



Linking Harvard for Clinical and Translational Science

*Powered by SPIN:
Shared Pathology Informatics Network*



Primary support: NCI, NLM, and the DF/HCC



SPIN addresses (3) pervasive issues

1. Linking routine care systems
2. Protecting patient privacy
3. Ensuring that hospitals remain in control

... and has been deployed for:

- Translational Research requiring human specimens
- Public Health Surveillance
- With the potential for other clinical applications



How SPIN works

1. Link existing databases

- Extract from existing hospital systems
- Transform the data into common HIPAA-safe vocabulary
- Load into hospital controlled “SPIN peer”

2. Protect Patient Privacy per HIPAA

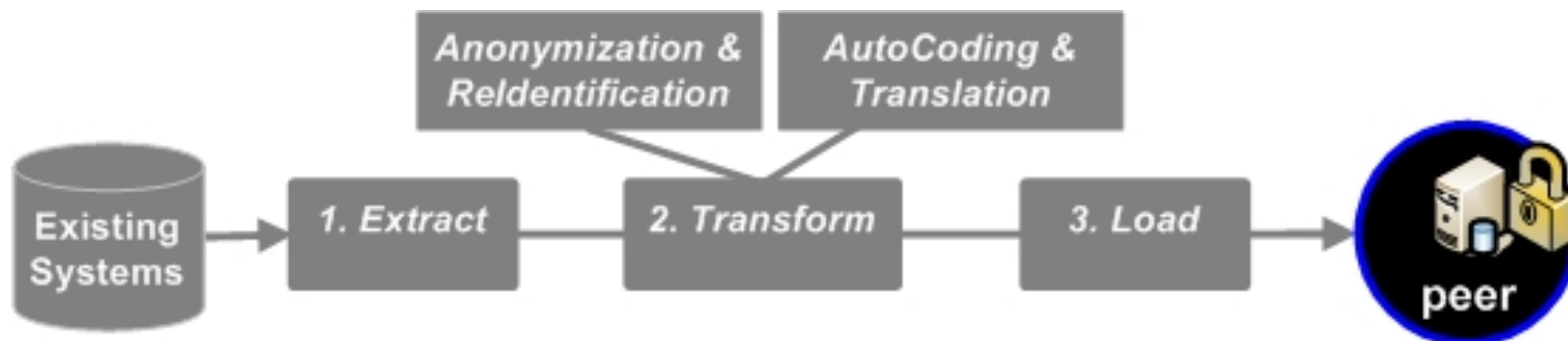
- De-Identified: Statistical Level query
- Limited: When authorized for individual cases
- *PHI is rarely used, and only with IRB from the Hospital (s).*

3. Hospital Control

- No central governing body
- Remain in control over disclosures at all times



(1) Linking routine care systems



- Extract from routine care delivery systems
 - Databases *or* XML
- Transform free text reports
 - “Scrub” patient identifiers (per HIPAA)
 - Autocode into controlled vocabularies such as UMLS
- Load into the hospital controlled PEER database
 - Assign a randomly generated ID to each case



