ENGELBERG CENTER for Health Care Reform at BROOKINGS

Brookings Roundtable on Active Medical Product Surveillance

Some Initial Housekeeping

- To minimize feedback, please confirm that the microphone on your telephone is muted.
- To mute your phone, press the mute button or '*6'. (To unmute, press '*7' as well.)
- **There will be several opportunities for questions and discussion throughout today's session. Please use the Q&A tab at the top of your screen to submit your questions into the queue at any point and we will call upon you to state your question.**
- We will open up the lines for questions from those participating only by phone at the end of each Q&A session.
- Call the Brookings IT Help Desk at 202-797-6193 with technical problems.
- Thank you! We will be starting the webinar momentarily.

ENGELBERG CENTER for Health Care Reform at BROOKINGS

Learning from SCHIEx:
South Carolina's Statewide Distributed Data Integration

Sue Veer, Chair, Lakelands Rural Health Network
David Patterson, Chief, Health & Demographics, ORS, SC B&CB
Vik Kheterpal, Principal, CareEvolution, Inc.

March 15, 2010

Ensuring Patient Privacy While Meeting Public Health Needs

What approaches are available for protecti… …ile also ensuring that the potential public h… … fully realized?

- Initially, little or no transfer of pers… evaluate many safety questions.
- For the near term, what are t… …ssure data privacy in compliance…
- As Sentinel becomes more sophisticated, there may be benefits to confirming results through chart reviews and/or linking databases (eg, claims, EHRs, registries) at the patient level.
- Is it possible to protect patient privacy while linking datasets and validating findings from Sentinel?

Jan 11, 2010 Sentinel Workshop
Dr. McClellan's Opening Remarks

Can We Link Datasets While Protecting Patient Privacy?

SCHIEx: Different Objectives But Similar Privacy Challenges as Sentinel

- To provide for the timely and accurate sharing of electronic patient information to improve the quality and efficiency of care.
- To allow for improved patient education, disease management, and patient outcomes.
- To allow for enhanced population health improvement, assessment, and management.
- To empower patients with a valuable, standards-based, transferable and longitudinal personal health record.

SCHIEx : An Exchange Infrastructure to Provide A Low Cost Public Utility For Subscribers

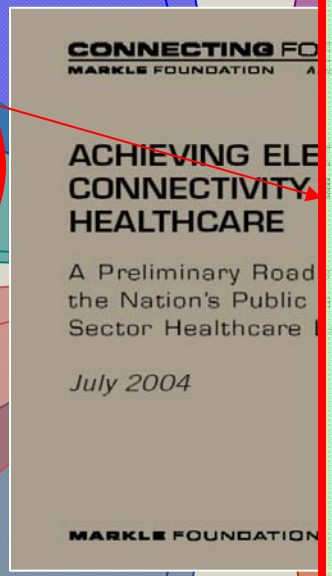
- Conceived as a "Public Good", SCHIEx was initiated in 2006, with production pilots in 2007 and production use in 2008
- "Standards based" and designed around privacy, security and ease of use
- Costs kept low and value added by leveraging extensive data holding and technical expertise existing within SC B&CB's Office of Research and Statistics
- Development based on a "practically focused" collaboration between public, private, and non-profit partners that included health care providers from the very beginning

…Good Intentions Not Enough

Had to Engage Stakeholders and Reassure Citizens

Designed for Privacy **and** Data Sharing

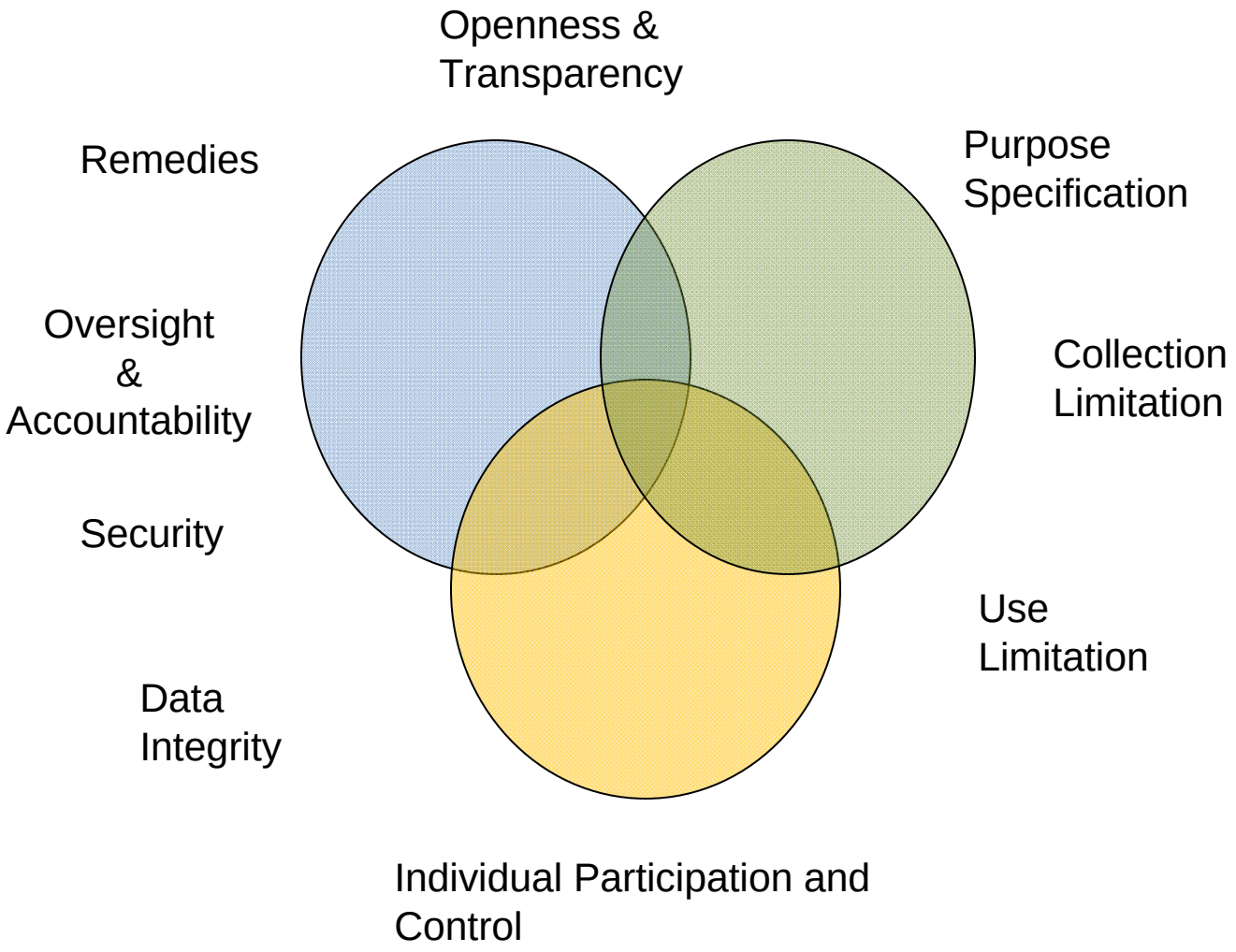
1. Engaging the American Public
2. Designing for Privacy & Security
3. The Infrastructure—technical architecture & approach
4. Accurate Linking of Patient Information
5. Data Standards
6. Clinical Applications
7. Funding & Incentives
8. Legal Safe Harbors



