

Data Linkage: Within, Across, and Beyond PCORnet

Keith Marsolo, PhD

Instructor, Duke Department of Population Health Sciences

Co-Investigator, PCORnet Coordinating Center

Thomas W. Carton, PhD, MS

Chief Data Officer, LPHI

Principal Investigator, REACHnet



pcornet

The National Patient-Centered Clinical Research Network

Presentation goals

- 🌐 Describe PCORnet experience to date
 - Within and across network linkages
- 🌐 Outline a global PCORnet-wide approach
 - Full network linkage
- 🌐 Present some potential extensions
 - Beyond current PCORnet partners

Presentation outline

- 🌐 PCORnet 2.0
- 🌐 Introduction to hashed linkage
- 🌐 PCORnet linkage
 - Within
 - Across
 - Full
 - Beyond
- 🌐 Technology, governance, and use cases


Snapshot of PCORnet 2.0


9 Clinical Research Networks (CRNs)

- 47 DataMarts
- >65M patients with an encounter in the past 5 years
- >30M patients with an encounter in the past year

2 Health Plan Research Networks (HPRNs)

- 2 DataMarts
- >40M patients with an encounter in the past 5 years
- >20M patients with an encounter in the past year

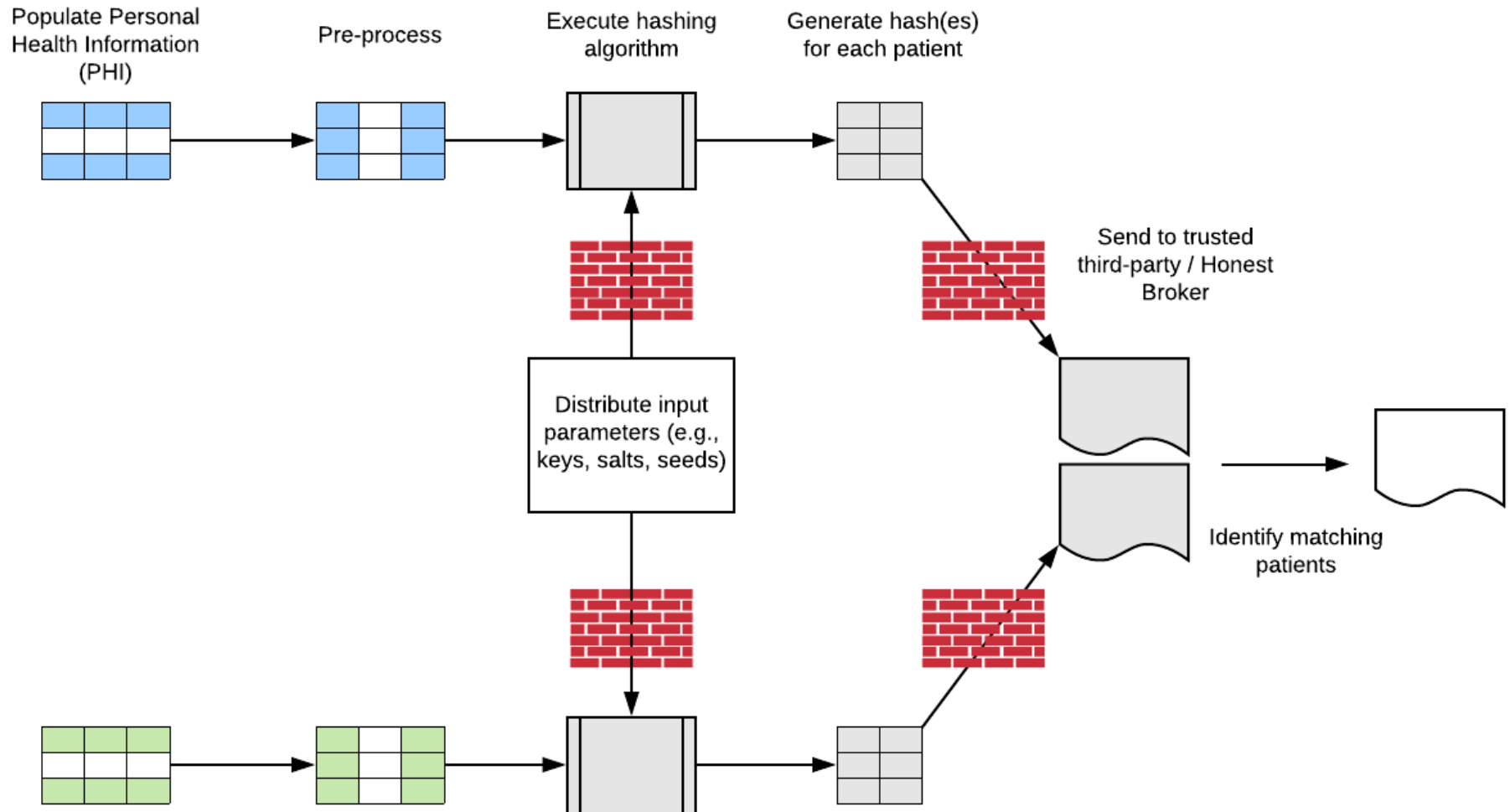
 The patient overlap between CRNs and HPRNs is unknown but expected to be high.

