Data Provenance Initiative Charter

Background

The term "provenance" in the context of Health IT refers to evidence and attributes describing the origin of health information as it is captured in a health system. The requirements for data provenance information must support the full lifecycle and lifespan of health IT data. As the exchange of health data increases, so does the demand to track the provenance of this data over time and with each exchange instance. Confidence in the authenticity, trustworthiness and reliability of the data being shared is fundamental to robust privacy, safety, and security enhanced health information exchange. Truth and trust may be improved by means of a standardized way to capture and express the provenance of the data and by the expectation that systems have the ability to recognize and validate the provenance information. This in turn can lead to uses such as "chain of trust" and "chain of custody" and other business requirements/applications (for example: records management, evidentiary support and clinical decision support).

A significant amount of literature has discussed the value, importance, and legal necessity of provenance.

The Data Provenance Initiative aims to establish a standardized way of capturing provenance (including inbound, system generated, and outbound provenance), retaining and exchanging the provenance of health information. (Provenance Capabilities)

Challenge Statement

While there are efforts to address data provenance, no existing authoritative specification, standard, or model for provenance has been commonly adopted to-date within the context of HIT. This is exemplified by the variance in how HIEs, EHRs, and PHRs currently capture, retain, and display provenance. This variability is problematic for the interoperable exchange (system interoperation), integration, and interpretation of health data.

Even current health information standards, such as the CDA, lack guidance for handling data provenance. Additionally, the receipt and integration of provenance information is equally variable and dependent upon system capabilities. Further challenges are presented if one system can share detailed provenance data but those receiving it cannot process the level of detail exchanged.

Purpose and Goals

It is important for healthcare stakeholders to have confidence in the authenticity, reliability, and trustworthiness of shared data (trust). This trust requires a clear trail of data provenance in order to validate and determine how to use the data.

To help meet this challenge, the goals of this initiative include:

- Establish guidance for handling data provenance in content standards, including the level to which provenance should be applied
- Establish the minimum set of provenance data elements and vocabulary
- Standardize the provenance capabilities to enable interoperability

Although significant challenges in addressing data provenance have been identified, there are some out of scope items that will not be directly addressed during this initiative. A Tiger Team will be established in addition to use case development to work with SDOs in support of this initiative. This group will serve to accelerate the standards analysis activity and provide recommendations back to the initiative. It is an explicit goal of this effort to identify those components of data provenance that can be expanded to support multiple content specifications. The initial focus of the Tiger Team will be a review of CDA and development of reusable building blocks. Subsequent application to other technologies using these building blocks will be prioritized by the initiative.

Scope Statement

To identify and define guidance on use of standards to facilitate provenance capabilities by specifying the following (to the extent specified by the Use Case):

- Standards for the provenance (e.g. origin, source, custodian(s), etc.) to the extent they can be supported by the use case
- Supportive standards (e.g. integrity, non-repudiation) to the extent they can be supported by the use case
- Standard metadata tags for data provenance
- Standards that define how a system can exchange and integrate data provenance and initially focus on defining the provenance for the CDA standard
- Variance in the level of granularity to which data provenance can be collected and how that provenance is communicated to consuming systems

Out of Scope

Out of scope items include the following:

- Patient identity matching
- Third party mechanisms for checking patient consent and the relative merits of existing policies or regulations (such as privacy policies or jurisdictional considerations)

Value Statement

The Data Provenance initiative will improve the confidence in the integrity of health information from creation to exchange and integration across multiple health information systems and between parties. Ultimately these standards will improve trust in healthcare data and its applications, which may include clinical care, interventions, analysis, decision making and clinical research, and others.

Data Provenance Proposed Technical Standards

Note – The below standards are a starting list for consideration and will be updated based on community feedback during Charter discussions and as use case requirements are further identified. This is not a full list of all possible standards.

- Cross Enterprise Document-Sharing (XDS)
- Simple Object Access Protocol (SOAP)
- Representation State Transfer (RESTful)
- HL7 Clinical Documentation Architecture Release 2 (CDA R2)
- HL7 Version 2 Vocabulary & Terminology Standards
- HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1
- HL7 FHIR DSTU Release 1.1 Provenance Resource
- W3C PROV: PROV-AQ, PROV-CONTRAINTS, PROV-XML
- HL7 Health Care Privacy and Security Classification System, Release 1
- HL7 Version 3 Standard: Privacy, Access and Security Services (PASS)
- HL7 EHR Records Management and Evidentiary Support (RM-ES) Functional Model, Release 2
- HL7 Digital Signature
- ISO 21089 Health Informatics: Trusted End-to-End Information Flows
- Personal Health Record System Functional Model
- HL7 Record Lifecycle Event Metadata using FHIR (project underway 2014)(Added October 23, 2014)
- ISO/HL7 10781 EHR System Functional Model Release 2 (2014)(Added October 23, 2014)
- HL7 EHR Lifecycle Model (2008)(Added October 23, 2014)
- ISO 21089 Trusted End-to-End Information (2004, currently in revision)(Added October 23, 2014)

Potential Risks and Challenges

This initiative may not address all of the data provenance challenges that exist today such as the following:

- Linkage of an individual's data across sources and organizations over time
- Insufficient "end-user" input (e.g. physicians, clinicians, technology vendors)
- Standards may not scale to small vendors and practices and/or accommodate clinical practice variations
- May be gaps in standards to support the initiative requirements
- Broad definition and use of provenance (we will need to define this clearly as part of this initiative)
- Gap to fulfill the roles and responsibilities of organizations to deploy data provenance solution (challenges due to implementation)
- It is important to focus not only on how provenance is expressed in data exchange artifacts (e.g., CDA, CCDA), but also to address what provenance is captured at the point of data origination (and managed thereafter).
    - The inability to capture provenance data at the point of origin
    - Provenance data being captured in the exchange my vary in granularity
    - Exchange standard reporting any degree of granularity around provenance
- It may not support additional applications of provenance supported in other domains (e.g. protect authenticity of data, security of data)

Stakeholders

Any entity can create or change the health care artifact to include:

- Patients and patient advocates
- Health Care Providers and Business Associates
- PHR/EHR/ HIEs/ Other Health Care Information System Vendors
- State HIEs
- Local, State, Federal Government Health Agencies
- Health Organizations (Research and Quality)
- Standards Organizations
- Healthcare Payers
- Ancillary Health Care Services (i.e. labs)