(2) Protecting Patient Privacy

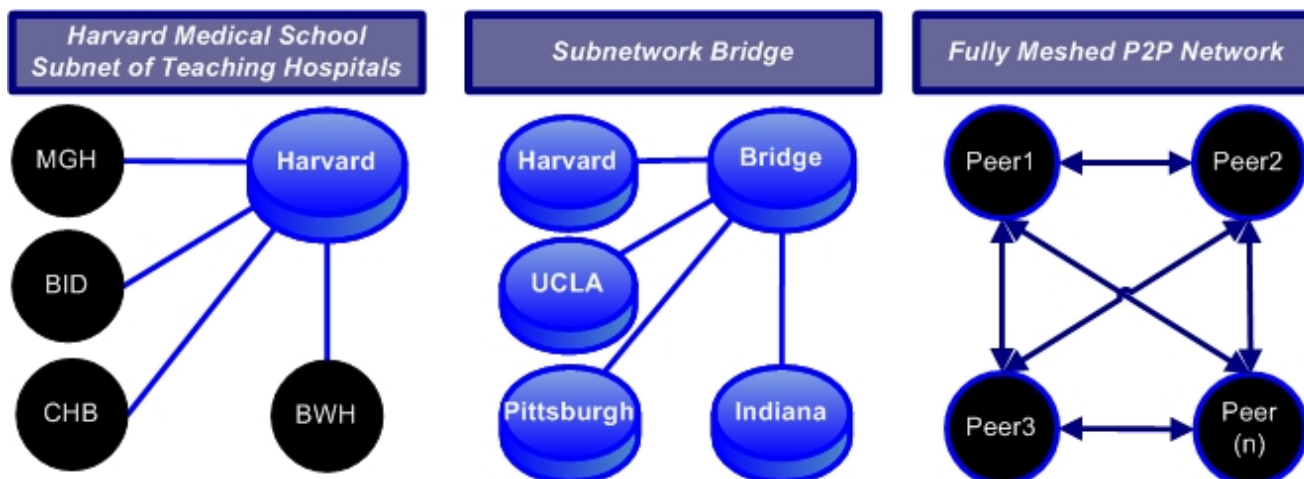
Increasing levels of investigator access commensurate with authorization by the hospital & investigator demonstrated need.

SPIN has enabled	<i>Statistical Queries</i>	<i>Non Identifying</i>	<i>PHI</i>