Approach

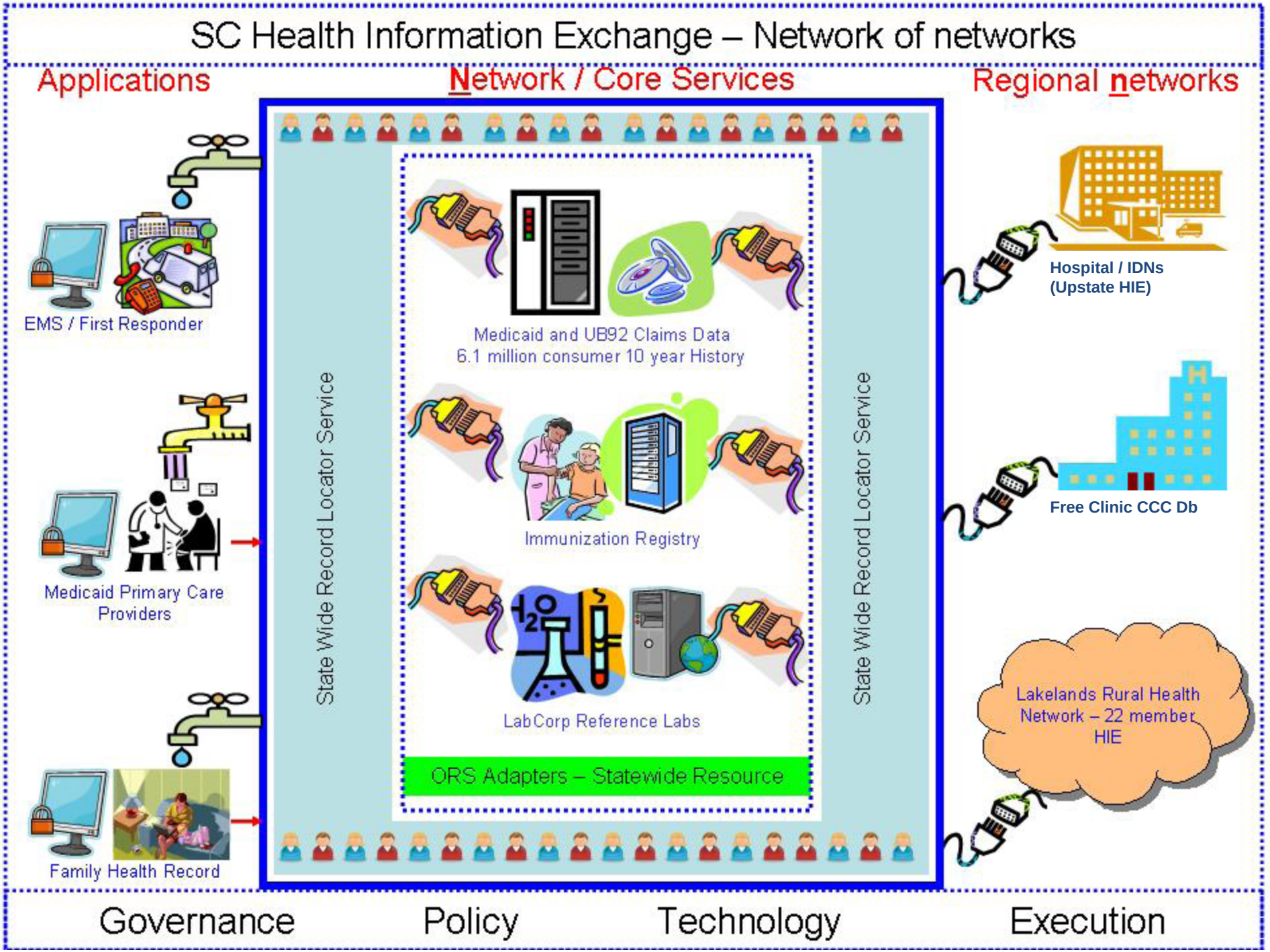
- The Basics
  - Separate Clin Data from Dem
  - Peer to Peer Data Exchng
  - No Centralized CDR
  - NHIN Stdards Compliance
- No Aggregated Central Store

The Privacy Principles Intersect To Create Core Operating Principles



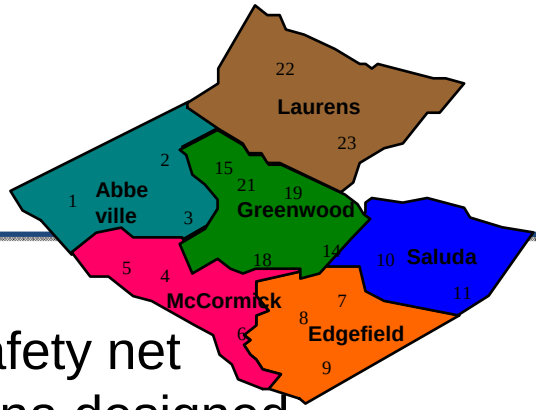
Core Operating Principles…Analogous to Sentinel

- The HIE is merely a new modality for exchanging information that is already being shared in less efficient ways, with greater privacy and security protections.
- HIE does not own or store any medical data – ownership rests with the source record where that data lives.
- The HIE governing body creates the minimum standards for participation - responsibility and accountability for compliance is continually pushed out to the user/provider organizations.
- The principle of individual participation and control creates opportunities for “holes” in the information – i.e. the HIE is not the patient’s complete medical record; therefore
- Use of the HIE does not alter the provider’s responsibility for thorough evaluation and appropriate clinical decision making.
- The HIE can be used to enable clinical decision making at the individual patient level as well as at the population level.


SC Health Information Exchange – Network of networks
Applications
Network / Core Services
Regional networks
EMS / First Responder
Medicaid Primary Care Providers
Family Health Record
State Wide Record Locator Service
Medicaid and UB92 Claims Data
6.1 million consumer 10 year History
Immunization Registry
LabCorp Reference Labs
ORS Adapters – Statewide Resource
State Wide Record Locator Service
Hospital / IDNs (Upstate HIE)
Free Clinic CCC Db
Lakelands Rural Health Network – 22 member HIE
Governance
Policy
Technology
Execution

Lakelands

Rural Health Network



A six-county partnership of key players and safety net providers in the Lakelands area of South Carolina designed to foster collaboration and strengthen the health care delivery system as a whole.

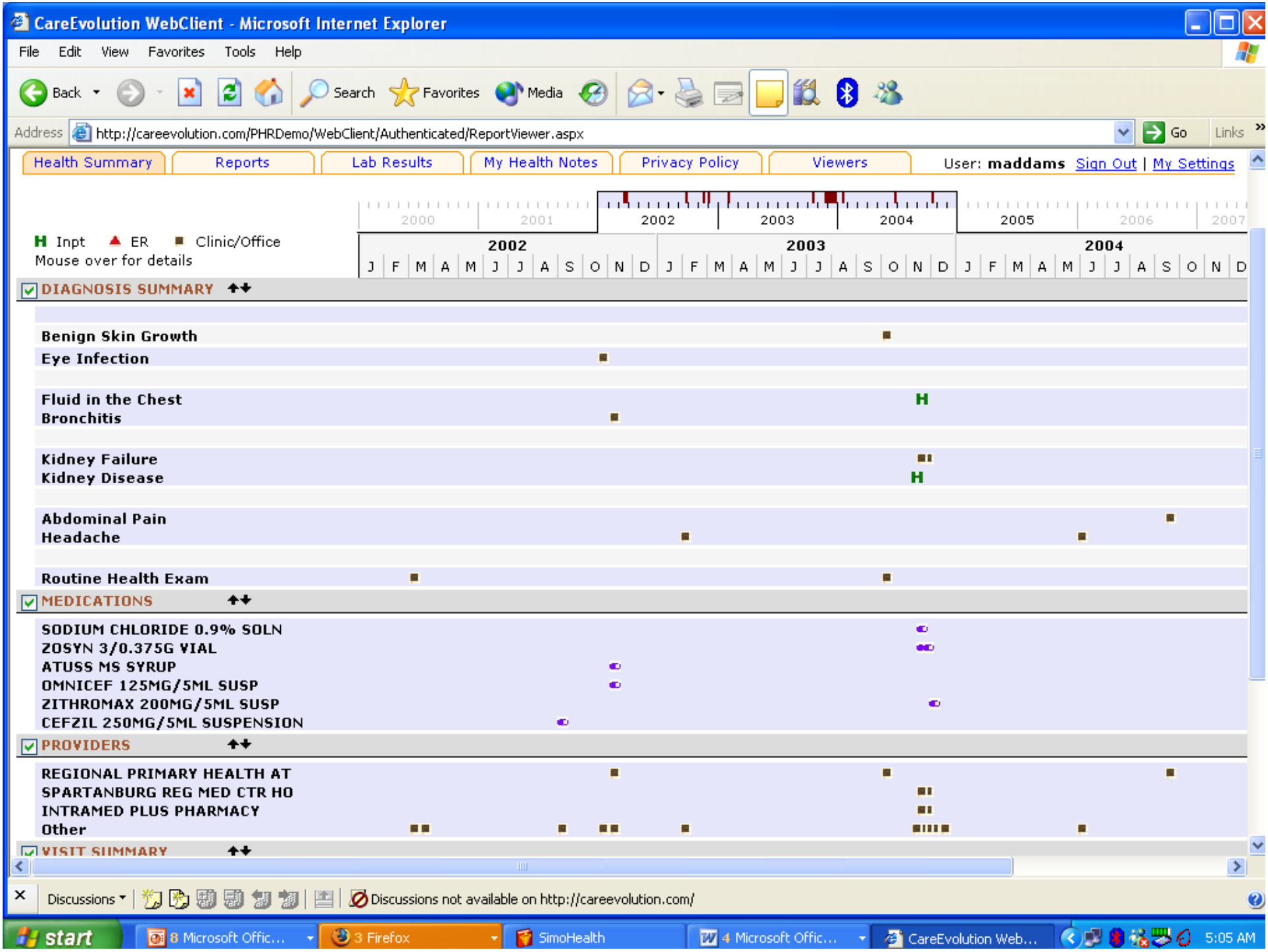
Mission:

