A LifeCycle Model for Privacy Preserving Record Linkage

Bradley Malin
Vanderbilt University

Summarized by Brett Walenz
Duke University

Some slides from Able Kho (http://ohrpjuly15.illinoisstate.edu/agenda/Clinical_Data.pdf)

Background

Tasked with counting aggregate medical statistics about a population by using visit information
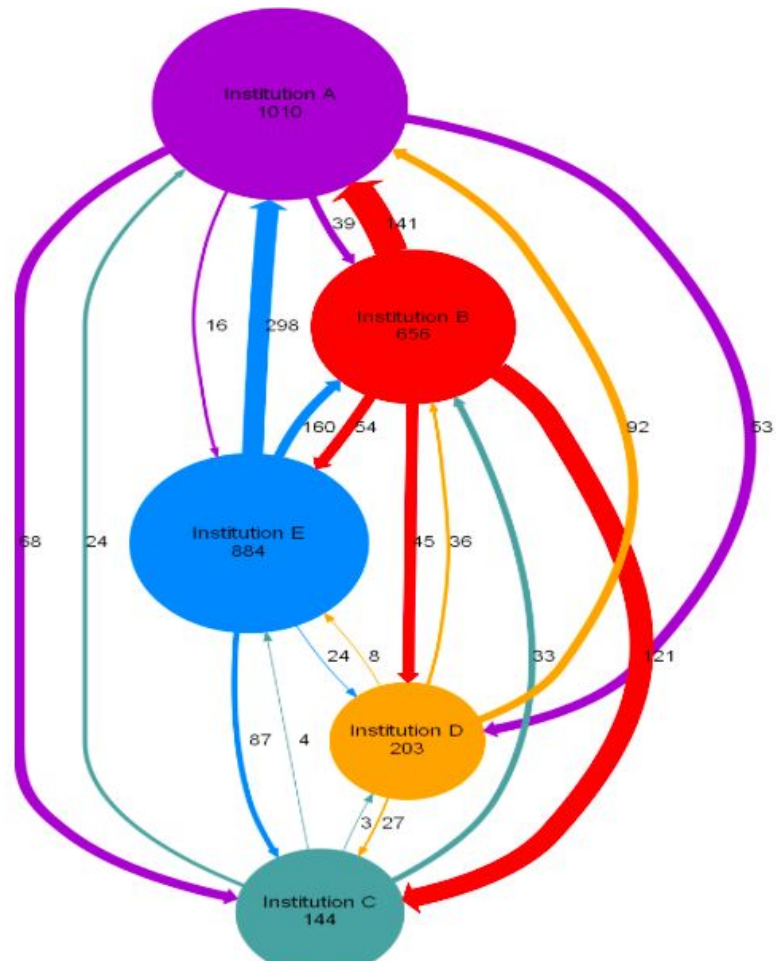
1. How many people have diabetes?
2. How many asthma events occur in a year?
3. 'Track' drugs on market for adverse effects

Problems:

1. No universal id for health (resisted in US)
2. Fragmentation of medical records
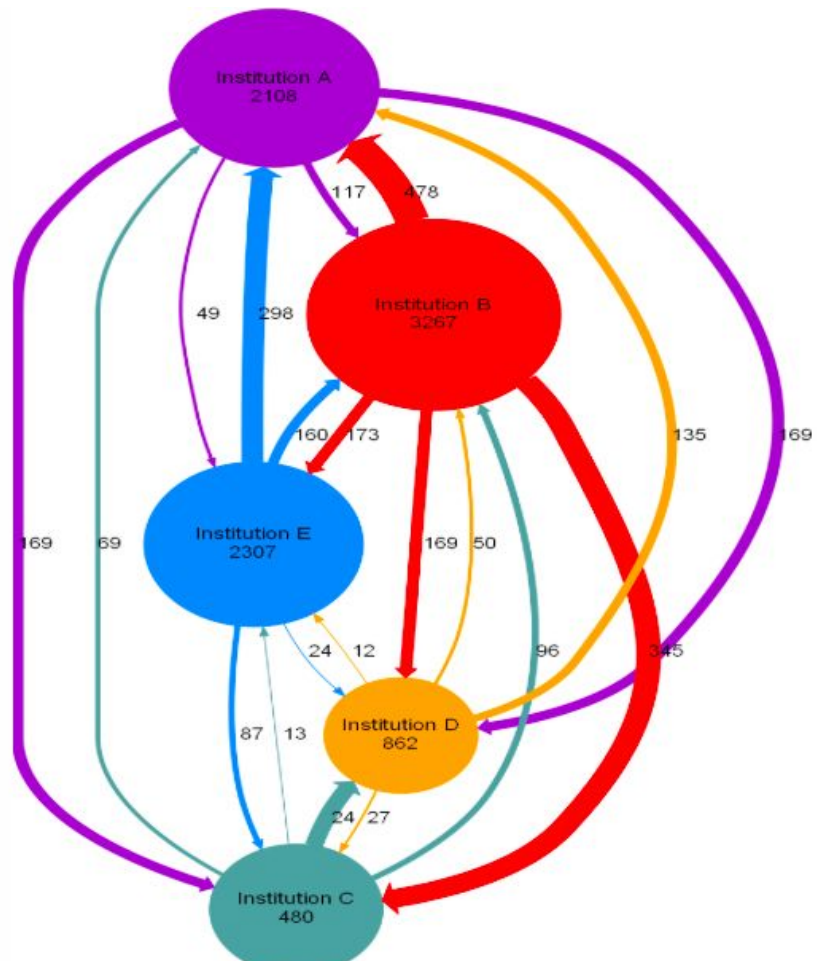
MRSA Admissions 2007-2010

VRE Admissions 2007-2010

Privacy Enhanced Record Linkage

Hash and Bash technique

1. Given two records A and B, hash each attribute in A and B using a cryptographic hash (SHA-1)

2. Compare for exact matches only

3. Aim for very high precision (specificity)

4. In a private setting, A and B can't know anything about the other. Need third parties to support matching, cryptography, and methodology to handle secure blocking to prevent leakage.

