The Office of the National Coordinator for Health Information Technology

S&I Data Provenance Initiative

Questions for the HITSC on the S&I Data Provenance Initiative

November 18, 2014

Julie Anne Chua, PMP, CAP, CISSP
Office of the Chief Privacy Officer
Office of the National Coordinator for Health Information Technology
julie.chua@hhs.gov

Johnathan Coleman, CISSP, CISM, CBRM, CRISC
Initiative Coordinator
OCPO/OST/ONC/HHS (CTR)
jc@securityrs.com

Putting the I in HealthIT
www.HealthIT.gov

Agenda

- Data Provenance Initiative Purpose & Goals
- Initiative Progress
- Candidate Standards (work in progress)
- Data Provenance Initiative Use Case Summary
- Questions for the HITSC:
  1) Did the Data Provenance Initiative community miss something that's potentially more impactful?
  2) Where in the Use Case should we start in terms of evaluating standards to meet Use Case requirements?
  3) Are there any architecture or technology specific issues for the community to consider?
- Response/Guidance from the HITSC on Questions

Initiative Purpose and Goals

- To establish a standardized way for capturing, retaining, and exchanging the provenance of health information.
- The community will:
  - Define an initial set of provenance metadata and vocabulary.
  - Create technical specifications to standardize data provenance:
    - At creation (i.e., point of origin);
    - When its exchanged; and
    - When data is integrated across multiple health information systems.
  - Develop guidance for handling data provenance in content standards, including the level to which provenance should be applied.

Initiative Progress

- Achieved consensus on Charter (June 2014)
- Achieved consensus on Use Case (October 2014)
- Participated in development of *HL7 Implementation Guide for CDA® Release 2: Data Provenance (DSTU – September 2014 Ballot)*
- Identified Candidate Standards for consideration during Harmonization Phase

Candidate Standards

Putting the I in HealthIT
www.HealthIT.gov

Note – This list will continue to be updated based on community feedback.

- Cross Enterprise Document-Sharing (XDS)
- Simple Object Access Protocol (SOAP)
- Representation State Transfer (RESTful)
- HL7 Clinical Documentation Architecture Release 2 (CDA R2)
- HL7 IG for CDA R2: Data Provenance – Sep 2014 Ballot
- 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 Record Lifecycle Event Metadata using FHIR (project underway 2014)
- HL7 EHR Records Management and Evidentiary Support (RM-ES) Functional Model, Rel 2
- HL7 EHR System Functional Model Release 2
- HL7 EHR Lifecycle Model (2008)
- ISO/HL7 10781 EHR System Functional Model Release 2 (2014)
- HL7 Digital Signature
- ISO 21089 Health Informatics: Trusted End-to-End Information Flows
- Personal Health Record System Functional Model

Data Provenance Initiative
Use Case Summary

Putting the I in HealthIT
www.HealthIT.gov

Scenario 1: Start Point -> End Point.

Describes simple provenance requirements when transferring healthcare data from a Start Point (sending system) to an End Point (Receiving System).