Translational Research	Feasibility Studies	Case Selection	Specimens
Public Health	Automated Analysis	Investigation	Emergencies
With potential for Clinical Applications	Research (CTSA)	QA/QC	Informed Care



(3) Hospitals remain in control

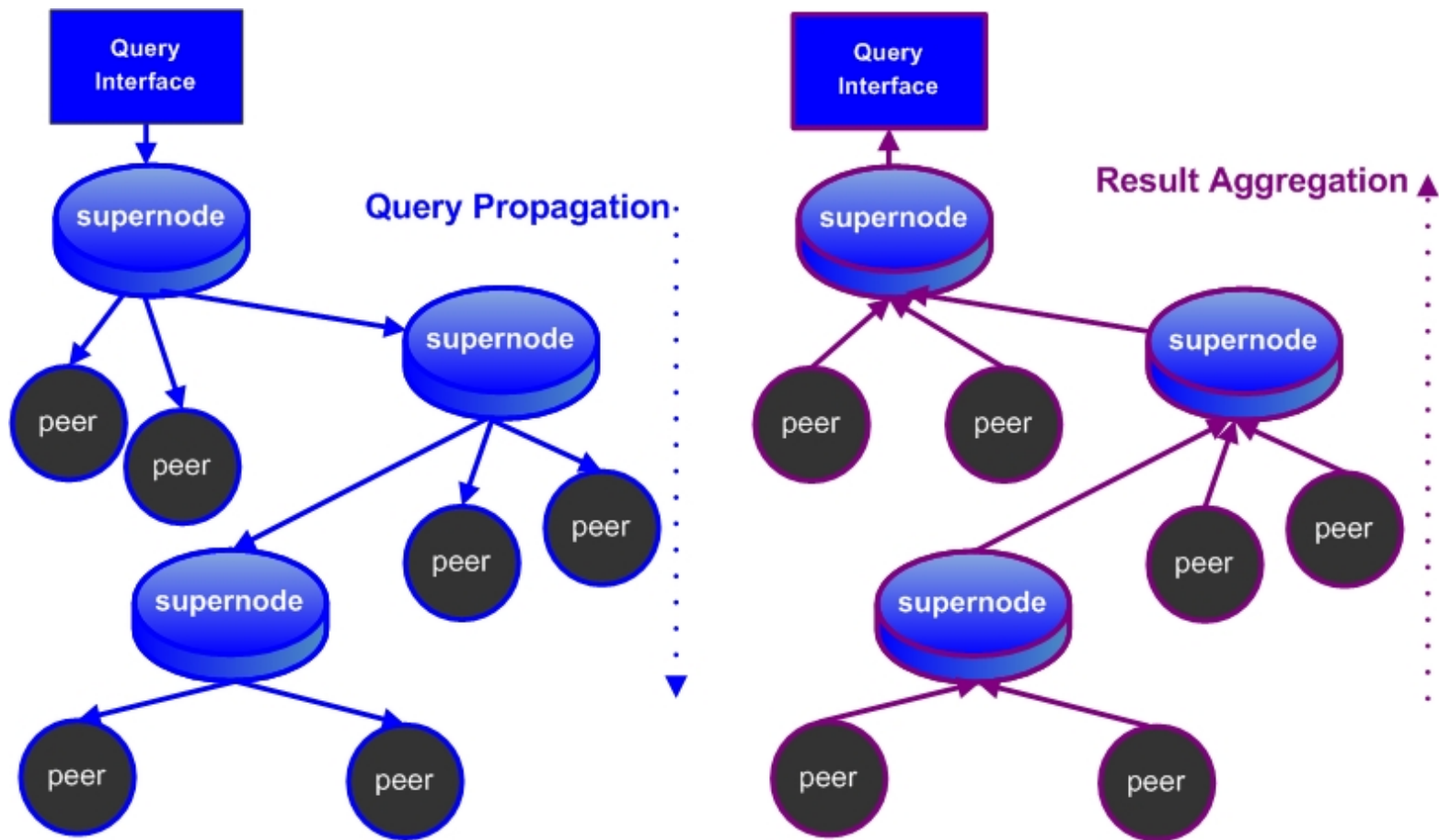
- Each hospital (Peer) chooses **who** to share with



