Confidentiality and Secondary Use of Data

Richard Steen, Healthcare Business Lead, NHIN Team, IBM

Monday, October 16th

2:45 – 4:00 PM

Defining Secondary Data Use – Confidentiality Requirements Vary Widely

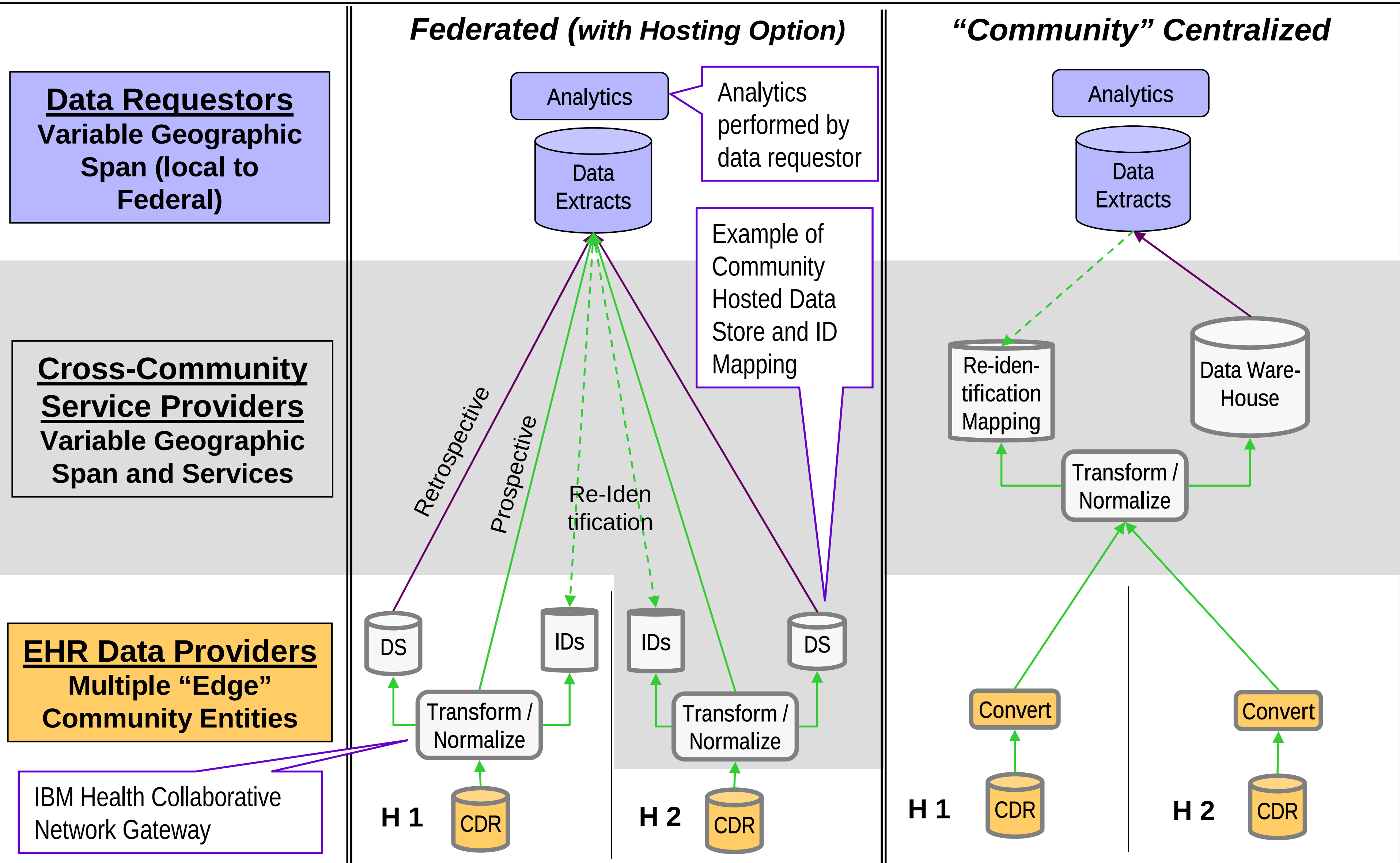
Varying Characteristics of Prospective and Retrospective Secondary Data Use

	1) Public Health - Biosurveillance *	2) Communicable Disease Reporting	3) Urgent & Targeted (VIOXX Recall)	4) QM & Clinical / Pharma Research
Live vs. Historical	Prospective Real-time	Prospective	Retrospective	Retrospective
Population Span	All Patients	Specific Patients with Disease	Population Sub-group	Research Qualified from Opt-in Patients
Data Extraction Requirements	Large Set; Defined Triggers	Smaller Set; Defined Type (a priori)	Specific; Types Defined by Event	Research Defined Data Sets
Confidentiality Standard	De-identified but Re-identifiable	Named Data for Limited Number	Partially De-identified but Re-identifiable	Typically HIPAA De-identified
Data Retention Requirements	Public Health Recipients	Source and PH Recipients	Data Custodian plus Regulatory Policies	Data Custodian plus Regulatory Policies
Re-identification Capability Required	Yes	N/A	Yes (potentially opt-in patients only)	Not Normally

Key: □ NHIN Phase 1

* Note: Future phases of Biosurveillance, e.g., Emergency Response will additionally require Secure Messaging.

