

Privacy Preserving Record Linkage (PPRL)

Previously Referred to as Blinded Record Index (BRI)

The Need

- Link data sources to get better outcome data
 - Currently available
 - Electronic health record, anesthesia record, ACS-NSQIP, STS locally using MRN
 - Michigan Value Collaborative payer data using date of service, age, hospital
 - Allow persistence of linkage for future datasets
 - State-specific mortality databases
 - BCBS of Michigan
 - Medicare, other payers
 - Pharmacy databases
 - Ensure no single point of security or privacy failure

A wise man once told me

- “When you have two choices, the one that is more work for you is almost always the right one” – Mike Englesbe
 - Applies to clinical decisions
 - Also applies to privacy and security
- We must NOT communicate or store real identifiers at the MPOG central server, even if the IRB allows us to do so
 - Do you want your family’s SSN sitting at the coordinating center?

The Solution

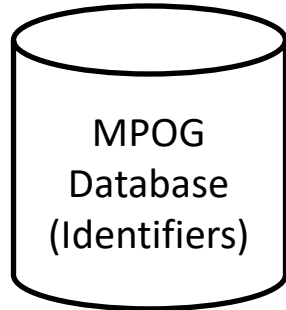
- Blinded Record Linking (aka Privacy Preserving Record Linkage)
 - For merging datasets across sites without PHI
 - Using US National Institutes of Standards and Technology Secure Hashing Algorithm
 - Incorporate RSA secure key
 - Keep the hashed codes centrally, no source identifiers
 - Hashed codes can be linked across data sources
- Established
 - Public domain hashing algorithm
 - “Blessed” by federal government in registries manual as non-PHI
 - Approved by IRBs, DUA at all MPOG institutions

What is Hashing?

- Publicly available one-way mathematical function
 - Input: a string (identifier)
 - Output: a really long alphanumeric code
- Cannot be “undone”
 - It is NOT encryption (which is meant to be decrypted)
- There are no “collisions”
 - No two strings result in the same hash code
 - Source uniqueness is maintained
 - Small changes in input string result in large changes in output code
- Example: Sachin →

