

Comments to: **Standards and Interoperability (S&I) Framework – Data Provenance Initiative (DPROV)**

Comments on: **Decision to Draft and Ballot an HL7 Implementation Guide (IG) for CDA Release 2 (CDA R2) without regard to, nor basis in, DPROV Community Use Cases**

Comments by: Gary Dickinson  
Director, Healthcare Standards, CentriHealth  
Co-Chair, HL7 EHR Work Group

Date: 28 July 2014 (submitted to S&I and DPROV Leads)  
10 September 2014 (revised and submitted to HITSC)

- Data provenance is about truth (authenticity) and trust (assurance).
- Data provenance represents (embodies) the source of truth – the point of data/record origination.
- Data provenance, if properly captured, retained, managed and conveyed from the point of origination forward ensures trust to all downstream users and for all purposes to which health information may be applied.

From the outset, there's never been an S&I (or prior AHIC/HITSP) Use Case Initiative where the subject matter is more self-evident (intuitively obvious) to the average person:

- 1) Data is anchored at the point of origination – always the source of truth – and bound PAIR-wise with its provenance from that point forward. Provenance is the who, what, when, where and why of each data element or collection originated together.
- 2) Each PAIR – data + provenance – is indivisible.
- 3) Each PAIR is immutable.
- 4) Each PAIR is:
  - a. created at the point of origination;
  - b. securely retained and managed by the source system over time;
  - c. securely conveyed in exchange artifacts (e.g., CCD/CCDA);
  - d. securely retained and managed by the receiving system over time...
- 5) As the function (result) of 1-4, each PAIR (and its chain of trust) evidences truth (authenticity) and ensures trust (assurance) to each downstream user and application.

Our comments follow...

- (A) One would anticipate that establishment of the S&I Data Provenance (DPROV) Initiative would be a strong positive step to anchor health data/records explicitly at their source of truth.

After review of all S&I Initiatives and as members of the Health Information Technology Standards Committee (HITSC) were polled on 17 June 2014, 8 of 18 (who registered a preference) identified DPROV as their first priority.

**Comments on: S&I Data Provenance (DPROV) Initiative – Decision to Draft and Ballot an HL7 Implementation Guide for CDA Release 2 without regard to, nor basis in, DPROV Community Use Cases**

Gary Dickinson, Director, Healthcare Standards, CentriHealth  
10 September 2014

- (B) One would anticipate, given the S&I Framework's solid reputation of deliberative analysis, S&I DPROV would logically start at the point of data origination, continue in sequence (per 4a-d above), and encompass the complete chain of trust (full data/record lifespan and lifecycle).

Initial statements from DPROV Leads indicated their intent was to focus entirely on the point of exchange. After DPROV kickoff it took 6 weeks of perseverance – to the point of Charter consensus – before DPROV Leads agreed to scope the Initiative starting at the point of data origination. Finally(!) the agreement that data provenance is actually about what happens at the point of data/record origination, as the essential precursor for what/how data provenance might be conveyed in an exchange artifact (e.g., CDA R2).

This decision (more seemingly a concession) followed concerns voiced by a number of DPROV Community members including representatives of the American Health Information Management Association (AHIMA), HL7 Electronic Health Records (EHR), HL7 EHR Interoperability and HL7 Records Management and Evidentiary Support (RM-ES) Work Groups, several EHR/PHR vendors, and others.

- (C) One would anticipate, with agreement of DPROV Leads and with the new Community Charter in place, the issue was settled and that DPROV work would proceed apace and aligned with that objective.

Instead a newly formed DPROV Tiger Team was commissioned to proceed on it's own, without regard to the DPROV Community Charter agreement, and began drafting an HL7 Clinical Document Architecture (CDA) Release 2 Implementation Guide (IG). The only thing clear in this separation is that the Tiger Team has produced an HL7 CDA IG derived independently from, and clearly at odds with, DPROV Community developed Use Case Requirements. Thus:

- DPROV Tiger Team focuses solely on point of exchange (4c above) – on Monday teleconferences (through July); whereas
- DPROV Community focuses on data and its provenance from point of origination and throughout its full lifespan/lifecycle (4a-4d above, including 4c) – on Thursday teleconferences (through to the present time).

- (D) One would anticipate that the DPROV Initiative would follow the same logical progression as the other 20+ S&I Framework Initiatives undertaken since 2011. In all other Initiatives, Use Cases are developed as a neutral statement of User Stories, Scenarios, Event Steps, Actors, Roles, System Functional Requirements and Data Requirements, ultimately agreed by Community Consensus. On such basis, standards and IGs are then selected or developed in the subsequent S&I Harmonization Phase.

As noted previously, the DPROV Tiger Team was commissioned to draft an HL7 CDA IG, prior and without regard to fully developed DPROV Community Use Cases specifying scenarios and requirements.

The DPROV Community (on Thursday calls) has not yet completed Use Case Requirements specification. It is of interest to note that, at this stage:

- All Community User Stories have a point of exchange – but several have two – in the narrative sequence.