 The patient overlap between CRN DataMarts is unknown but expected to be low in most cases (except select markets).

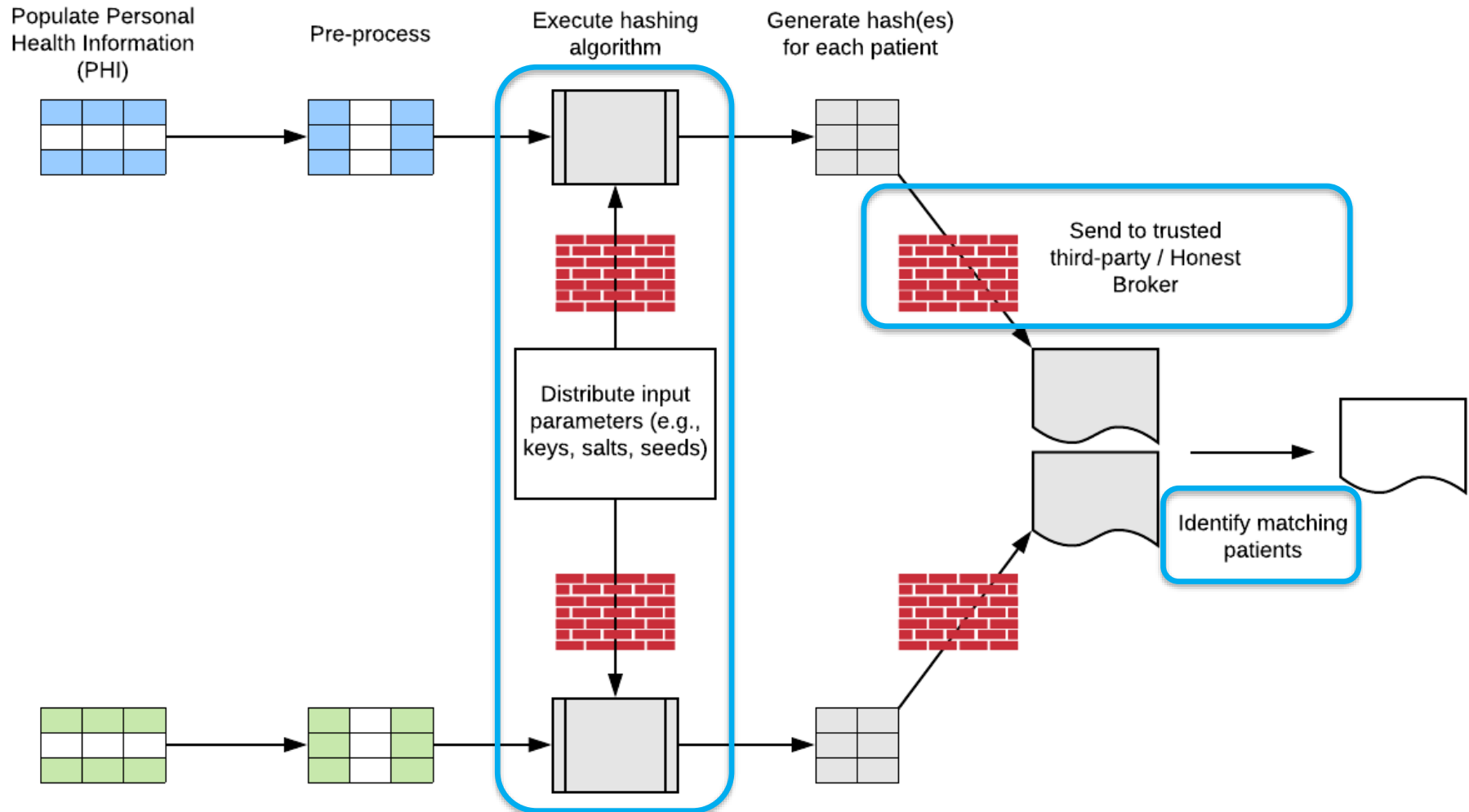
Introduction to hashed linkage: Terminology