"To develop a collaborative, economically viable health network to improve the quality of care in our service area."

The network was conceived and designed to foster collaboration and strengthen the area health care system as a whole.

… LRHN a SCHIEx "subscriber"

Gains Broad/Shallow Claims Data … Contributes EMR Data


CareEvolution WebClient - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://careevolution.com/PHRDemo/WebClient/Authenticated/ReportViewer.aspx
Go Links
Health Summary
Reports
Lab Results
My Health Notes
Privacy Policy
Viewers
User: maddams Sign Out | My Settings
2000 2001 2002 2003 2004 2005 2006 2007
H Inpt ▲ ER ■ Clinic/Office
Mouse over for details
2002 2003 2004
J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D
DIAGNOSIS SUMMARY
Benign Skin Growth
Eye Infection
Fluid in the Chest
Bronchitis
Kidney Failure
Kidney Disease
Abdominal Pain
Headache
Routine Health Exam
MEDICATIONS
SODIUM CHLORIDE 0.9% SOLN
ZOSYN 3/0.375G VIAL
ATUSS MS SYRUP
OMNICEF 125MG/5ML SUSP
ZITHROMAX 200MG/5ML SUSP
CEFZIL 250MG/5ML SUSPENSION
PROVIDERS
REGIONAL PRIMARY HEALTH AT
SPARTANBURG REG MED CTR HO
INTRAMED PLUS PHARMACY
Other
VISIT SUMMARY
Discussions
Discussions not available on http://careevolution.com/
start
8 Microsoft Offic...
3 Firefox
SimoHealth
4 Microsoft Offic...
CareEvolution Web...
5:05 AM

SCHIEx Federated Architecture

David Patterson
Section Chief, Health & Demographics
Office of Research and Statistics
SC Budget & Control Board

Data Strategy – Claims Data to Jump Start … Clinical Data Supplements Later

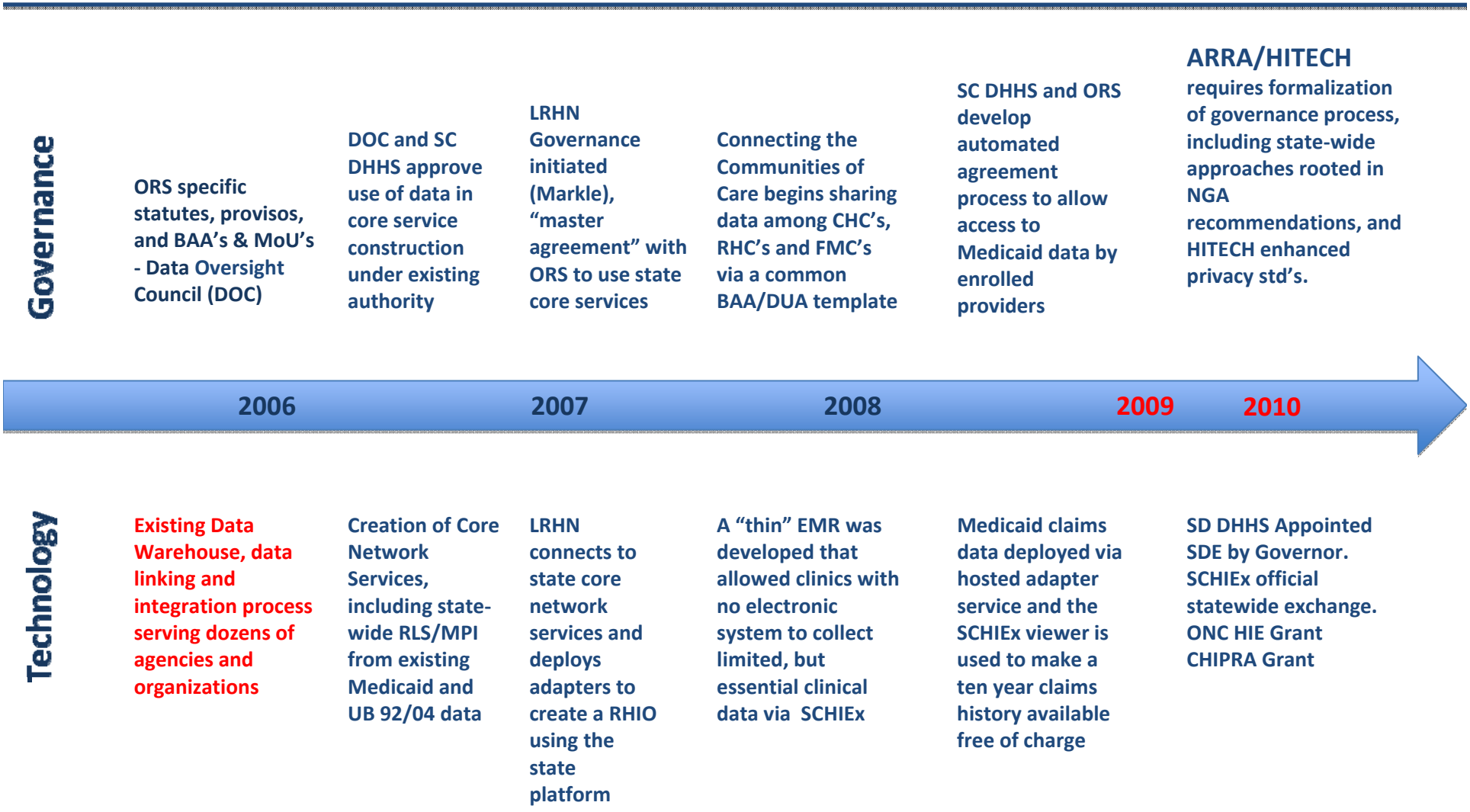
- Leverage Existing Claims Data (10 years, 6.1 million consumers)
  - AllMedicaid (1.1 million subscribers)
  - All UB92 (state statute requiring all acute care, ER, urgent care, and ambulatory surgery centers)
  - Lots of data pulled together to get clinician buy-in

- Use it to Jumpstart the HIE effort in SC
  - Statewide record locator service (6.1 million consumers) covering well over 95% of the state population
  - Use Claims data to create a longitudinal health record summary for citizens
  - Then, combine with clinical data sources (hospital EMRs, ambulatory EMRs)

- Build it and See if Anyone Comes… Follow the emerging national standards but put them in practice to iterate through various cycles

- Yogi Berra…. "In theory, theory and practice are the same but…. In practice, they are different"

SCHIEx: Leveraged Claims Data Holdings While Ensuring Could Link to "Deeper" Data Later