Privacy Enhanced Record Linkage

Requires two third-party brokers: key server and honest broker. Key server provides cryptographic keys, honest broker performs the actual matching.

Data appears to be high quality: most false positives comes from uncommon events, such as SSN changes and last name changes.

Blocking can be difficult under private settings: can't release exact information about size of blocks or contents of blocks.
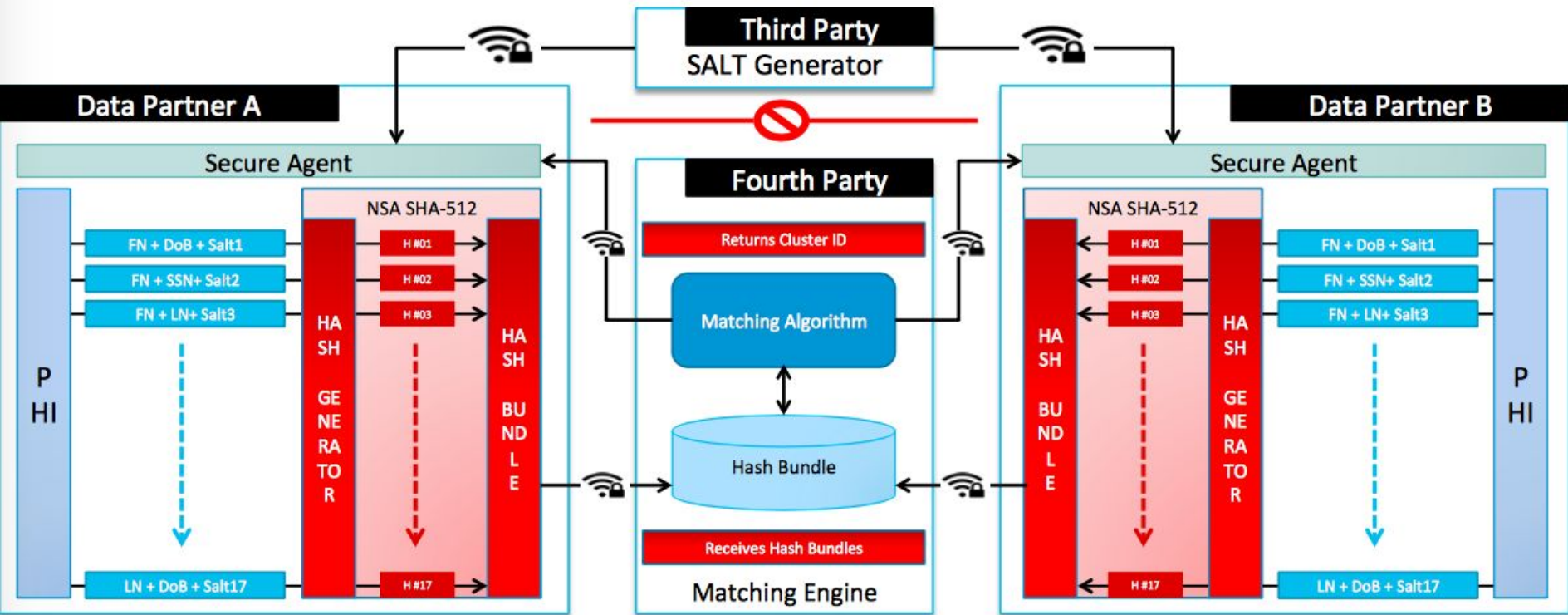
Reduction in counts with de-duplication for a sample of conditions

	Non Deduplicated	Deduplicated
Diabetes (Type II only)	**n=135,779**	**n=103,177** **24.0% reduction**
Asthma	**n=110,640**	**n=79,563** **28.0% reduction**
Myocardial Infarction	**n=6,049**	**n=5,384** **10.9% reduction**

Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, Humphries JE, Kominers SD, Hota BN, Sims SA, Malin BA, French DD, Walunas TL, Meltzer DO, Kaleba EO, Jones RC, Galanter WL. Design and Implementation of a Privacy Protecting Electronic Health Record Linkage Tool in Chicago. JAMIA 2015.

Privacy Preserving Record Linkage in Practice

1. Most existing software does not implement private record linkage(PRL), makes using in practice difficult

2. Extended OpenEMPI for PRL

3. http://www.openempi.org/

4. Secure Open Enterprise Master Patient Index - http://www.ncbi.nlm.nih.gov/pubmed/25954421

Key Server	Alice	Bob	DAN	PAM

1a — get salts

2a — message digest fields

1a — get salts

2b — message digest fields

3a — send encoded dataset

3b — send encoded dataset

4 — compute new parameters

5a — get new parameter advice

5b — get new parameter advice

6a — BF encoded fields

6b — BF encoded fields

7a — send encoded dataset

7b — send encoded dataset

8 — match record pairs