- And **what** to share (Pathology Reports, ED feeds, ..)



Forming a decentralized network of subnetworks





SPIN Applications & Timeline

- **(2001 – Present) Translational Research**
- **(2006 – Present) Public Health Surveillance**
- **(2007 - Future) Clinical Applications**



2001 to present: SPIN for Translational Research

➤ **Motivation:**

Vast collections of human specimens and clinical data exist all over the country, yet are infrequently shared for cancer research.

➤ **Results:**

- ◆ Virtual Specimen Locator (all HMS hospitals)
- ◆ National prototype including HMS, UCLA, Indiana
- ◆ Later adopted by caBIG (caTIES is directly from SPIN)
- ◆ Direct influence on Markle's CFH Common Framework



2001 to present: SPIN for Translational Research

(1) Link Existing Pathology Databases

- Extract Pathology Reports (coded XML) from each site
- Scrub HIPAA identifiers & Autocode diagnosis for UMLS
- Generate random ID and load into local SPIN peer

(2) Patient Privacy: Increasing levels of investigator access

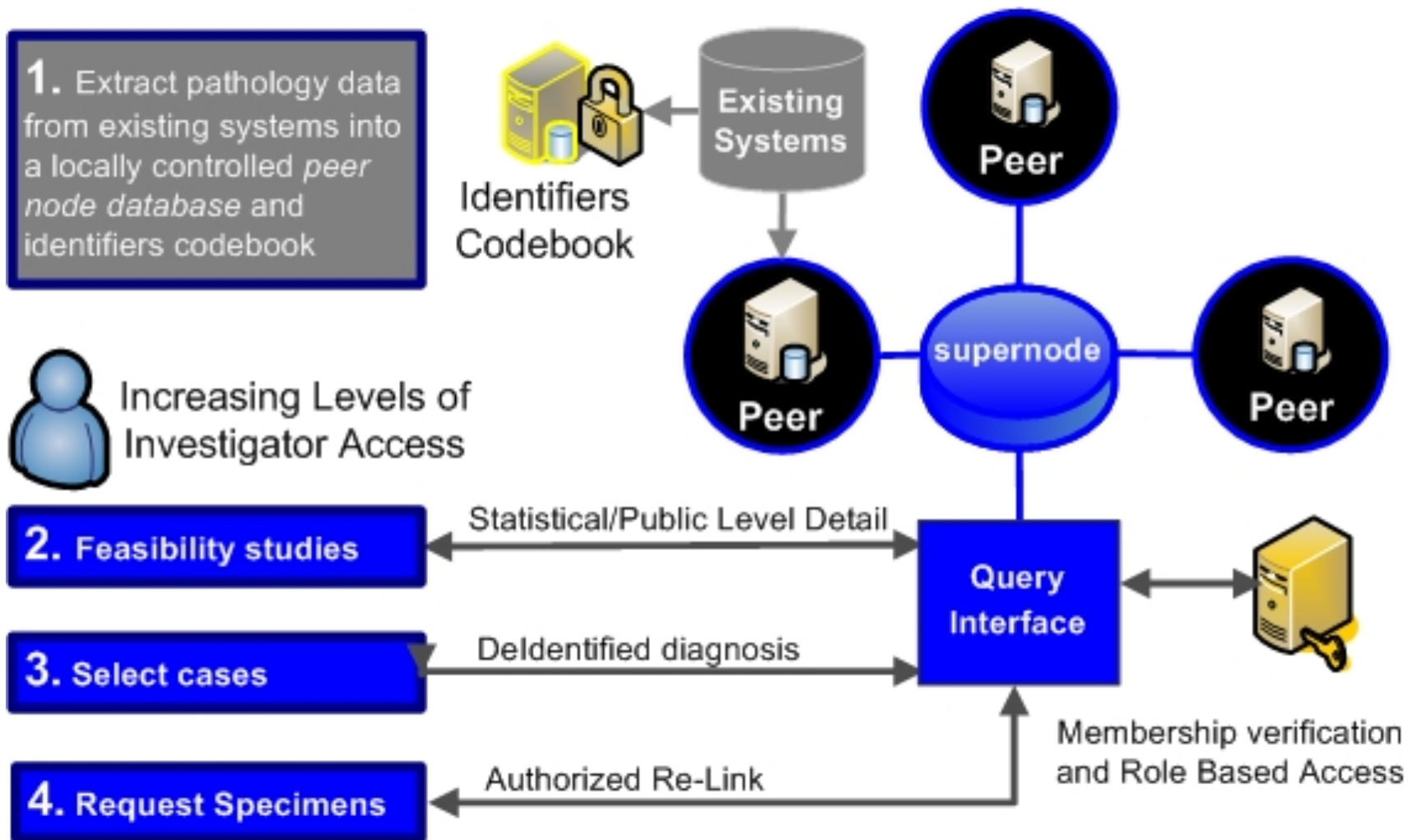
- Statistical query (With UMLS and keywords)
- Individual case query (De-Identified Path Reports)
- Specimen request (Remap case UUID to accession #)

(3) Hospital Control

- De-Identified reports for Harvard researchers (ecommons)
- PHI (to my knowledge) has never been used within VSL



2001 to present: SPIN for Translational Research





2006 to present: SPIN for Public Health Surveillance

➤ **Motivation:**

- ◆ Track the spread of influenza
- ◆ Search for patterns of disease activity
- ◆ Need ability to re-identify patients *as fast as possible* during public health emergency

➤ **Results:**

- ◆ Satisfies AHIC biosurveillance Use Case
- ◆ One of four federally funded NHIN architectures
- ◆ Working with Mass DPH to broadly deploy across the state
- ◆ Expressed interest: CDC & Australian Government
- ◆ Enables our existing biosurveillance application
AEGIS (aegis.chip.org)



2006 to present: SPIN for Public Health Surveillance

(1) Link Routine Care Delivery Systems

- Extract Emergency Department visits from each site
- Anonymize patient addresses & Autocode “Chief complaint”
- Generate random identifier and load into local SPIN peer