Separate Information Linking Patients From Clinical Information – 2 Phased Query

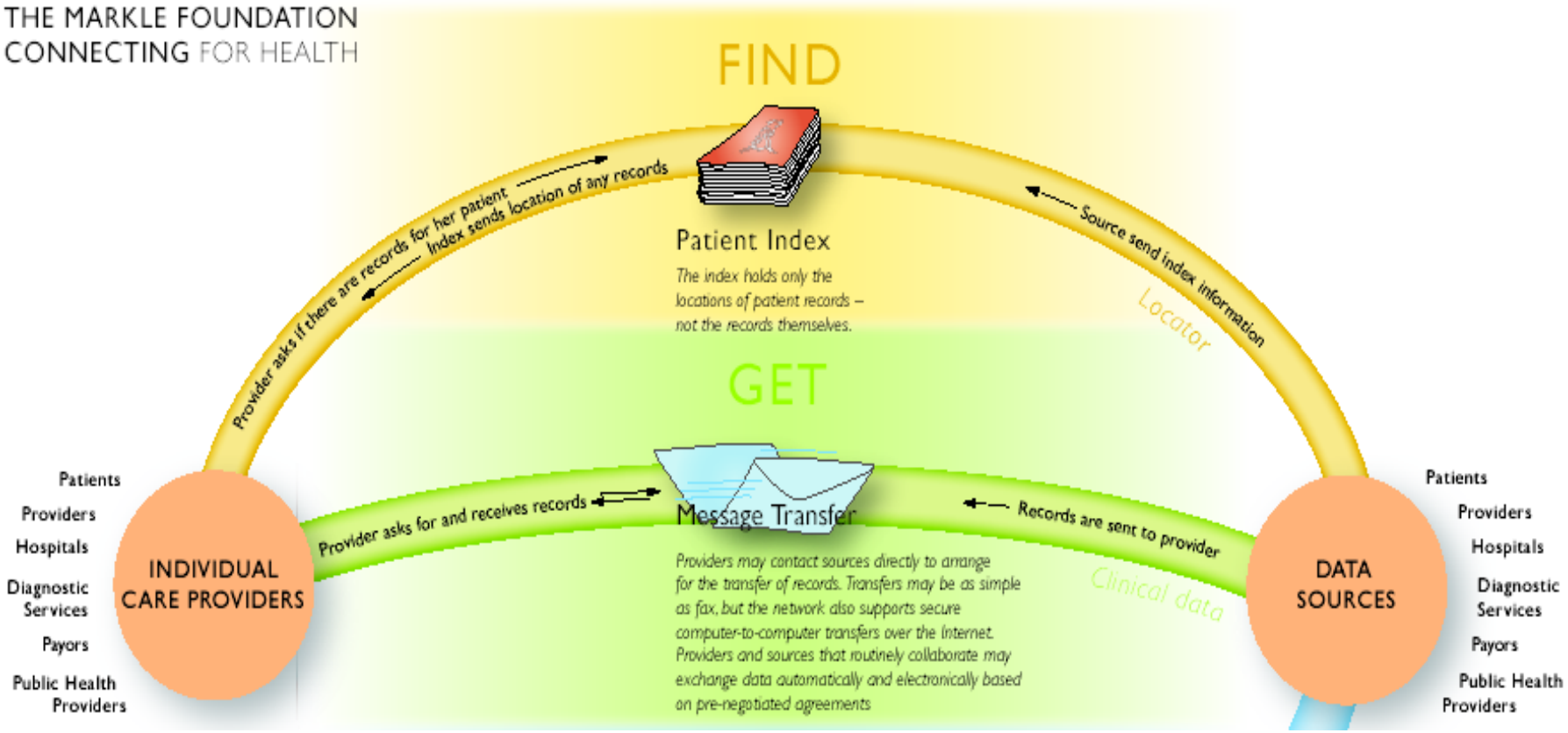
Step 1. “Where are records for Patient X”

Step 2. “How can I get Them?”

Follow CfH, HITSP/CCHIT, & NHIN Standards – so the **n**etwork will be interoperable, privacy protecting, and scalable…for other uses

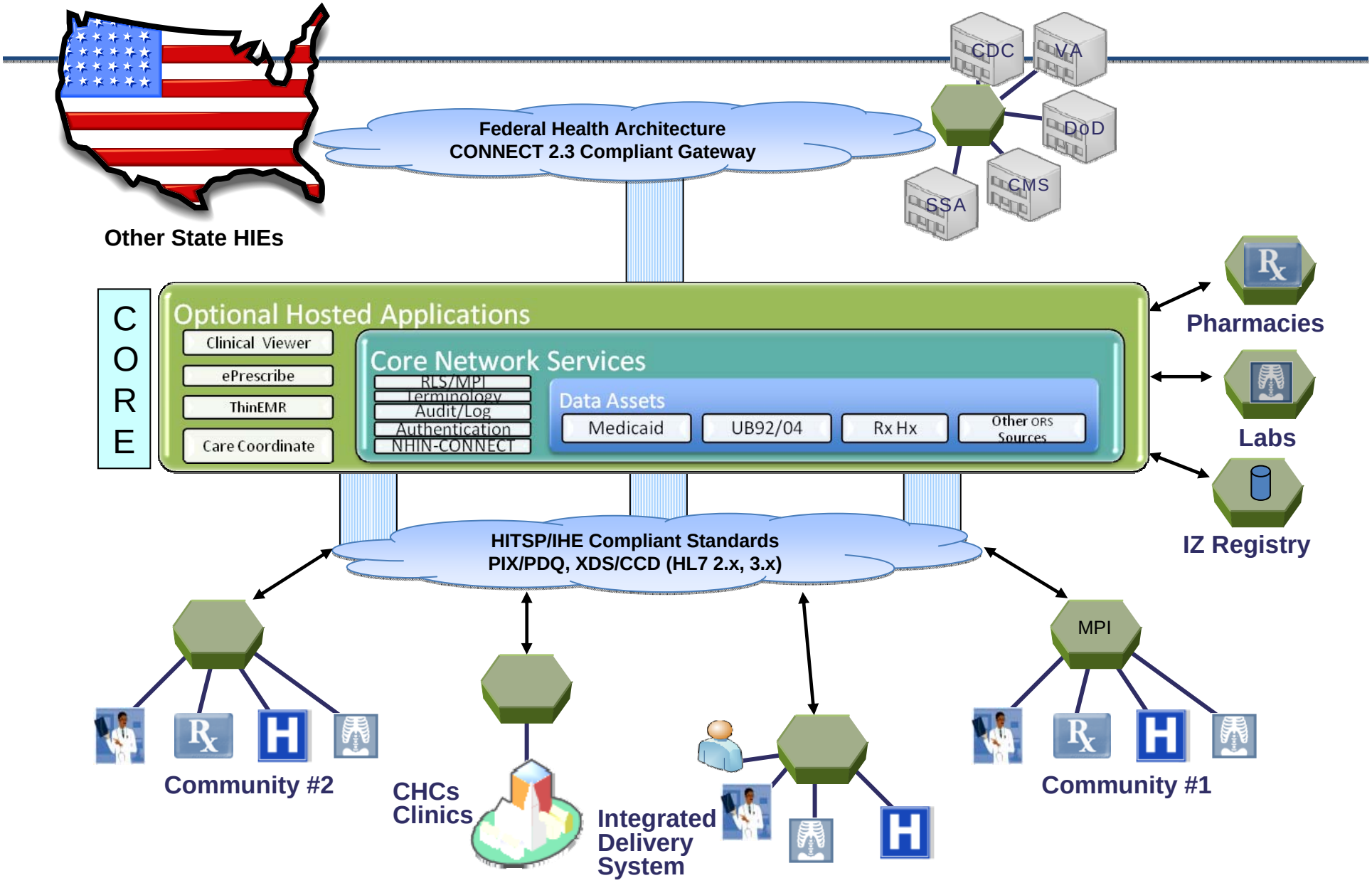
Courtesy: Connecting for Health
Markle Foundation

Step 1 : Patient Linking – Centralized
Step 2 : Clinical Information - Federated



Courtesy: Connecting for Health
Markle Foundation

Core : Trusted Linking For Diverse Adopters



The Problem: Centralized RLS (MPI) Critical Weak Link for Privacy

1. Data aggregation increases value to attackers – one stop … all citizens
2. Large number of entities with legitimate need to access the RLS… increases vulnerability
3. “discoverability” of information by government agencies
4. Threat from within
5. Fostering “trust” amongst competing provider entities.