Architectural Options range from Federated to Centralized - IBM's approach is federated with optional centralized hosting



Data Requestors
Variable Geographic Span (local to Federal)

Cross-Community Service Providers
Variable Geographic Span and Services

EHR Data Providers
Multiple "Edge" Community Entities

IBM Health Collaborative Network Gateway

***Federated** (with Hosting Option)*

Analytics

Analytics performed by data requestor

Data Extracts

Example of Community Hosted Data Store and ID Mapping

Retrospective

Prospective

Re-Iden tification

DS

IDs

IDs

DS

Transform / Normalize

Transform / Normalize

H 1

CDR

H 2

CDR

"Community" Centralized

Analytics

Data Extracts

Re-iden-tification Mapping

Data Ware-House

Transform / Normalize

Convert

Convert

H 1

CDR

H 2

CDR

Issue #1 - Inter-organizational coordination required to carry out the data gathering, anonymization and data delivery activities

- **Inter-organizational Challenges to**
  - Coordinate requestors (PH, QM, Research) and providers (entity providers, service intermediaries)
  - Coordinate data custodians (entity providers and service intermediaries) and data owners (clinicians, patients, consumers)
- **Complexity of managing data aggregates requires elegant but flexible solutions to**
  - Protect data confidentiality
  - Simplify inter-organizational coordination
  - Ensure compliance with patient and provider consents
  - Address ownership interests

Issue #1 - Inter-organizational coordination required to carry out the data gathering, anonymization and data delivery activities

- **Current approach - Hybrid (adapted Federated)**
  - Data gathering, transformation and data custodianship are federated with optional centralization at service provider hub
  - Re-identification is currently federated but can be centralized through a secure Patient ID Cross-reference service

- **<u>Pros</u>**
  - Source entity in direct control of data use and re-identification requests
  - Easier to ensure data integrity/currency

- **<u>Cons</u>**
  - Greater complexity in gathering and transforming data at each “edge” provider
  - More difficult to ensure reliable delivery of data from multiple “edge” providers

Issue #2 - Standardization of the methodologies needed to accomplish this intent

- **Standardization of methodologies needed for**
  - Data origination
  - Data extract requests (specification language)
  - Normalization tools (consistent coding, terminology mapping)
  - De-identification (by requestor type)
  - Anonymization (sometimes re-identifiable)
  - Data owner consent
  - Transmission protocols
  - Storage formats
  - Secure messaging

- **Pros (Hybrid - adapted Federated approach)**
  - Source entities directly influence how standards are implemented
  - Source entities manage patient and provider consent processes for new data requests

- **Cons**
  - Difficult to implement consistent normalization and terminology mapping across individual entities

Issue #3 - Data persistence requirements of the various approaches

- **Data Persistence Challenges to**
  - Guarantee synchronization of retained data stores
  - Maintain consistent sun-setting retention policies by data type
  - Guarantee data redundancy and resiliency
  - Ensure proper data retention
  - Respect data owner consents
  - Control access for re-identification

- **Pros (Hybrid - adapted Federated approach)**
  - Supports in-house or outsourced data retention
  - Secure Patient ID Cross-reference service correlates patient encounters across multiple providers (helps eliminate duplicates)

- **Cons**
  - Community-level re-identification service requires authorization management
  - Requires ability to match "Just in Time De-identification" services to authorization rights of requestor