(2) Patient Privacy: Increasing levels of investigator access

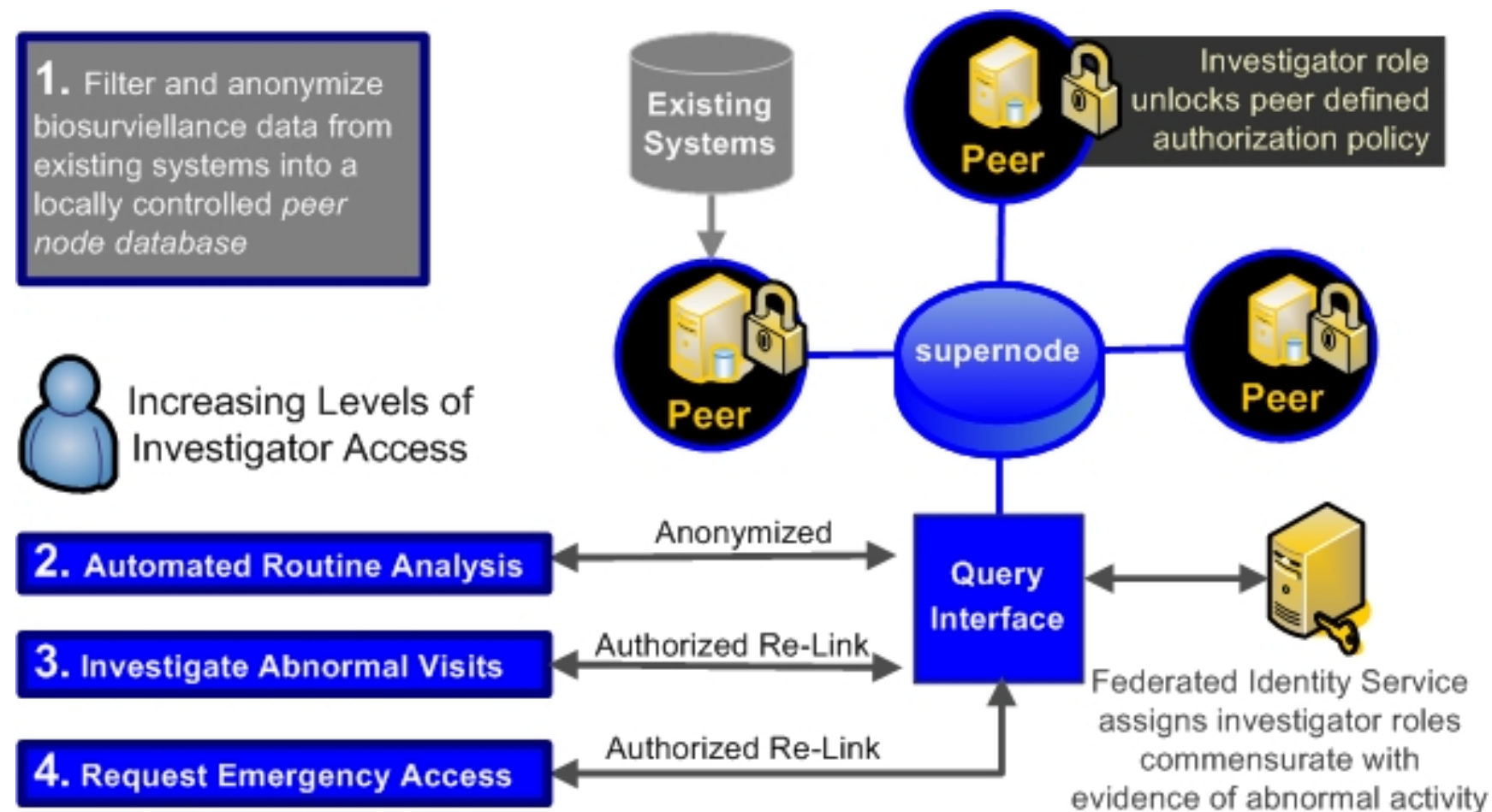
- Statistical query (Automated routine analysis)
- Limited Disclosure (DPH Investigation)
- Patient Re-Identification (Emergency Investigation)

(3) Hospital Control

- Which public health agencies do I trust? (CDC, DPH, HSPH)
- What do I want to allow in each investigation scenario?