Solution: “Blindfolded” Record Linking – Link Patients Without *knowing* the Demographics

1. Just as the Bank does not know the contents of the safety deposit box, Crypto-RLS™ can provide linking without ever “KNOWING” the contents
2. Use a “likeness” of patient demographics not the actual demographics (a SHA hash based “fingerprint”)
3. No risk in managing entire regional population
4. No clinical data centralization
5. Protects from
   - Internal threats
   - Disgruntled employees
   - External hacks
   - Inadvertent loss (theft, backup distribution)

How Do We Know It Works?

Note: The statistical analysis was prepared using a large-sample confidence interval with a degree of confidence of 99%. The formula used in calculating the lower and upper bounds is given below.

$$\frac{x}{n} \pm 2.58 \sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

Unique ID Sampling Results Using 99% Confidence Interval

Summary for All Sources

Sample ID	Sample Description	Percent of Correct Matches	Percent of Incorrect Matches
221	VRB (pq3406)		
222	DAODAS (Jul 2007)		
223	LFAF (2006)		
224	DOEPRECODE (07)		
225	DOEPACT 2006-2007		
226	ALZREG (ctlc0506)		
227	DMH (0807)		
228	HL (Sep 2007)		
229	MCAIDRECIP (117)		
230	SHP (Sep 2007 Eligibility)		

Summary graphs

- Weekly Statistical Sample Manually Reviewed
- Statistical Process compared with Clinical Process ; clinical RLS has higher specificity
- LRHN conducted prospective audit of 100,000 records with known links and known links
- No reported false positive in 3 years of production use.

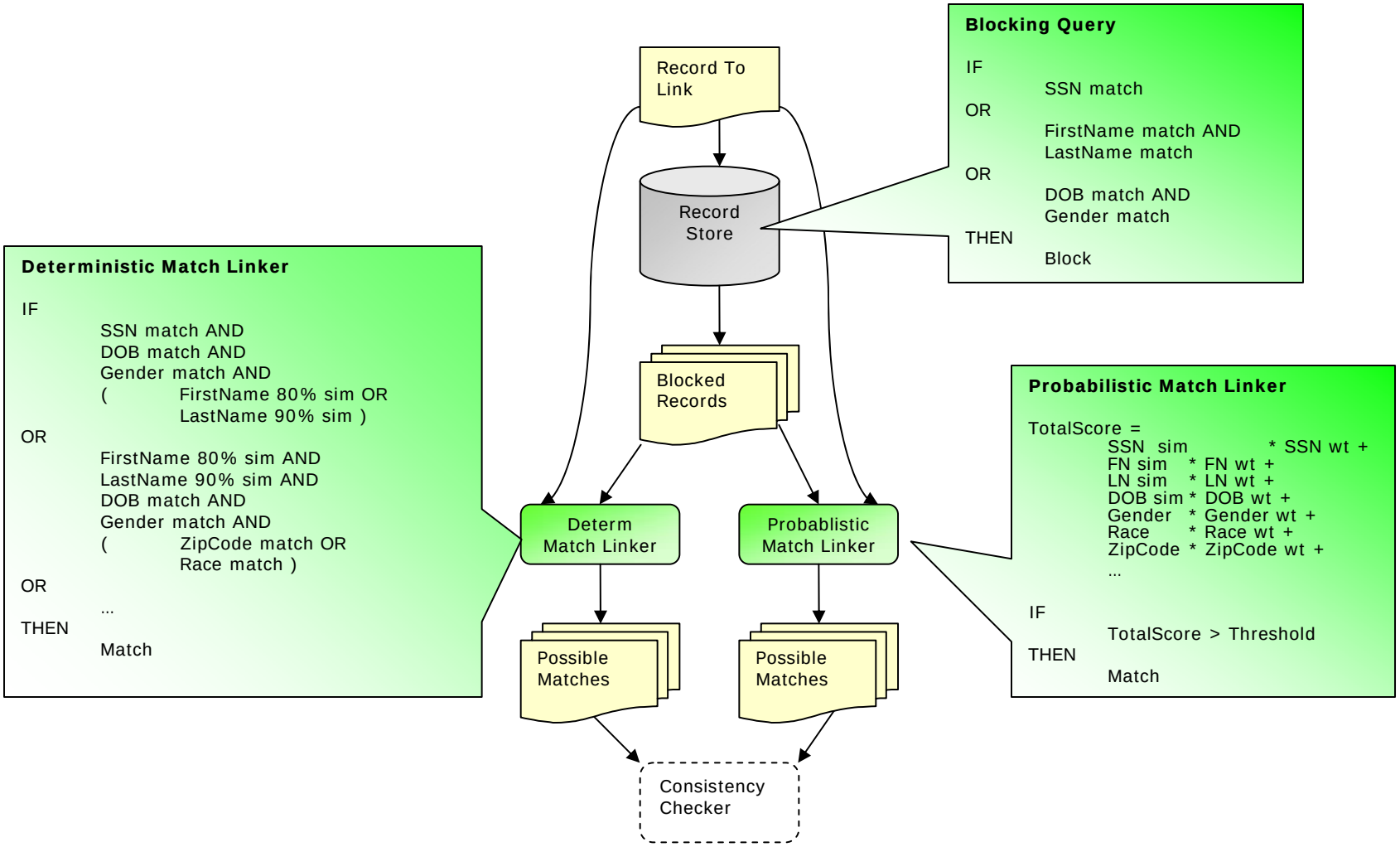
Vignettes of Innovations Enabled by SCHIEx Architecture – All Use the Trusted Core

- LRHN – HRSA FLEX grant winner to improve diabetes care across population
- CCC – Linking all Free Clinics in State together
- Medicaid Claims EHR – more than 4000 providers eligible to see a claims based history
- CHIPRA Grant Award - one of 10 CHIPRA Innovation Grant winners just recently
- AccessNET – HRSA Patient Navigation demo. One of 6 in country
- Telepsychiatry Initiative – linking 65 SC EDs to central Psych consulting hub

Blindfolded Record Linking

Vik Kheterpal
Principal, CareEvolution, Inc.

Start with the Best Linking Algorithm Known in Industry… SAMSHA/Census Bureau



Use Cryptographic Techniques Like Hashing to Obfuscate The PHI…

" The **SHA hash functions** are a set of cryptographic hash functions designed by the National Security Agency (NSA) and published by the NIST as a U.S. Federal Information Processing Standard. SHA stands for **Secure Hash** Algorithm." (wikipedia)

"The four hash algorithms specified in this standard are called secure because, for a given algorithm, it is computationally infeasible 1) to find a message that corresponds to a given message digest, or 2) to find two different messages that produce the same message digest" NIST FIPS140-2 Standard Published 2003.

How To Perform Effective "linking" and hashing – Generate Permutations Prior to Hashing

- Misspellings would generate completely different "hashes" or fingerprints
- Different permutations depending on type of identifier
  - Names: Bigrams, Nicknames, NYSIS codes, etc
    - Bigram: All subsets of the set of 2 consecutive characters in a string.
    - Example: Pete -> {"pe", "et", "te"}, {"pe", "et"}, {"pe", "te"}, {"et", "te"}, {"pe"}, {"et"}, {"te"}
  - Numeric: Transpositions, Off-By-One, etc
  - Date: Month-day swap, Off-by-one, etc
- Permutations provide ability to partial-match identifiers even though we're blinded.