- ❁ **Deterministic linkage** – two records match if all / some identifiers match above a specific threshold
- ❁ **Probabilistic linkage** – weights are assigned to each identifier & used to calculate probability that two records match
- ❁ **Privacy-preserving record linkage (PPRL)** – allows linkage across databases while preserving privacy of entities in them. Can be deterministic or probabilistic.
- ❁ **Trusted third party / honest broker** – a neutral third party that performs sensitive activities within a PPRL linkage method. Can also be achieved with technology.
- ❁ **Hashing algorithm / hash function** – used to convert an input string into an alpha-numeric string of fixed length (the hash). Two different strings should not generate the same hash.
- ❁ **Salt** – data appended to input of a hash function as protection against attack (e.g., storing passwords). In general, a random salt is used for every record. When linking, the same salt needs to be used across all databases.

Introduction to hashed linkage: General approach



Introduction to hashed linkage: General approach



Introduction to hashed linkage: Example uses

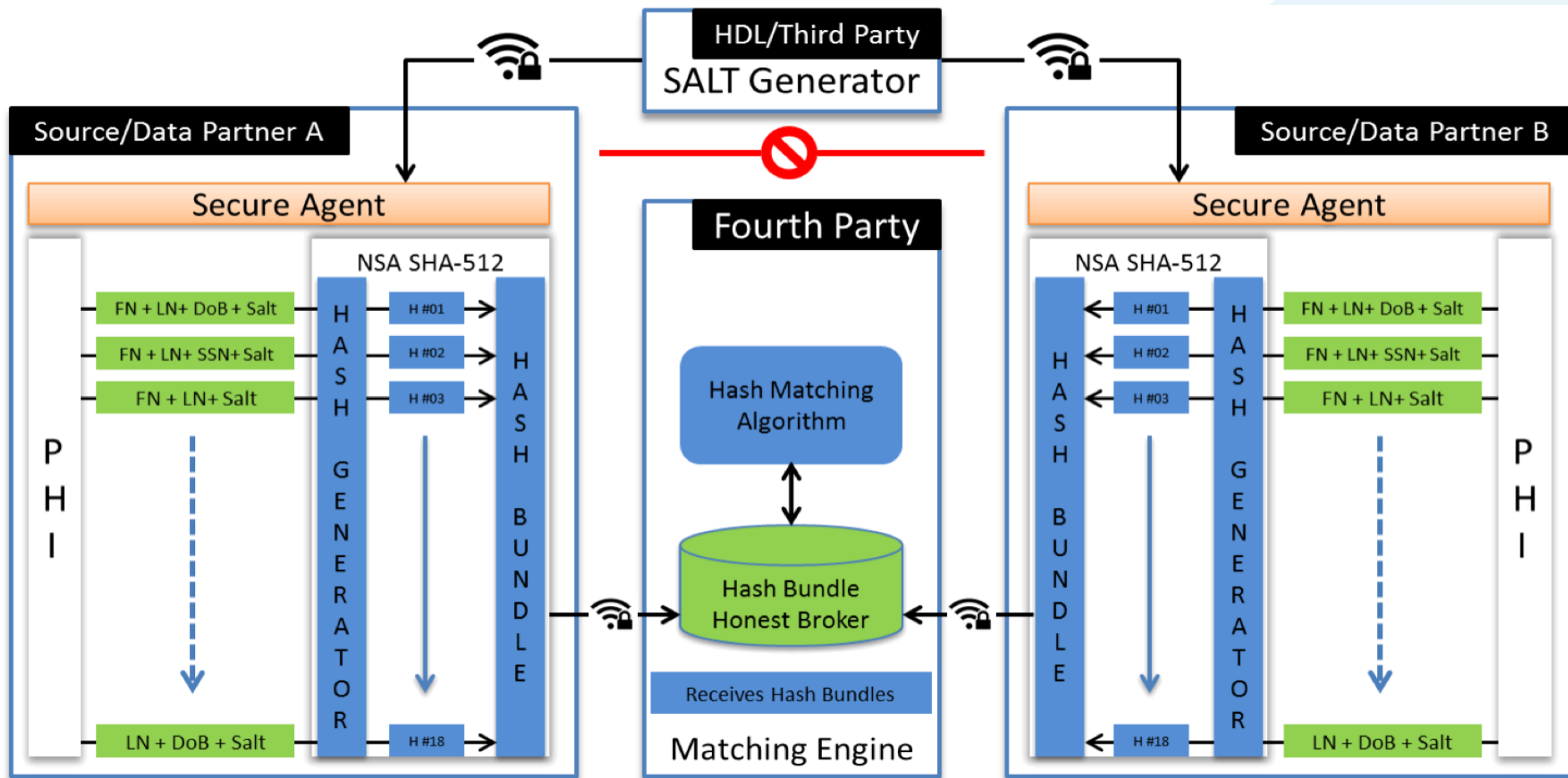
- 🌐 Link claims & EHR
 - Non-PCORnet example: All of Us
- 🌐 Link claims & claims
 - Western Australia & New South Wales
- 🌐 Identify overlap in rare-disease registries
 - Rare Diseases Registry Program (RaDaR) Global Unique Identifier (GUID) – utilizes National Database of Autism Research GUID program
- 🌐 Master Patient Index / Health Information Exchange