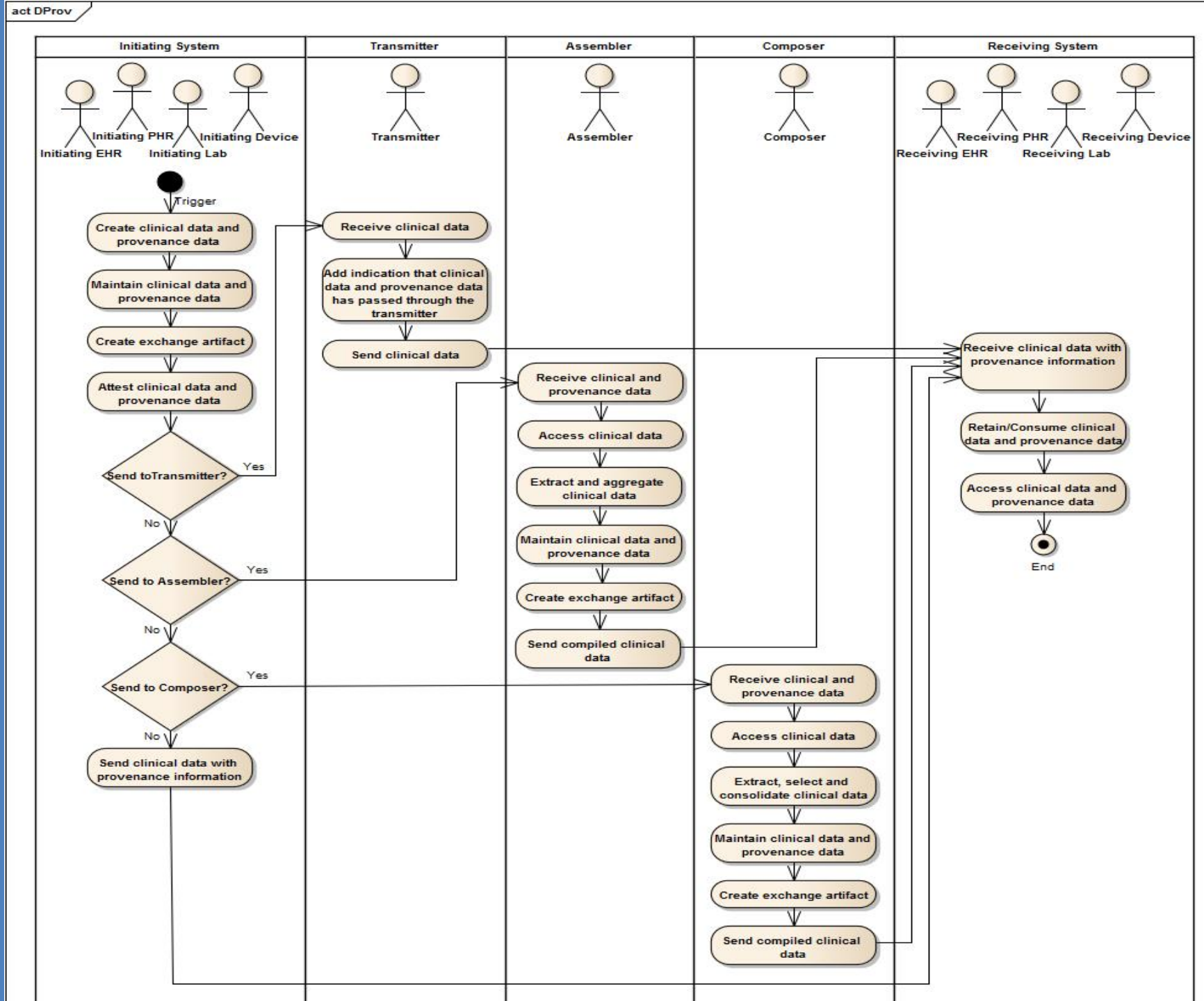
Scenario 2: Start Point -> Transmitter -> End Point.

Includes use of a third party as a conduit/transmitter to transfer information from Start Point to End Point. There may be use cases where it is important to know how the information was routed, as well as who originated it and who sent it.

Scenario 3: Start Point ->Assembler / Composer -> End Point.

Uses a third party system to aggregate or combine information from multiple sources, either in whole or in part, to produce new healthcare artifacts. The new artifacts may contain information previously obtained from multiple sources, as well as new information created locally.

Data Provenance Initiative Use Case Summary



Questions for the HITSC

1) Do the 3 scenarios in the Use Case, and the Use Case's identified scope, address key data provenance areas or is something missing?

   a) Yes, the scenarios address key provenance areas
   b) No, some key data provenance areas are missing

Questions for the HITSC, continued

2) The Use Case is broad and spans a lot of challenges. Where in the Use Case should we start in terms of evaluating standards to meet Use Case requirements?

   a) At the point of data creation in a Patient Controlled Device (PCD) or PHR?
   b) At the point of origin/data creation in an EHR or HIE?
   c) With the transfer of data from a PCD/PHR to an EHR system?
   d) With exchange of data between EHRs?

Questions for the HITSC, continued

3) Are there any architecture or technology specific issues for the community to consider?

a) Content: Refining provenance capabilities for CDA/C-CDA while supporting FHIR?

b) Exchange: Push (e.g. DIRECT), Pull (SOAP and REST-based query responses)?

c) Others?

Response/Guidance from HITSC

- Other questions/comments from the HITSC?
- Thank you

The Office of the National Coordinator for Health Information Technology

BACKUP SLIDES

Slides from Presentation to the HITSC on Data Provenance
September 10, 2014

Putting the I in HealthIT
www.HealthIT.gov

Why do we need data provenance standards?

- Health care providers need confidence in the authenticity and integrity of health data they review/access/receive.

- Ever expanding role for individuals to contribute data toward their health and care through the use of health IT.

- Trends away from documents and toward “atomizing” data.

Challenge

- While there are several existing efforts to address data provenance, no authoritative specification, standard, or model for provenance has been universally adopted to-date, within the context of HIT.
- The variability in how HIEs, EHRs, and PHRs currently capture, retain, and display provenance is problematic for the interoperable exchange, integration, and interpretation of health data.

