Privacy Preserving Record Linkage: Linkja

Linkja is designed to uniformly de-identify patients and securely link them across sites e.g., health systems, public health records, criminal justice system etc. without exposing underlying PHI. It is open source and available under General Public License. There are 3 main modules:

1. **Salt engine and project, user and key management interface**
   The salt engine generates encrypted shareable salt files. The engine can be used as a stand-alone utility or coupled with a web interface. The web interface enables project managers to create projects, add users and authenticate users. Authenticated users upload their RSA public keys[1] and download the asymmetric encrypted salt file.
   Language: Java

2. **De-identification: Data standardization, exclusion, and hashing**
   This module includes a data pipeline to digest and validate data, standardize data, manage data exclusions, create composite identifiers from patient identifiers, and hash them using SHA512[2] algorithm and the shared Salt.
   Language: Java, SQL-batch scripts

3. **Disambiguation**
   This module allows the aggregator to merge files, disambiguate the hashes and assign master (global) patient ID to matched and non-matched patients using deterministic algorithms.
   Language: Java (with SQLite)

Each module is available as a separate tool in multiple formats and can be used independently as long as the inputs for each module meet the general requirements.

[1] https://en.wikipedia.org/wiki/Public-key_cryptography
[2] https://en.wikipedia.org/wiki/Secure_Hash_Algorithms