Blindfolded Record Linking Steps (CryptoRLS™)

Within Each Federated Site

PHI at participating institutions with identifiers (First Name, Last Name, SSN, DOB, …) is processed by a "local" service as follows

1)Standardizer Pass (delimiters, junk values, spaces)

2)Create permutations of identifiers with associated similarity score (NYSIS, Soundex, transpositions, Nicknames, …)

3)One-way hash the permutations

4)Send hashed permutations to the RLS (record locator service)

5)RLS matches received set of hash permutations with others

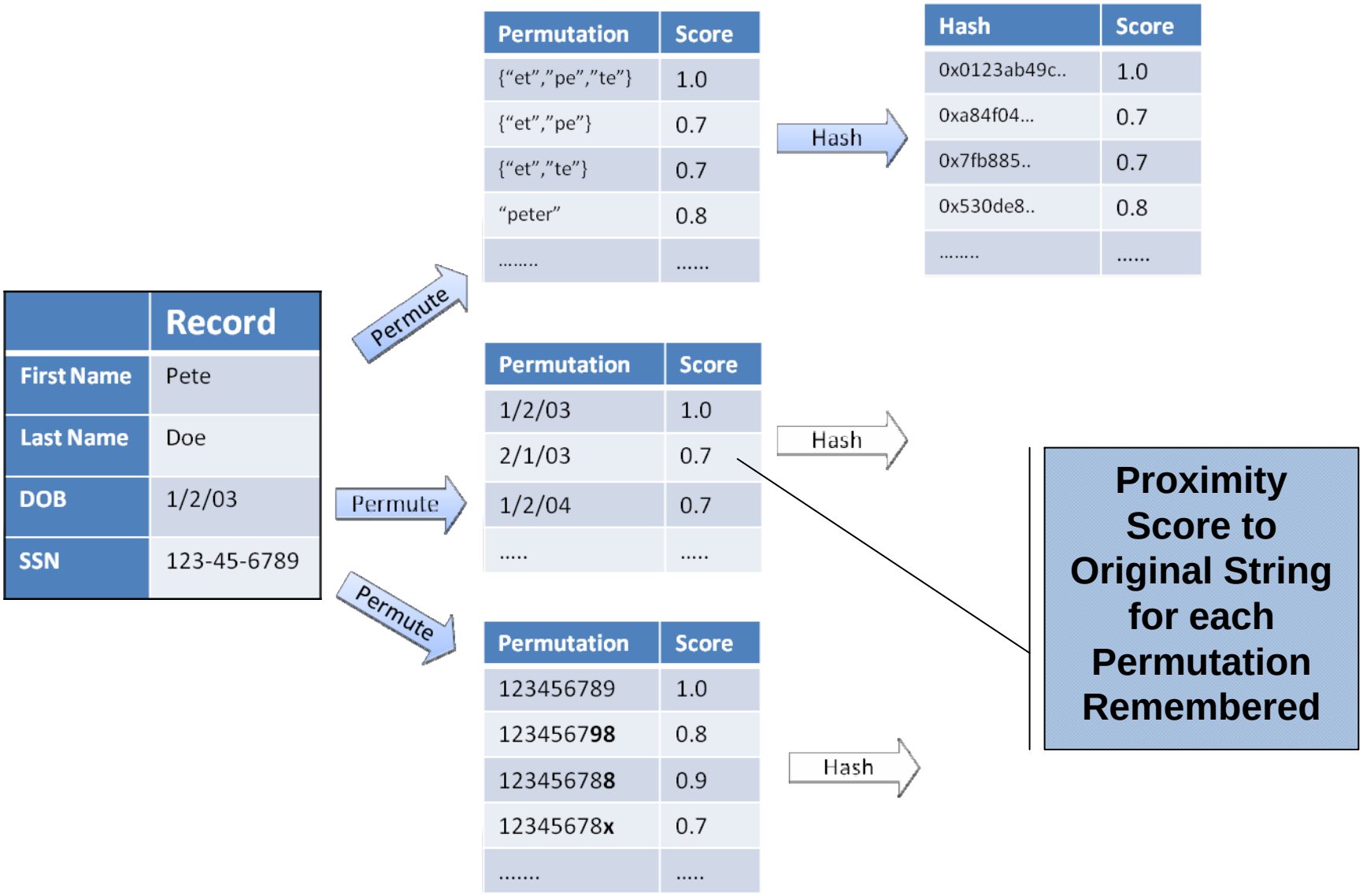
6)Link to other records based on identifier similarities