2006 to present : SPIN for Public Health Surveillance





2006 to present : SPIN for Public Health Surveillance

Three upgrades:

(1) Modular submission process (ETL)

- Extractors fill values into a plain old object
- Transformers (anonymize & code free text)
- Loaders submit the results into the SPIN peer

(2) Distributed Identification and Credentialing

- Peers choose which agencies to trust
- Peers specify the disclosure policy per each ROLE
- Peers and patients can audit disclosures

(3) Peer database encryption

- (Optional) DRM style data storage with derby



2006 to present : SPIN for Public Health Surveillance

Separate Authentication (CDC) and Authorization (Hospital)

	Routine Analysis	Alarm Investigation	Emergency Investigation
Visit ID	Anonymize	Permit	Permit
Gender	Permit	Permit	Permit
Chief Complaint	Permit	Permit	Permit
Location	Anonymize	Anonymize	Permit
Disposition		Permit	Permit
Temperature		Permit	Permit
Checkin Time		Permit	Permit
Discharge Time		Permit	Permit
MRN			Permit

(One Example of Hospital Authorization Policies)



2007: Clinical Applications using SPIN

Motivation

- Linking clinical data for pathology in VSL (spin.chip.org)
- Pharmacovigilance (aegis.chip.org)
- Personally Controlled Health Records (indivohealth.org)

- Harvard CTSA?



2007:

Clinical Applications using SPIN

(1) Link Routine Care Delivery Systems

- Extract clinical data from each site
- Perform anonymization & autocoding
- Generate random ID and load into local SPIN peer

(2) Patient Privacy: Increasing levels of investigator access

- Statistical query
- Limited Disclosure
- Patient Re-Identification (very rare)

(3) Hospital Control

- All disclosures authorized by the hospital NOT 3rd party
- PHI only with approval from the host hospital (very rare)



2007: Clinical Applications using SPIN

We have the principles, now here are the challenges:

- *For each* participant and *for each* type of data exchange, we need to **map heterogeneous databases**
- **Agreement to share!** (IRBs and the political will)



2007: Clinical Applications using SPIN

Applying lessons learned: mapping heterogeneous DBs

The screenshot shows the Altova XMLSpy interface with an XSD schema loaded. The schema defines a root element 'query' and various sub-elements and types. Key elements include:

- query**: The root node containing all query and return data specifications.
- IdentityType**: A data type for the identity node.
- EvaluationSetting**: An abstract data type for a single 'global' query evaluation setting.
- ConditionConjunction**: A logical conjunction for evaluating criteria on database values.
- ConjunctionType**: A data type for logical conjunction combining criteria within a single condition.
- Comparison operators**: Elements like 'eq', 'neq', 'gt', 'geq', 'lt', 'leq', 'contains', 'matches', and 'condition'.

VS

The tree diagram illustrates the mapping of criteria to specific data types and values. The structure is as follows:

- Criteria**
 - StatisticsType**
 - BIN_ON_AGE:StatisticsType
 - BIN_ON_AGE_AND_GENDER:StatisticsType
 - Range**
 - Range(int, int)
 - min:int
 - max:int
 - Gender**
 - male:boolean = true
 - female:boolean = true
 - transgender:boolean = true
 - unknown:boolean = true
 - ReturnVariables**
 - ageAtCollection:boolean = false
 - tissueDate:boolean = false
 - gender:boolean = false
 - pathReport:boolean = false
 - isReturningDetail():boolean
 - toString():String
 - ageAtCollection:Range = new Range(0, 100)
 - yearOfCollection:Range = new Range(1970, 2007)
 - gender:Gender = new Gender()
 - returnVariables:ReturnVariables = new ReturnVariables()
 - statsType:StatisticsType = StatisticsType.BIN_ON_AGE
 - freeText:String = ""
 - umls:String = ""
 - searchName:String = ""
 - cancerType:String = ""

Start SMALL : Grow the number of common terms!



2007: Clinical Applications using SPIN

Applying lessons learned: mapping heterogeneous DBs

1. Request for Capabilities & Statistics (What is available?)
2. Availability limits scope of the vocabulary
3. What big questions can be asked with only a few identifiers?
 - ◆ Pathology: age, gender, collection, free text “diagnosis”
 - ◆ Public Health: age, gender, location, free text “complaint”
 - ◆ CTSA: age, gender,, free text mining
4. Parallel tracks: autocoding and standard vocabulary approach
 - Different low hanging fruit: diagnosis *vs* MRN
5. Quick End-To-End lifecycles
 - Question, development, research, new question