Within Network Linkage

Survey of within network approaches

Network	Method	Type	Proprietary	Hashing
CAPriCORN	GPID	Weighted deterministic	Licensed	Yes
INSIGHT	GPID	Deterministic and probabilistic	Licensed	No
MidSouth	PPRL	Deterministic	Open source	Yes
OneFlorida	De-Duper	Deterministic	Open source	Yes
PEDSnet	CURL	Deterministic, probabilistic, or both	Licensed	Yes
pSCANNER	Garbled circuit	Deterministic	Open source	No
REACHnet	GPID	Deterministic	Licensed	Yes

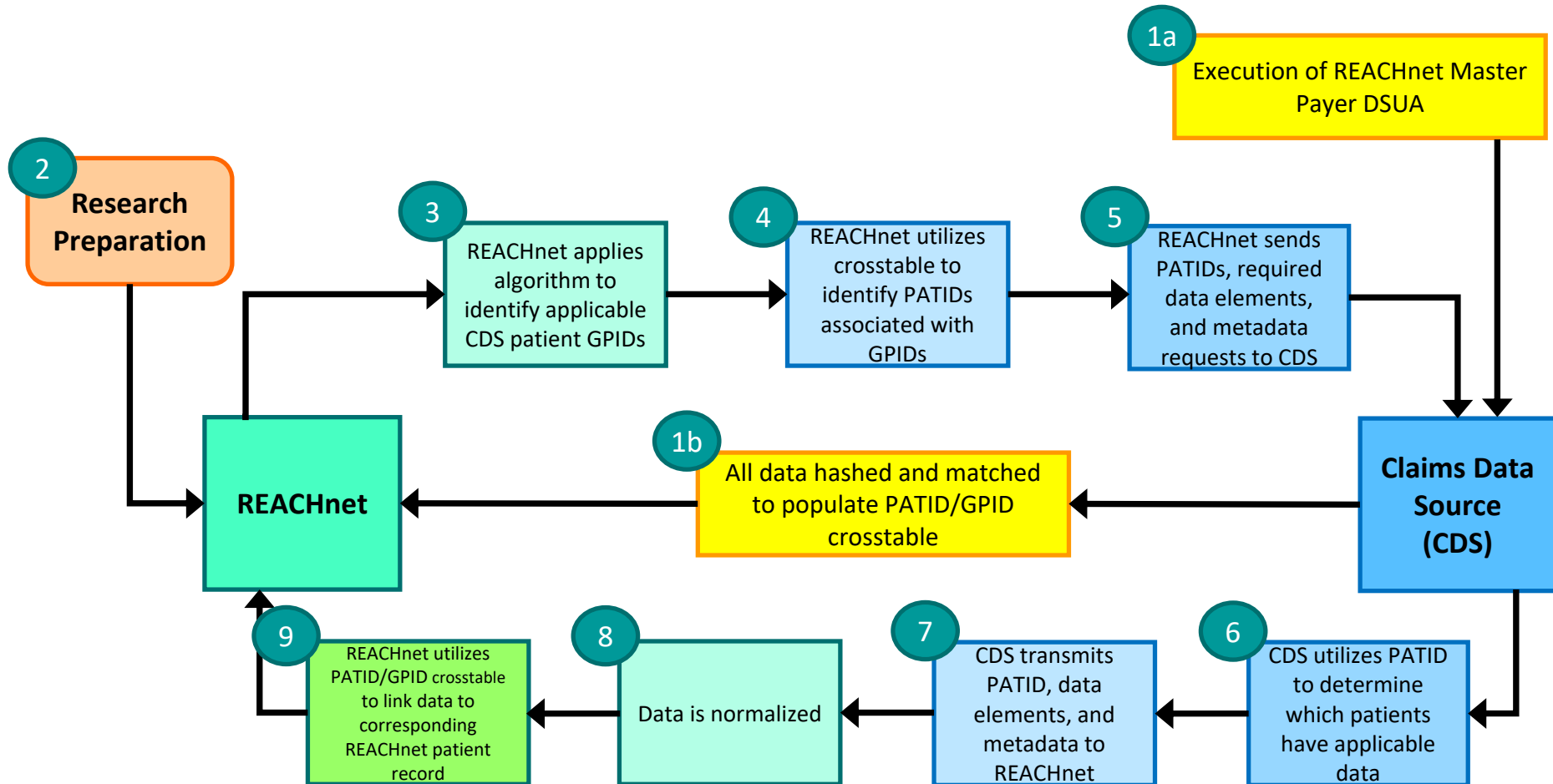
Within network example: REACHnet technology