Initiative Purpose and Goals

- To establish a standardized way for capturing, retaining, and exchanging the provenance of health information.
- What will the community create?
  - Technical specifications to standardize data provenance:
    - At creation (i.e., point of origin);
    - When its exchanged; and
    - When data is integrated across multiple health information systems.
  - 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.

Data Provenance – Phase 1

- The scope of Data Provenance is broad and there are differing perspectives surrounding priorities and expectations for provenance capabilities.
- For Phase 1, we will tackle the following challenges:
  - *(1) When healthcare data is first created, what is the provenance information that should be created and persisted?*
  - *(2) Can a receiving system understand and trust that provenance information?*
  - *(3) Do we need to know who touched it along the way?*
  - *(4) When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?*
  - *(5) When multi-sourced data is assembled and sent to another system, how do we convey the provenance of the multiple data sources as well as for the system doing the assembly?*
    - *Is this considered new data?*
    - *What if the assembling system "cherry picks" from multiple sources, or adds some new health information of its own?*

Initiative Activities:
Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information

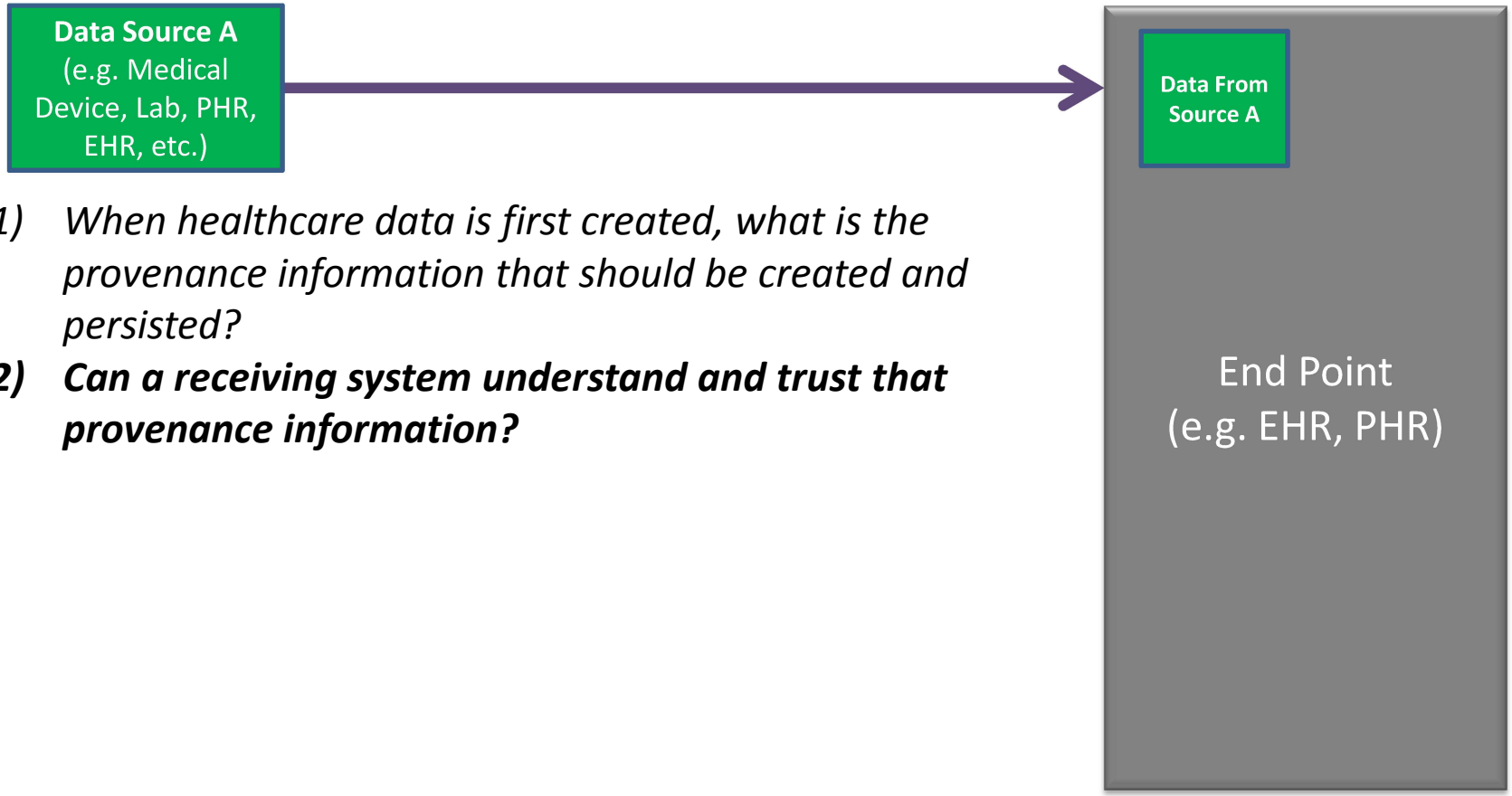
Data Source A
(e.g. Medical Device, Lab, PHR, EHR, etc.)