7)Group linked records into Patients

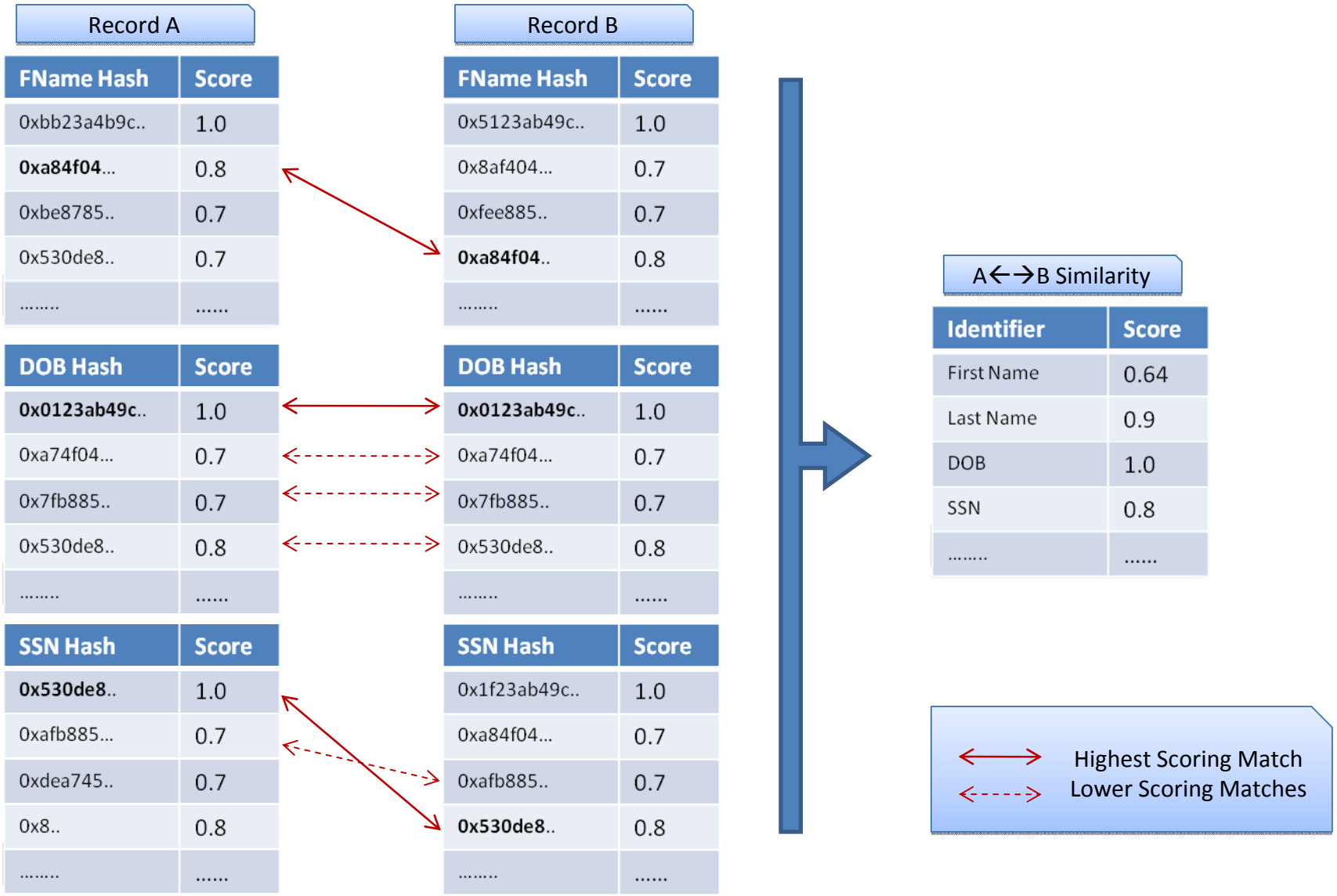
Shared Across Sites

Best Fit Permutation Strategy

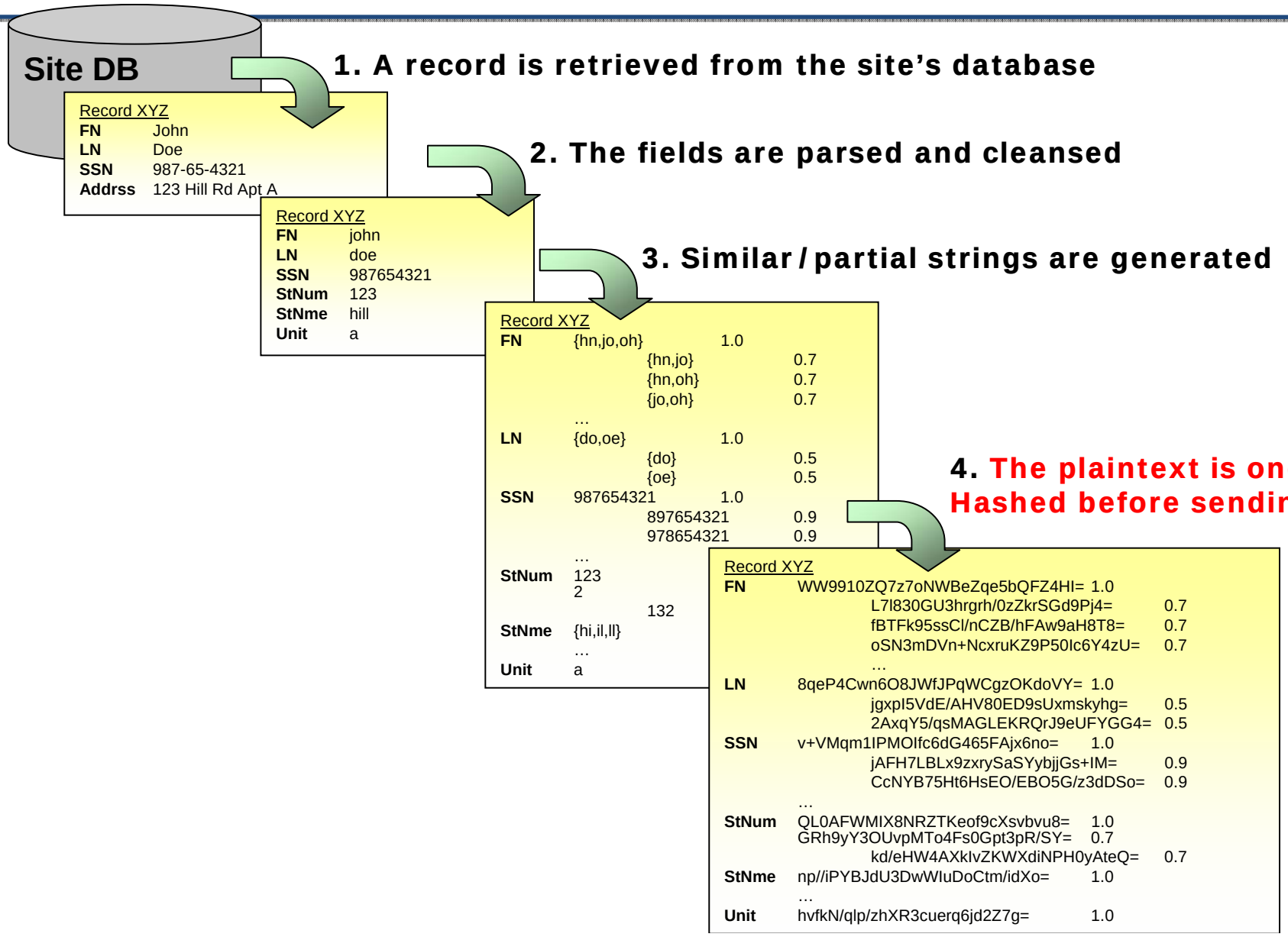
(bi-grams and nicknames for names, Digit transpositions for DOB/SSN etc.)



At the RLS/MPI – Hashes and Proximity Scores Received and Compared



Summary of Steps



At The RLS : Links Created

A←→B Similarity

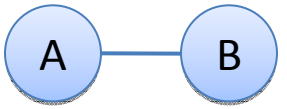
Identifier	Score
First Name	0.64
Last Name	0.9
DOB	1.0
SSN	0.8
……	……

Rule Based

Condition	Outcome
FName>.6 + LName>.8 + SSN>.7 + DOB=1	Definite Link
SSN>.7 + DOB>.9 + LName>.8	Possible Link
……	……

Result

Definite Link

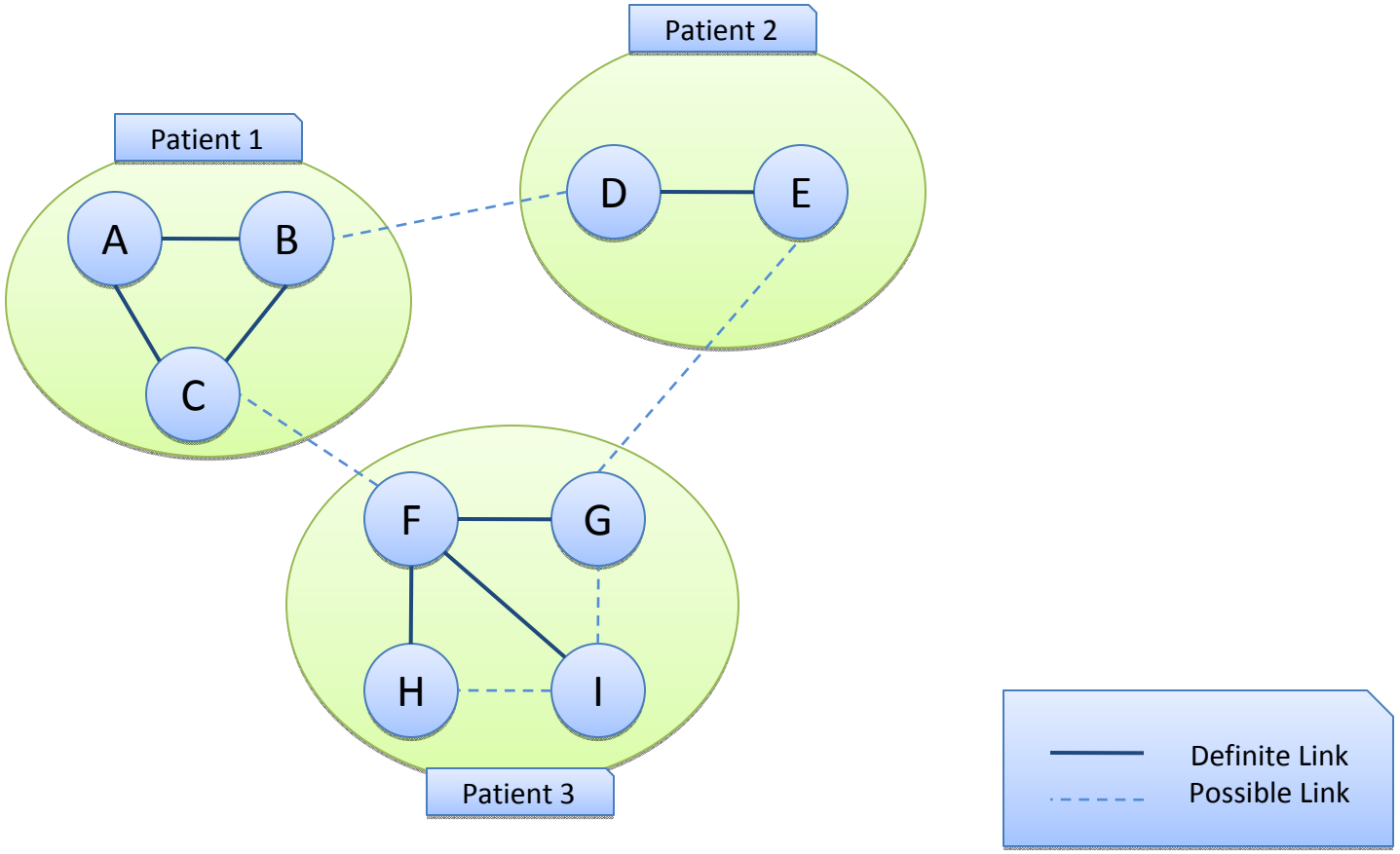


Weights Based

Identifier	Agreement Wt	Disagreement Wt
First Name	1	-5
Last Name	1	-5
SSN	10	-8
DOB	3	-4
…..	…..	