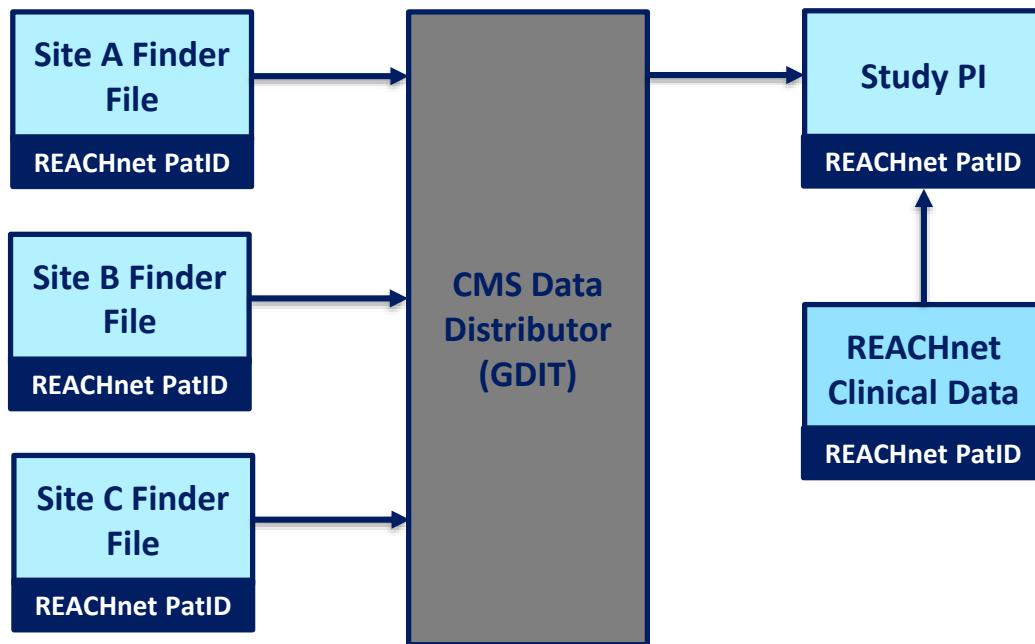
Within network example: REACHnet governance

- Site-level Common Data Model IRB
 - Governs systems sending hashes periodically with CDM elements to REACHnet Coordinating Center.
- Network-level Master Reliance Agreement (MRA)
 - Governs sharing of hashes for study specific use cases (under their own regulatory agreements).
- Network-level master payer data sharing and use agreement (DSUA)
 - Governs global hashing/matching to support specific research use cases (nested as amendments).

Within network example: Health plan linkage



Within network example: Medicare linkage



Requirements

1. Evidence of Funding Letter
2. IRB Common Rule and HIPAA Waiver Approvals
3. Part D Attestation
4. Research Methods
5. Research Identifiable File Cost Estimate/Invoice
6. Research Identifiable File Data Use Agreement
7. Research Identifiable File Executive Summary (including site-specific Data Management Plans)
8. Research Identifiable File Request Letter for New Study
9. Research Identifiable File Specifications Worksheet
10. Research Identifiable File Study Protocol
11. Submission of beneficiary finder files with the following data elements (as available): 1) Beneficiary IDs; 2) Health Insurance Claim Numbers; 3) SSNs; 4) Resident ID/State Code; 5) Unique Physician Identification Numbers; 6) National Provider Identifiers; 7) Employer Identification Number/Tax Identification Number.

