determine the record length of the original file.
  + Unload the file to a sequential file (processing will take place on sequential files to avoid issues with keys that do not allow changes).
  + Open the sequential file and process each record
  + Identify each PII field (this will be done via a new CodeGen field loop expression token <IF CONTAINS\_PII>)
  + Replace the PII fields field’s with new values by calling the randomization function.
  + Write updated records to the sequential file.
  + When all records have been processed, close the sequential file and convert it back to the original format (ISAM or relative).

# Tagging of PII Fields

Fields containing PII must be “tagged” by adding a special token to either their user-defined text or long description fields in the repository. These tags will convey three things:

1. That the field contains PII.
2. The type of PII ( First Name, Last Name, Full Name, Phone Number, etc.)
3. Whether a new random value can be generated, or whether it is necessary to persist the replacement data so that the same data can be injected into other (related) files

The format of the tags shall be:

PII\_RANDOM=*type;*

PII\_PERSIST=*type;*

Where type is replaced by the type of data. For example:

PII\_RANDOM=PHONE\_NUMBER;

PII\_PERSIST=CUSTOMER\_ID;