2007: **Clinical Applications using SPIN**

Agreement to share: IRBs and political will

- Statistical level queries easy are OK by IRBs
- Difficulty arises going to the next step
 - ◆ HIPAA limited data set
 - ◆ PHI
- Derivative of VSL IRB could potentially be used for CTSA



2007: Clinical Applications using SPIN

Agreement to share: IRBs and political will

- ANY use of patient data for research imposes SOME risk
- Minimize risk, show that research benefit is overwhelmingly in the best interest of patients



2007: Clinical Applications using SPIN

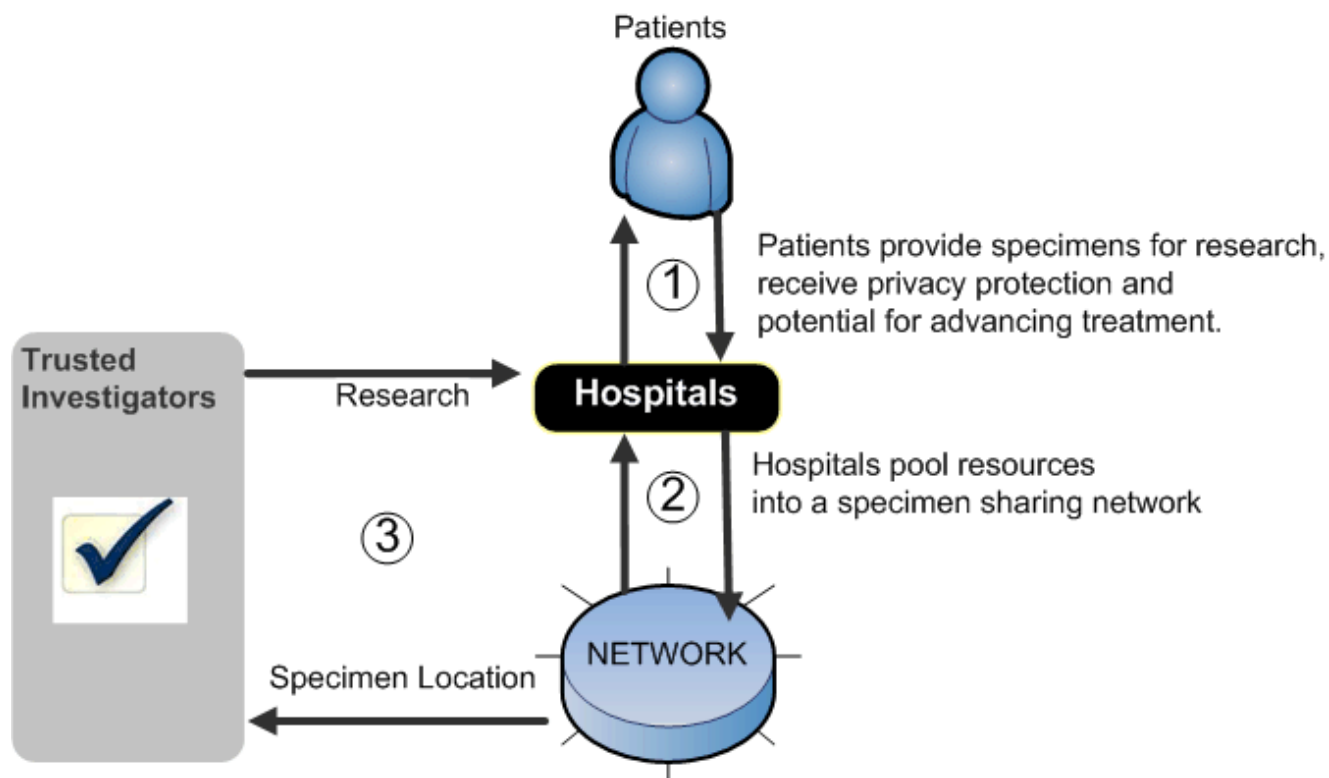
Agreement to share: IRBs and political will

