The Office of the National Coordinator for Health Information Technology

Data Provenance

Environmental Scan

July 2013

Office of the Chief Privacy Officer

The Office of the National Coordinator for Health Information Technology

Data Provenance

- What is "Provenance"?
  - Origin of clinical information when first created
    - Information about the source of the data
    - Information about processing/transitions the data has undergone
- Why is it important?
  - Enables segmentation of information based on source
  - Enhances provider trust in information being exchanged between providers
  - Enhances provider trust in information received from a patient (e.g., data from PHR)

About the Environmental Scan

- Explore how EHRs, PHRs and health information exchanges (HIEs) track data provenance
- Focus on marking and retaining provenance as systems aggregate data from multiple sources and records are exchanged
- Focus on provenance within CDA documents
- Includes landscape analysis, gap analysis

Key Provenance Considerations

- Who/What to list as source?
  - Organization
  - Provider
  - Data entry staff
  - Device details
- At what level?
  - Entire document (full record has one source listed)
  - Section (ex. medication section lists its own source)
  - Individual data element (ex. each individual medication lists its own source)
- How to update or modify source when importing/exporting?
  - List receiving organization as source
  - List original creator as source

Current Provenance Practices

- How are systems documenting provenance today?
  - Considerable **variability** in how provenance marked and retained in EHRs, PHRs, and HIEs
    - Different levels of granularity (document level vs section level vs data element level)
    - Different practices regarding source (organization vs provider vs individual entering the data)
    - Different practices when importing/exporting data (list original source vs modify source to list importing org)

Flow Down of Provenance Data

- Common response to variability is to utilize "flow down" of provenance data
  - Mark provenance at **document level only**
  - Provenance data inherited at data element levels
    - Example: the source for an individual medication will be the same source listed at document level
  - Results in insufficient granularity of detail; creates integrity issues; undermines trust

Implications for Patient-generated Data

- Generally, provenance at data element level is lacking as records are shared with providers
  - EHRs often export with sending organization listed as source on **document level only** (not section or element)
  - PHRs may import with provenance at document, section and/or data element level, but often export with PHR as source on **document level only**
  - HIEs can export with documentation at data element level, but sufficient information often **not available**
- Resulting flow down of document level provenance data may not meet provider need for granular detail

Challenges and Opportunities

- Currently no dominant provenance model within the HIT community
  - No uniform way of handling data provenance when sharing data
- No harmonized standard currently in place
  - Upcoming work includes HL7 data provenance and privacy support in C-CDA initiative
- Current flow down practice does not meet provider need for granularity regarding origin of data


Office of the Chief Privacy Officer
The Office of the National Coordinator for
Health Information Technology


HealthIT.gov