	Threshold
Definite Link	12.4
Possible Link	5.6

Pair-wise Links Grouped Into “Patients”



Consistency Checker Pass

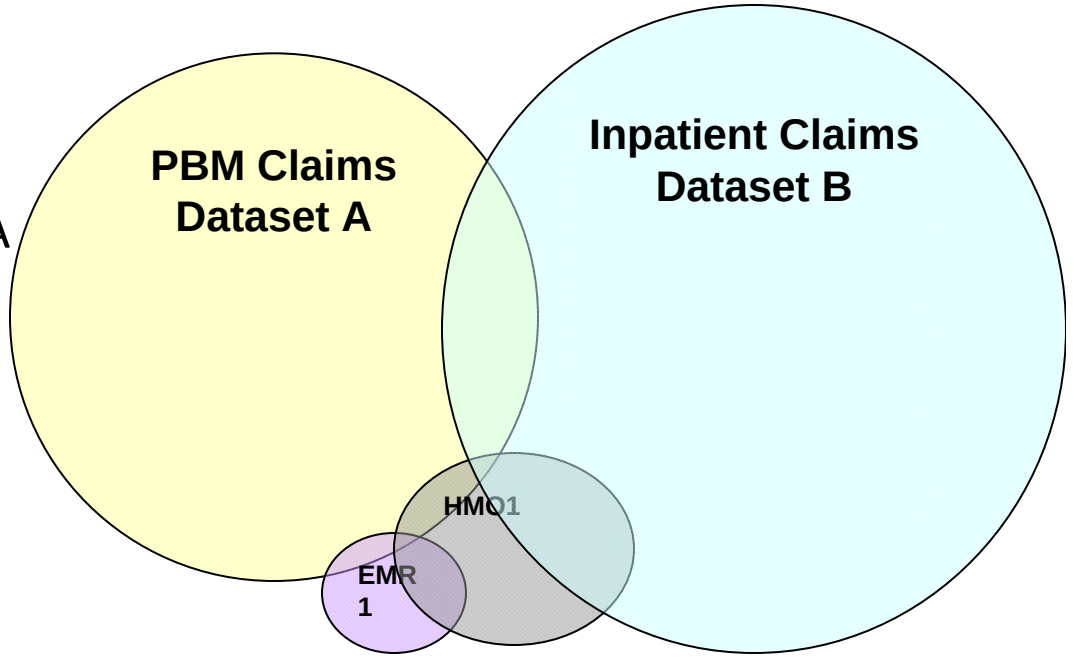
- Typical RL approach: pair-wise links (even with advanced Bayesian probabilistic algorithms)
  - Only as good as core algorithm
  - Weakest link creates issues
- Healthcare: n-way linkage
  - A1: Bithika S. Kheterpal, 11/8/68, F, SSN1
  - A2: Bithika S. Malhotra, 11/8/68, F, SSN1
  - A3: Bithika S. Malhotra, 11/8/68, F, missing SSN
  - A1=A2; A2=A3 but A1<>A3
- Consistency Checker will promote the A1 link.
- **Impact: HUGE improvement in sensitivity without sacrificing specificity**

Additional Data Model Characteristics of Relevance to Sentinel (Jeff Brown, 1/11/2010)

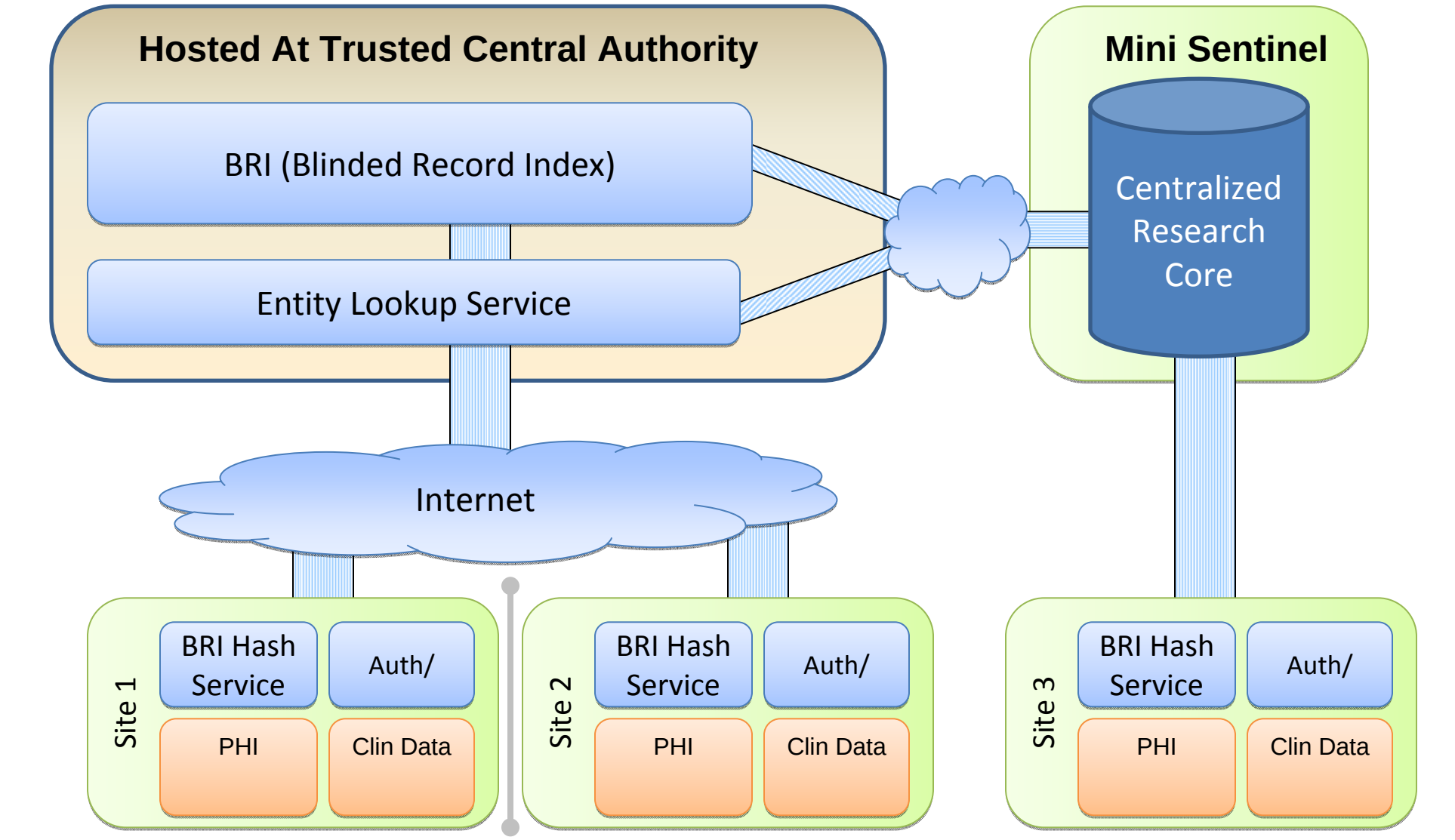
- Linkage: Medical Charts
  - Access to detailed information from the full text records is essential for primary users
    - Allows validation of outcomes & some exposures
    - Provides information on co-existing conditions, indications, and other data to elucidate findings
- Linkage: External data sources
  - Can provide data beyond that found in administrative & claims databases or EMRs
- Linkage: Between institutions
  - Identify individuals across different care settings
  - Longitudinally identify individuals across data holders
- Timeliness
  - Interval until data becomes available for analysis
  - Varies by data source and system

Another Way to Visualize Need

- Core Value Proposition
  - Address Subject Duplicity
  - Data Gaps
  - Longitudinal Follow up
    - Acute Events In Dataset B
    - Prescription Hx in Dataset A
    - Details in EMR1



Blindfolded RL Web Service for Sentinel



Summary

- Blindfolded record linking is a solution to maintaining privacy **and** achieving linking
- Blindfolded record linking is viable and practical
  - currently running in production
  - Meets clinical use case requirements : generally far more stringent
  - Large population sets
- Current efforts/architecture can be extended to include blindfolded linking

ENGELBERG CENTER for
Health Care Reform
at BROOKINGS

Roundtable Discussion and Questions