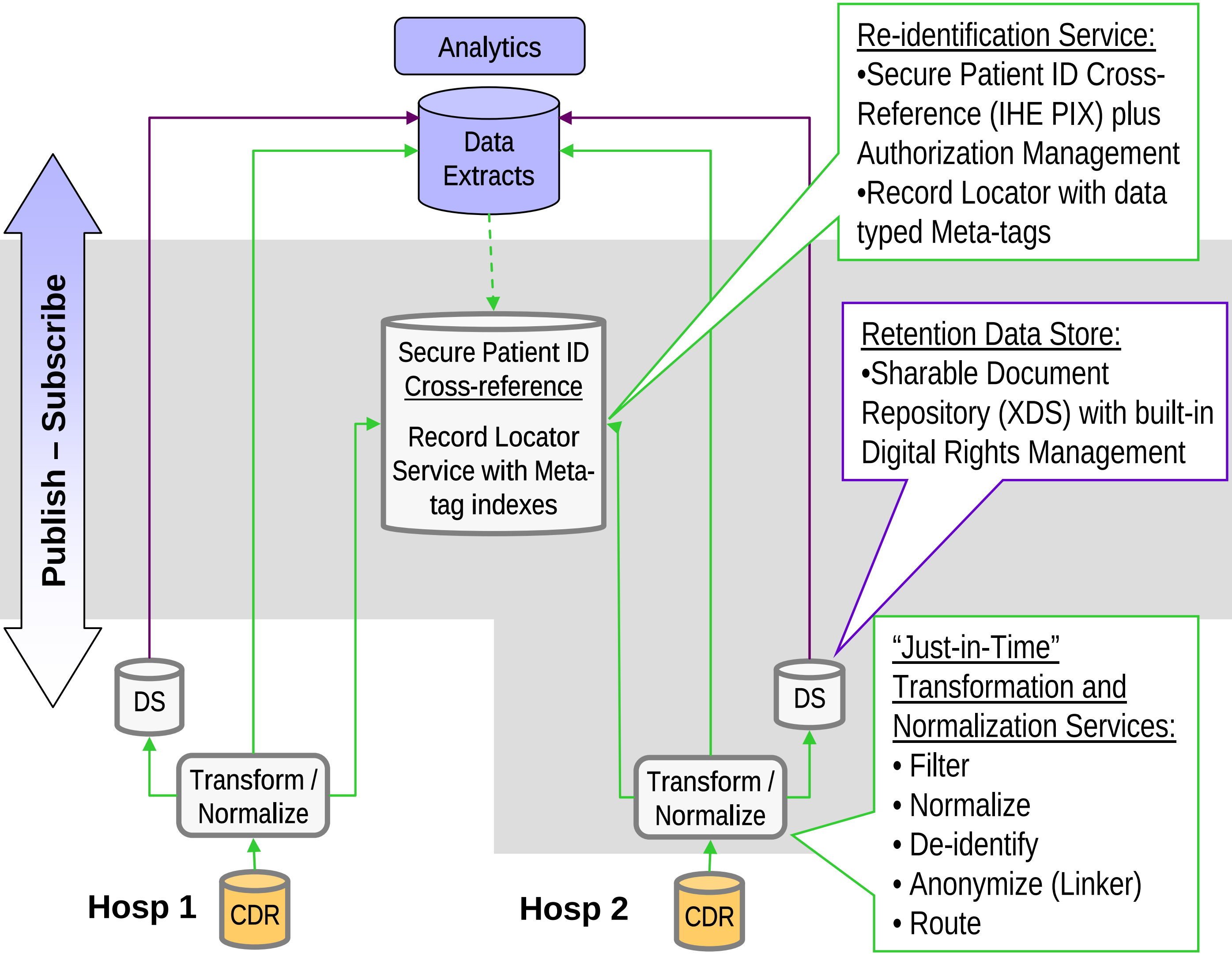
Issues #3 - Possible Future Vision for Data Retention: Intelligent data objects broker digital access rights, retention life, re-identification rights, etc.

Data Requestors
Variable Geographic Span (local to Federal)

Cross-community Service Providers
Variable Geographic Span and Services

EHR Data Providers
Multiple "Edge" Community Entities

Digital Rights – right data, right role, right time, any place



Analytics

Data Extracts

Secure Patient ID Cross-reference

Record Locator Service with Meta-tag indexes

Publish – Subscribe

DS

Transform / Normalize

CDR

Hosp 1

Hosp 2

Re-identification Service:
•Secure Patient ID Cross-Reference (IHE PIX) plus Authorization Management
•Record Locator with data typed Meta-tags

Retention Data Store:
•Sharable Document Repository (XDS) with built-in Digital Rights Management

"Just-in-Time" Transformation and Normalization Services:
- Filter
- Normalize
- De-identify
- Anonymize (Linker)
- Route

Issue #3 - Data persistence requirements of the various approaches

Possible Future Vision – Intelligent Data Objects

- Retained data as Sharable Document Repository (XDS) with built-in Digital Rights Management
- Publish and Subscribe model
- Record Locator with data-typed Meta-tags for cross-entity data identification

- **Pros**
  - Elegant but flexible solution
  - Intelligence resides with the data object
    - Portable
    - Breach Resistant
    - Redundant / Resilient
  - Supports on-going control of owner consent, and where appropriate, data usage fees

- **Cons**
  - Massive standards work required
  - Immaturity of Intelligent Data Objects and Digital Access Management paradigms