964FA766292D50C62019ED7D33232559C0AF511CDDD5E55938AA845698F771B9

Data Source 1



Secure Key

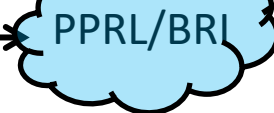


Multiple Hash

Hash	Code 1
Hash	Code 2
Hash	Code X

Fire Wall

https

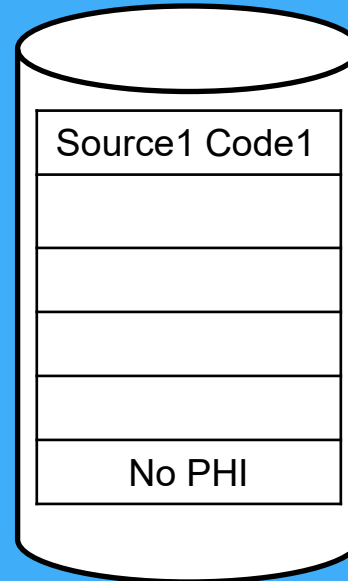


Multiple
Hash

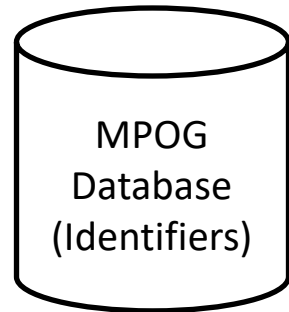
BRI (PPRL)
At University of Michigan

Fire Wall

Fire Wall



Data Source 1



Secure Key



Multiple Hash

Hash	Code 1
Hash	Code 2
Hash	Code X

https



Multiple Hash

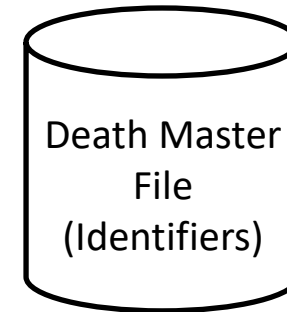
Secure Key



Multiple Hash

https

Data Source 2

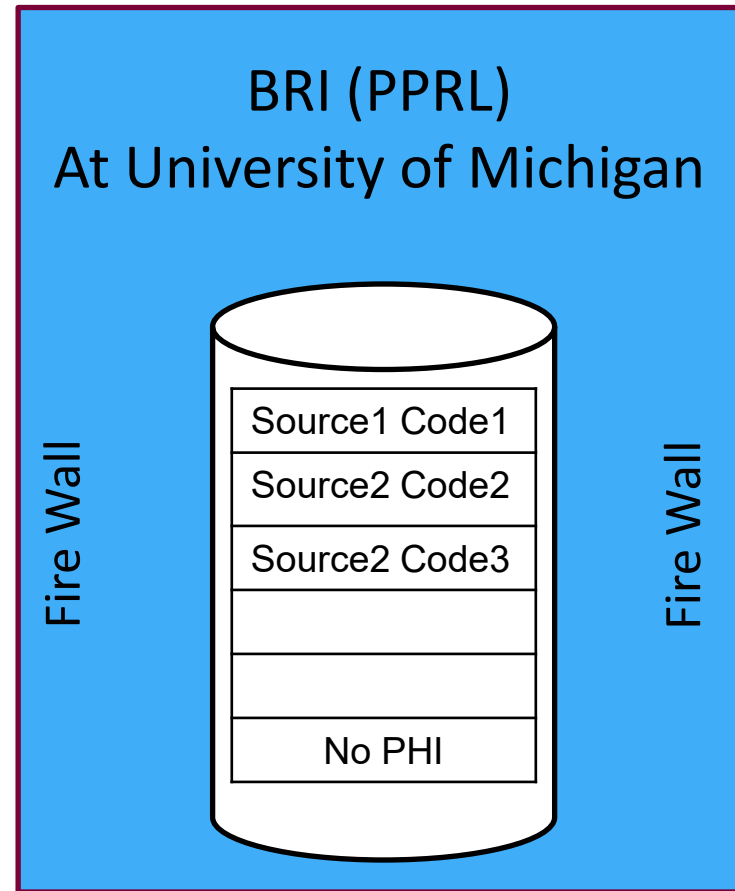


Secure Key



Multiple Hash

Hash	Code 1
Hash	Code 2
Hash	Code X



Example: Mortality Ascertainment

- Using Social Security Death Master File (DMF) (No longer available)
 - Treated DMF as a ‘data source’
 - Hashed each patient record in DMF since 2003
 - Kept Date of Death as plain text
 - Date of Death is publicly available
 - Need to keep unhashed for research purposes
- Each MPOG site’s patient records are also hashed
- Centralized Privacy Preserving Record Linkage (PPRL)
 - Previously referred to as Blinded Record Index (BRI)
 - Match clinical patient records against DMF patient records
 - Apply “Last Known Alive” check step to improve specificity

Study Goals

- Establish that Privacy Preserving Record Linking (PPRL) was
 - Feasible (computationally)
 - Economical (server quality / speed)
 - Accurate: acceptable match rates
- Evaluate the impact of
 - Varying ‘stringency’ levels on matching algorithm
 - Exact match of all (SSN, DOB, name)
 - Exact match of one and close match of others
 - Somewhere in between
 - Value of “Last Known Alive date” to improve specificity
 - Value of SSN

Methods

- 2400 patients across 2 distinct sites
 - University of Michigan
 - Suburban / rural patient population
 - 90% SSN fill rate
 - Thomas Jefferson University
 - Urban / metropolis patient population
 - 70% SSN fill rate
- Established “gold standard” dead and alive status
 - Manual review of patient records
 - Prospective patient registry
 - 30 day and 1 year mortality

Results

- Achievable
 - IRB approved each center's work
 - Standard commercially available servers
 - Commercially available software from 3rd party vendors
- 30 day mortality
 - Very Strict matching: Sens: 89.7% Spec:100%
 - Expanded matching: Sens: 93.3% Spec: 99.3%
- Impact of last known alive:
 - Worsens sensitivity
 - Improves specificity
- SSN is NOT necessary for high quality matching

What Next?

- Coordinating center
 - Signed contract with State of Michigan for Death Data
 - We will eventually apply Privacy Preserving Record Linkage to the state death data
- Each MPOG site
 - Work by your technical staff: < 1 minute
 - Work by your ACQR: < 5 minute
 - All software already installed and ready to go
- What we've learned
 - Do it in 2 phases: It takes a while 1 hour for 8,000 patients
 - Historical data first, then monthly data
 - It works

Historical Load

 MPOG Uploader

Upload | Status | Settings | Log

Case Selection

There are 403 cases that need to be PHI scrubbed.

☒ Cases awaiting upload

150679

☐ All cases

150679

☒ Specify Date Range

From  To 

Presets:

[Recommended](#)

[Previous Month](#)

[Historical](#)

Privacy Preserving Record Linkage

Note: You must be running a PPRL service in order to use this.

☒ Create/update the PPRL for these patients

☐ Update PPRL only (Do not upload case data)

Start Transfer

Monthly Load

 MPOG Uploader

Upload | Status | Settings | Log

Case Selection

There are 403 cases that need to be PHI scrubbed.

☒ Cases awaiting upload

150679

☐ All cases

150679

☒ Specify Date Range

From  To 

Presets:

[Recommended](#)

[Previous Month](#)

[Historical](#)

Privacy Preserving Record Linkage

Note: You must be running a PPRL service in order to use this.

☒ Create/update the PPRL for these patients

☐ Update PPRL only (Do not upload case data)

Start Transfer