- **Patients** are served with respect to both protections and research benefit.
- **Hospitals** can form collaborations with investigators they authorize.
- **Hospitals** retain authority of their valuable and cost intensive specimen collections.
- **Investigators** can easily locate specimens from multiple hospitals in a timely and affordable manner.
- *Keep patient interests paramount!*



2007: Clinical Applications using SPIN

Agreement to share: IRBs and political will





Summary

SPIN addresses (3) pervasive issues:

- Linking routine care systems
- Protecting patient privacy
- Ensuring that hospitals remain in control

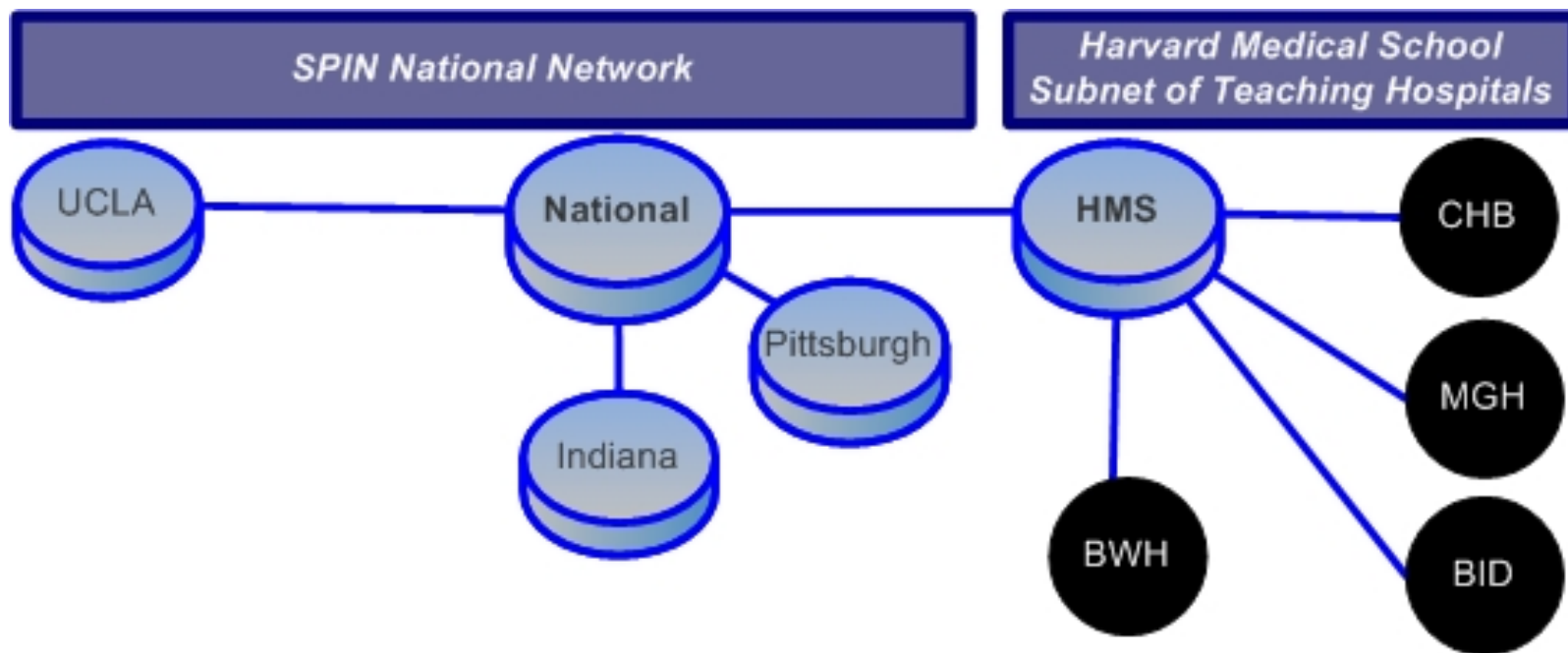
SPIN is broadly applicable for:

Research, public health, & clinical applications

SPIN is decentralized and builds agreement



SPIN Deployments





Final Comments

- **NOW:**
We know we can use routine care systems for research.
- **FUTURE:**
Why not use mature research systems to assist routine care?