Within network example: REACHnet use cases

- 🏥 GPID validation (clinical-to-clinical and clinical-to-claims)
 - *Current and Potential Effects of Cancer Screening on Health Outcomes*
- 🏥 Clinical-to-clinical linkages
 - *Real-world treatment patterns and outcomes of patients with T2DM*
 - *Real-world disease burden and treatment outcomes of patients with hyperkalemia*
 - *Louisiana Experiment Assessing Diabetes Outcomes*
- 🏥 Clinical-to-claims linkages
 - *T2DM Rapid Cycle Research Project (Tulane & BCBS)*
 - *PCORnet Antibiotics Study (Ochsner, Tulane & Humana)*
- 🏥 Clinical-to-Tumor Registry
 - *Investigating Social Determinants of Breast Cancer Disparities Using Cancer Registry and EHR Data*
 - *Social Determinants Role in Explaining Disparities in Hepatocellular Carcinoma*

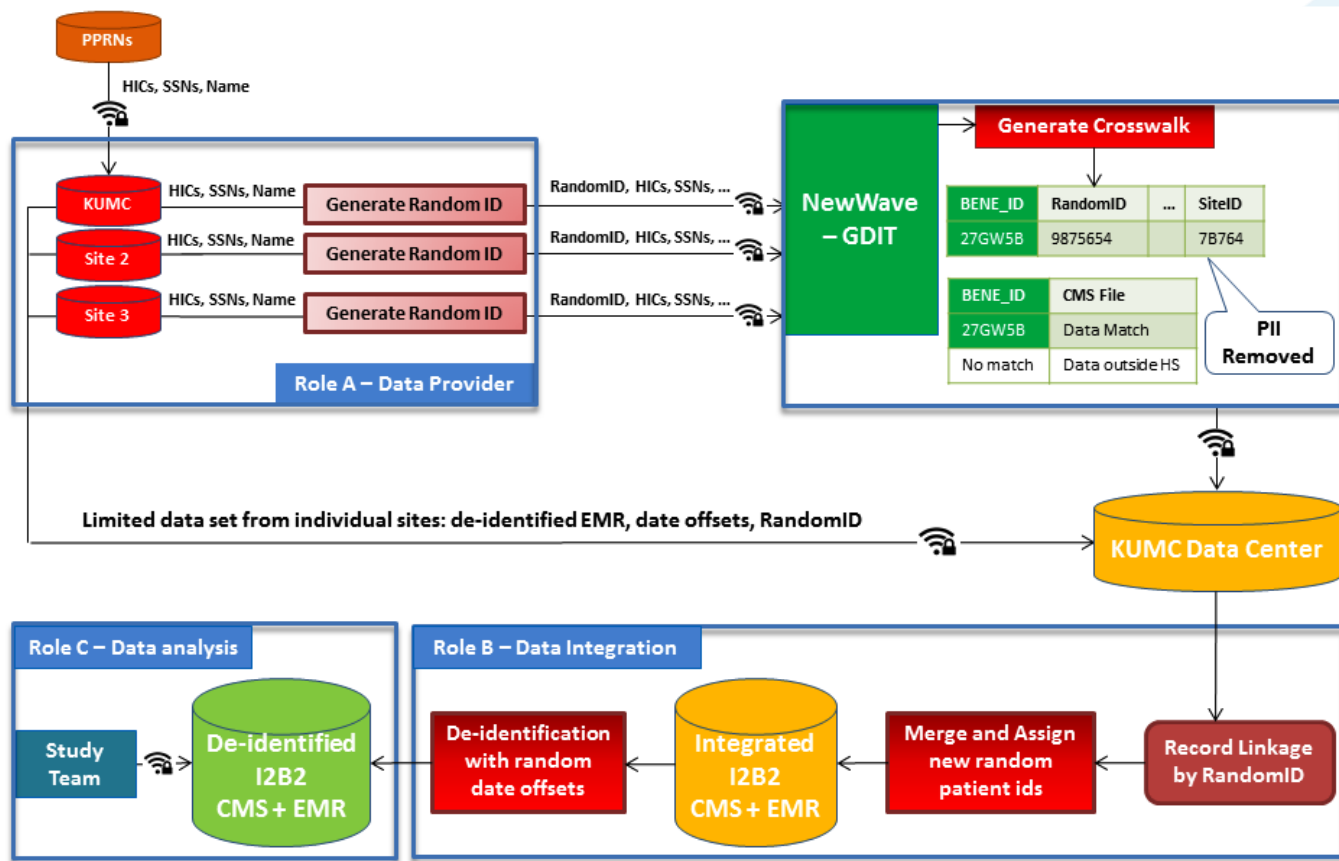
Research example: Cancer RCR

Aim 3. Completeness and Outcomes

- In a cohort of patients with first single breast cancer diagnosed during 2011-2015 with linked Medicare claims, assess the completeness of the EHR-derived data for identifying targeted therapy and molecular tests.

