Interoperability MATTERS

an initiative of The Sequoia Project

Sequoia Privacy and Consent Workgroup

Data Segmentation for Privacy

Mohammad Jafari

January 2024

the sequoia project

Background

- Data Segmentation:
  - Identifying data elements that are subject to additional privacy/security controls.

- Sensitive Data
  - Data elements the disclosure of which involves a higher risk to the patient due to social, cultural, or economic reasons.
  - Lack of confidence in the protection of the information can lead the patient not seeking care or not sharing information.
  - Social, cultural, and economic consequences may negatively impact the patient if the information is disclosed more than necessary.

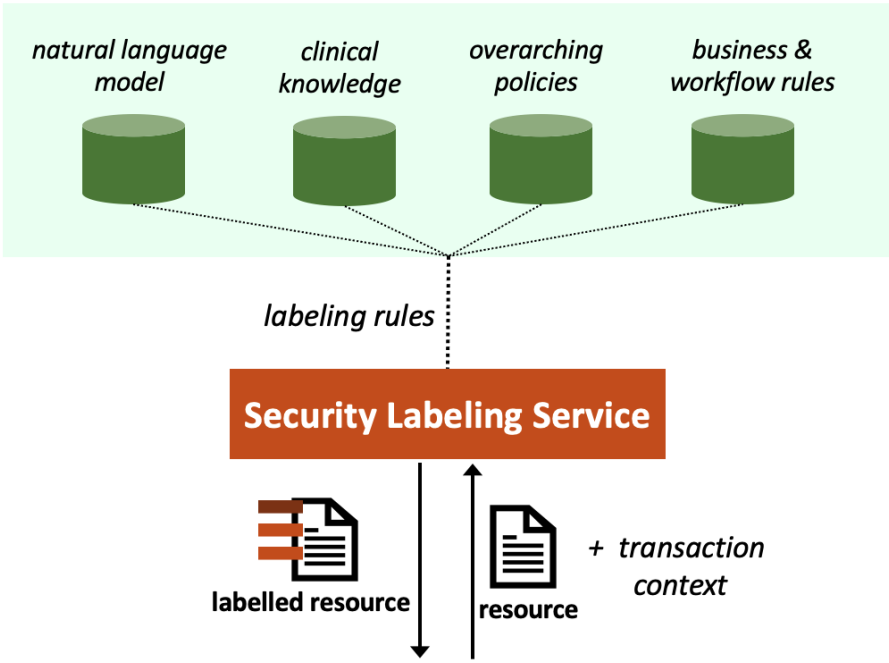
- The expectation of privacy leads to state and federal regulations:
  - *Substance Use Data*, *Psychotherapy Notes*, *Behavioral Health Data*, *Reproductive Health Data*

Key Components

- Tag data
  - Security Labeling Service
- Record labels
  - Standard specifications
    - CDA, V2, and FHIR
  - Standard codes for labels
    - HL7 terminology
- Process Labels
  - incorporate in authorization decision e.g., consent enforcement
  - incorporate in workflow, e.g., prevent sensitive information from access
  - incorporate in UI/UX, e.g., mark sensitive data

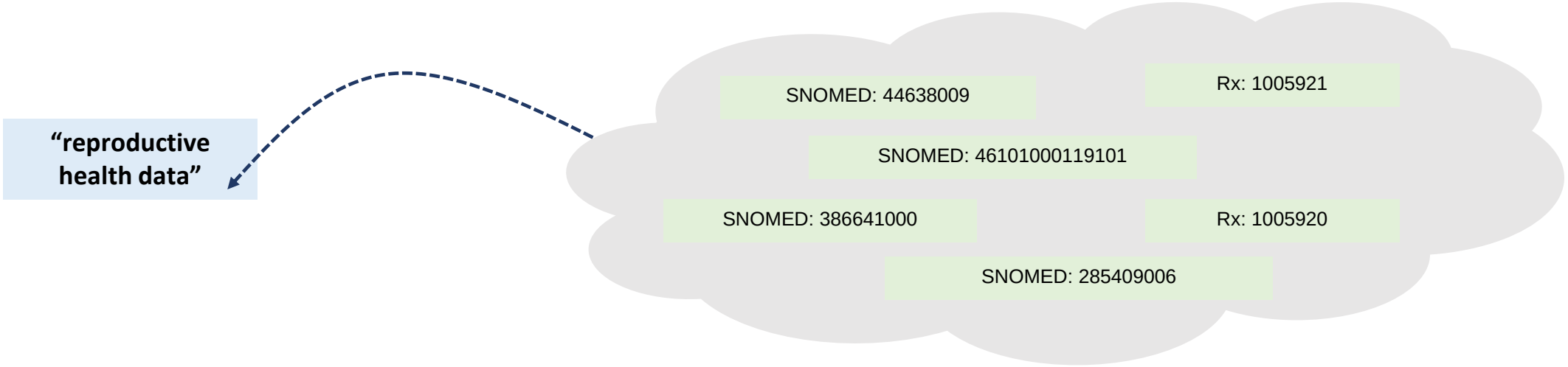
Security Labeling Service

- Rudimentary labeling plainly based on pre-determined value sets
- Probabilistic labeling
- More sophisticated processing
  - Related resources
  - Encounter context
  - Facility type
  - Unstructured text: NLP and LLM



Rudimentary Sensitivity Labeling

Mapping from detailed clinical concepts to a standard sensitivity code



Labeling Metadata

- Who performed the labeling
  - Identity of the entity that applied the label
- The basis for labeling
  - The law or regulation behind the labeling
- Time stamp
  - Determining whether the data has changed since the decision to label

Technical Architecture Considerations

- Where does the labeling service reside?
  - Participant or sub-participant vs. QHIN
- When does the labeling take place?
  - At the time of transaction
    - Always label the latest version of the documents, no need to persist labels or re-label
    - Response-time challenges
  - Offline
    - Batch or bulk labeling of data at rest and persist the labels
    - Advanced processing (e.g., unstructured text) is possible because of the offline nature.

Other Considerations

- Who is responsible for labeling the data
  - Participant or QHIN
    - Participants can still outsource this to the QHIN
- What sensitivity categories must be supported
- How to record labeling metadata
- A subset of sensitivity codes to be supported by all entities
- Redact vs. share with labels
- What are the rules for processing labeling data for the recipient

Challenges and Gaps

- HL7 specifications are available but need to be actively updated and maintained
- HL7 terminology for sensitive categories need to be overhauled
  - More granular codes
  - Deprecate old codes
  - Update definitions
- More implementation guidance on:
  - Standard HL7 codes to use for different classed of sensitive data identified in US regulations
  - Value sets (of clinical codes) tied to each sensitivity category