**Comments on: S&I Data Provenance (DPROV) Initiative – Decision to Draft and Ballot an HL7 Implementation Guide for CDA Release 2 without regard to, nor basis in, DPROV Community Use Cases**

- Several User Stories have more than one data provenance event prior to the point of exchange. For example, where source data is transformed into the exchange artifact there are two data provenance events, one for the source data (point of origination), one for the transformed data (preparatory to point of exchange).
- At least one User Story describes the Assembler as a Patient. An Assembler is a human or software (system, device) that gathers patient health information from multiple sources and creates a new artifact (e.g., patient summary). The Assembler must ensure that data/provenance PAIRs are properly assembled and conveyed in the new artifact (i.e., ensuring 2-3 above).

(E) One would anticipate that when the S&I Framework considered existing Standards on the Data Provenance topic, there would be a natural tendency to seek those that specifically address data provenance fundamentals (1-3 above) and health data/record management from the point of origination and thereafter, including a transparent chain of trust and the full lifespan and lifecycle (4a-d above).

Obvious choices, until now almost entirely ignored by DPROV Leads, include:

- ISO 21089, Trusted End-to-End Information Flows, 2004. 2014 revision just approved as ISO TC215 work item and is now in progress.
- ISO/HL7 10781, Electronic Health Record System Functional Model (EHR-S FM), Release 2, 2014.
- HL7 Records Management and Evidentiary Support Functional Profile (RM-ES FP), Release 1, 2009.
- HL7 EHR Lifecycle Model DSTU, 2008.

(F) One would anticipate that when S&I approached HL7 with intent to collaborate on data provenance, there would be a natural tendency to seek established Standards and expert groups already convened in the subject matter.

Obvious choices, as authors of and expert contributors to the cited Standards yet bypassed by DPROV Leads, include:

- HL7 EHR Work Group (active since 2000)
- HL7 RM-ES Work Group (active since 2006)
- HL7 EHR Interoperability Work Group (active since 2004)

Instead DPROV Leads chose HL7 Community-Based Collaborative Care (CBCC) Work Group, whose prior data provenance related Standards work is non-existent.

(G) One would anticipate that when S&I Framework Initiatives bring forward specific updates to HL7 Vocabulary these would be based on completed S&I Use Case analysis and requirements specification, as the result of consensus-based Use Case Requirements, in harmony with existing HL7/ISO Standards, and in consultation with HL7 Work Groups expert in the subject matter.

Instead the DPROV Tiger Team (via HL7 CBCC) brought forward HL7 Vocabulary proposals that are highly speculative and without formal basis as described above.

For example, one Tiger Team/CBCC update proposal is for the addition of document “Assembler” (see reference in D above) as a new term in the HL7 Vocabulary. Their proposal, now leveraged by CBCC into the HL7 Vocabulary, defines “Assembler” as

software but not human. For now, the term will not support basic human or combined human/software roles in document assembly/origination – even though these patterns are already evident in DPROV Community User Stories. This point of conflict will now have to be resolved going forward.

- (H) One would anticipate that, based on 1-4 above, the indivisible/immutable data/provenance PAIRs would be immediately evident in any DRAFT Implementation Guide.

The current IG DRAFT lacks specification of how these indivisible/immutable PAIRs are captured then conveyed in the exchange artifact (i.e., CDA R2). DPROV and CBCC Leads have stated their intent to allow each implementer to choose their own style (!).

- (I) One would anticipate that, based on DPROV Community analysis showing multiple Provenance Events occurring before the point of exchange, the DRAFT IG would specify how a full chain of trust (evidenced as multiple PAIRS) is conveyed in the exchange artifact.

The DRAFT IG carries no provision for multiple PAIR instances generated by multiple Provenance Event instances.

- (J) One would anticipate that, based on the choice to ignore DPROV Community Use Cases and Requirements, that the DRAFT IG would explicitly inform the reader of this basic fact.

It is suggested that the DRAFT IG offer a simple statement of fact such as:

*NO PART of this Implementation Guide is derived from consensus-based S&I Data Provenance Community Use Case Requirements.*

The current DRAFT IG includes no such statement nor any explanation regarding the disengagement and separate focus of the DPROV Tiger Team (along with CBCC) and the DPROV Community efforts.

- (K) We believe separation of DPROV efforts is confounding and creates a tension that degrades the value and potential benefit/usefulness of both efforts:
- DPROV Tiger Team/CBCC: i) by production of a CDA IG without supporting consensus-based Use Case Requirements, ii) by not starting at the point of origination, iii) by not accounting for multiple Provenance Events and corresponding PAIRs; and
  - DPROV Community: by development of Use Case Requirements starting at the point of data/record origination but saddled with a pre-ordained and now obviously incompatible target CDA IG for exchange.
- (L) We believe the DRAFT IG (for CDA R2) is not only premature but ignores fundamental chain of trust requirements that ensure data provenance PAIRS are properly captured, retained, managed and then conveyed at the point of exchange. We believe the DRAFT CDA IG should be withdrawn until its specification can be fully informed by DPROV Community Use Case Requirements, including chain of trust and traceability to source