Slides courtesy of Mary Schroeder (Ulowa), Russ Waitman (KUMC), Betsy Chrischilles (Ulowa) and the RCR Project Team

Research example: Cancer RCR technology



Slides courtesy of Mary Schroeder (Ulowa), Russ Waitman (KUMC), Betsy Chrischilles (Ulowa) and the RCR Project Team

Research example: Cancer RCR governance

- 🌸 Executive Summary: Describes the project and initial team members
- 🌸 Study Protocol: Describes the specific analyses and types of data required to support those analyses
- 🌸 Data Use Agreement: Stipulates data elements, linkage, and use
- 🌸 Data Management Plan: Describes environment to conduct this research
- 🌸 Supplemental Data Security Analysis: Helps move the project forward with CMS and sites

Slides courtesy of Mary Schroeder (Ulowa), Russ Waitman (KUMC), Betsy Chrischilles (Ulowa) and the RCR Project Team

Across Network Linkage




Antibiotics demonstration study: Overview

- ❖ Purpose – determine the associations of antibiotic use with weight outcomes in a large national cohort of children
- ❖ Quantitative aims – assess the association between antibiotic (ABX) use before age 2 and childhood weight outcomes:
 - Weight outcomes at age 5 & 10
 - Childhood weight trajectories
 - Variation according to maternal variables (subset)
- ❖ Qualitative aim – parent focus groups & provider interviews on association between ABX & childhood obesity
- ❖ Published findings (Aim 1) – ABX use at <24 months associated with slightly higher body weight at 5 years of age

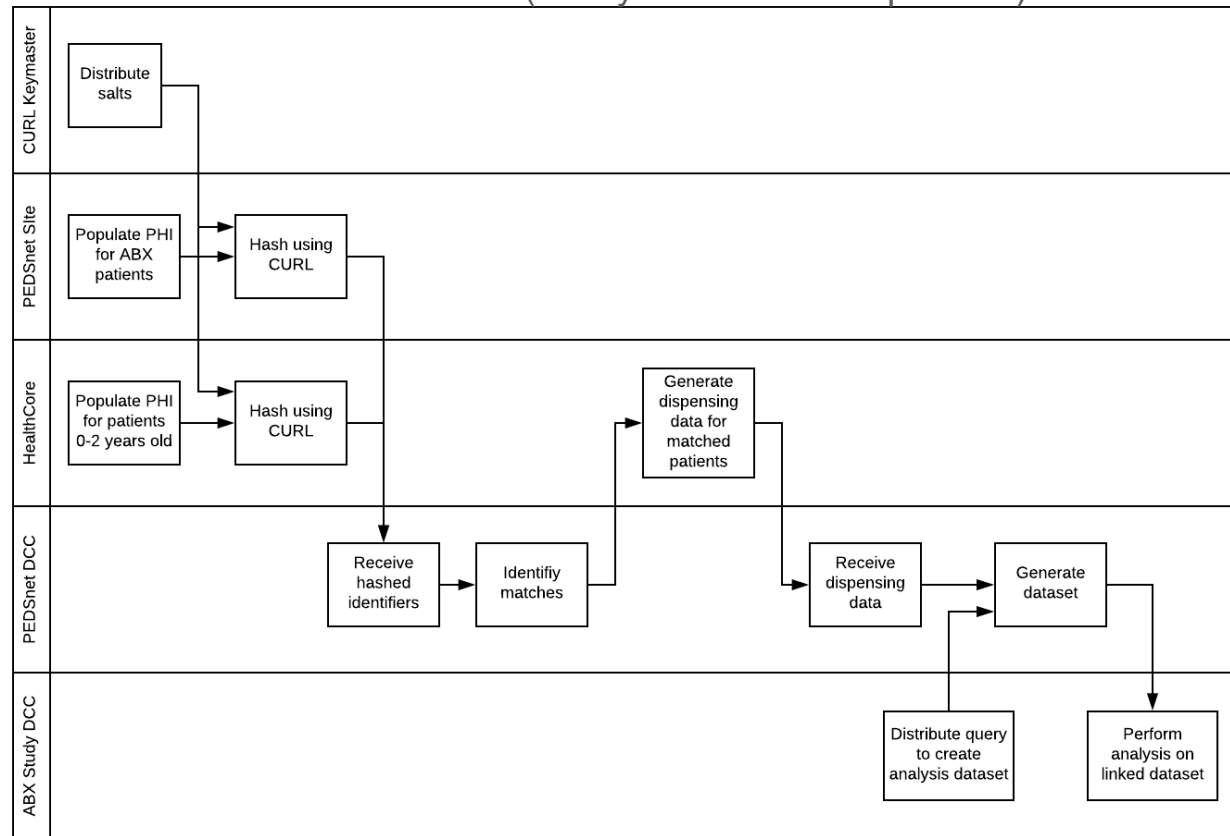
CDRN – Health Plan Linkage for ABX Study

- ⊕ Primary aim - Better capture of antibiotic exposure data before 24 months of age
- ⊕ Secondary aims
 - Develop technical process for linkage
 - Assess information gain
 - Extend prescribing – dispensing comparison
 - Potential added data on comorbidities
- ⊕ Linkage partners
 - PEDSnet/HealthCore
 - REACHnet/Humana

Across network example: PEDSnet/HealthCore technology

- 
 CURL (Colorado University Record Linkage) – developed by Toan Ong
- 
 Supports distributed & centralized linkage – centralized for this project
- 
 Publications on method forthcoming

Idealized data flow (reality was more complicated)



Across network example: PEDSnet/HealthCore governance

- 🏥 Data Use Agreements
 - PEDSnet members signed PEDSnet & PCORnet DUAs
 - PEDSnet – sharing between PEDSnet
 - PCORnet – sharing with PCORnet CC
 - HealthCore signed PEDSnet DUA & study-specific DUA with PCORnet CC

- 🏥 IRB – PEDSnet (CHOP as central IRB)
 - ABX study determined to be non-human subjects research (NHSR)
 - Use of linkage algorithm – NHSR
 - Linkage with HealthCore – NHSR

- 🏥 IRB – HealthCore
 - Had to submit local IRB – BAA with Anthem requires IRB approval with HIPAA waiver & DUA to release data

- 🏥 Editorial comment – NHSR determination may have actually slowed process

Within / across network summary

- 🌐 Networks selected the technology & governance they felt was most appropriate given their local context
- 🌐 Achieved local success, but lack of standardization has made it difficult to scale or rapidly execute new projects
 - If a health plan is linking with 5 networks, are they really expected to implement 5 methods?
 - Inconsistent governance means each new linkage discussion essentially starts from scratch
- 🌐 Recognition that a network-wide approach to linkage is needed
 - Networks can continue to utilize their local methods
 - Similar approach towards standardization as with the PCORnet Common Data Model and query tools

Full Network Linkage

Purposes

- ❖ Build distributed network linkage infrastructure (technology and governance)
 - For observational and population health surveillance research
 - Global agreement for the infrastructure
 - Scores of research use cases
- ❖ Classify the network
 - Overlap analysis
 - Number of unique patients
 - Table 1
- ❖ Support demonstration projects and RCRs
 - Antibiotics study
 - Opioid RCR
 - Scores of future use cases
- ❖ Develop the business model
 - Strong comparative advantage
 - Better, faster, cheaper technology and governance to link for specific projects
 - Scalable to other data sources

Governance

Global Linkage Workgroup

- Representatives from
 - Each CRN and HPRN
 - PCORI, Coordinating Center, PCRFB

CDM expansion

- Hash table

IRB

- Global agreement for the infrastructure
 - Update CDM IRB (one per network)
- Scores of research use cases
 - Individual study IRBs (one per study)

DSUA

- CDM expansion and study-specific use cases governed by current PCORnet DSUA v2.0 (so long as study results returned to Coordinating Center)

Technology

- ❁ Landscape analyses to inform
 - Technology
 - Most important attributes/metrics
 - Validation (formal validations)
 - Efficiency (time to implement)
 - Identifiers for linkage
 - Technical requirements for linkage (software requirements)
 - Proof of concept (real world implementation, peer review)
 - Governance
 - Agreements, partners, use cases
- ❁ Methodology
 - Develop RFP for hashing/matching solutions
 - Attribute list
 - RFP
 - Review process
 - Expect identified solution to provide salts/ hashes and support network implementation
- ❁ Queries
 - Develop a query that can be executed through PopMedNet by PCORnet Coordinating Center
 - Allow for linkage and de-duplication and replaces hashes with random patient IDs post-linkage

Beyond Network Linkage

Benefits of a scalable infrastructure

- 🌐 Reusable infrastructure
 - Global approach supporting scores of research use cases
- 🌐 Better, faster, cheaper linkage
 - Easy to add partners, data sources
- 🌐 Business model
 - Uniqueness of the asset

Potential extensions

- 🌐 Registries
 - E.g. Louisiana Tumor Registry
- 🌐 Commercial claims
 - E.g. Sentinel partners
- 🌐 Medicare claims
 - E.g. ResDAC
- 🌐 Patient reported outcomes
 - E.g. Patient Powered Research Networks

Questions / Discussion