(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

Initiative Activities: Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information



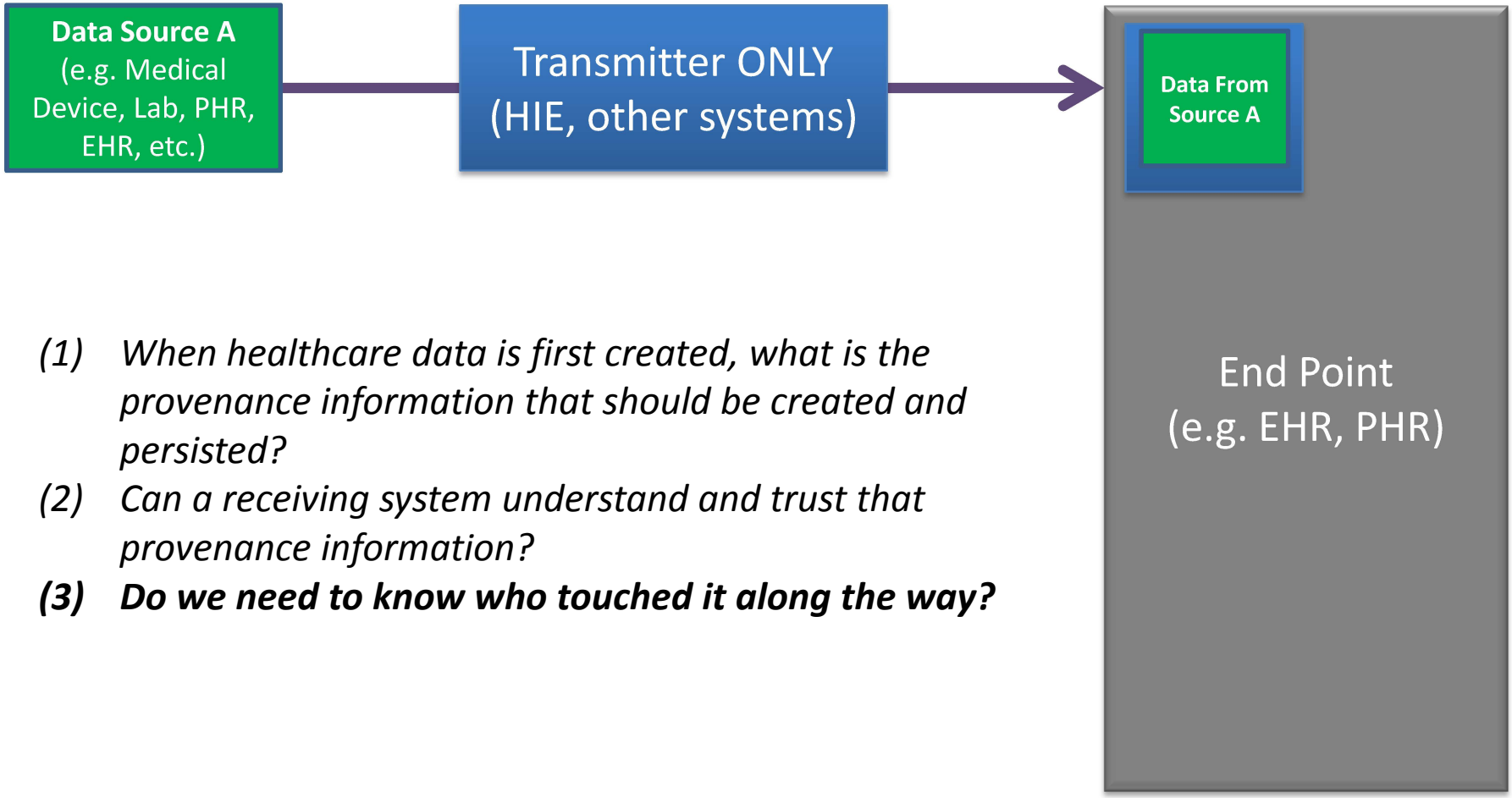
(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

(2) Can a receiving system understand and trust that provenance information?

Initiative Activities: Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information



(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

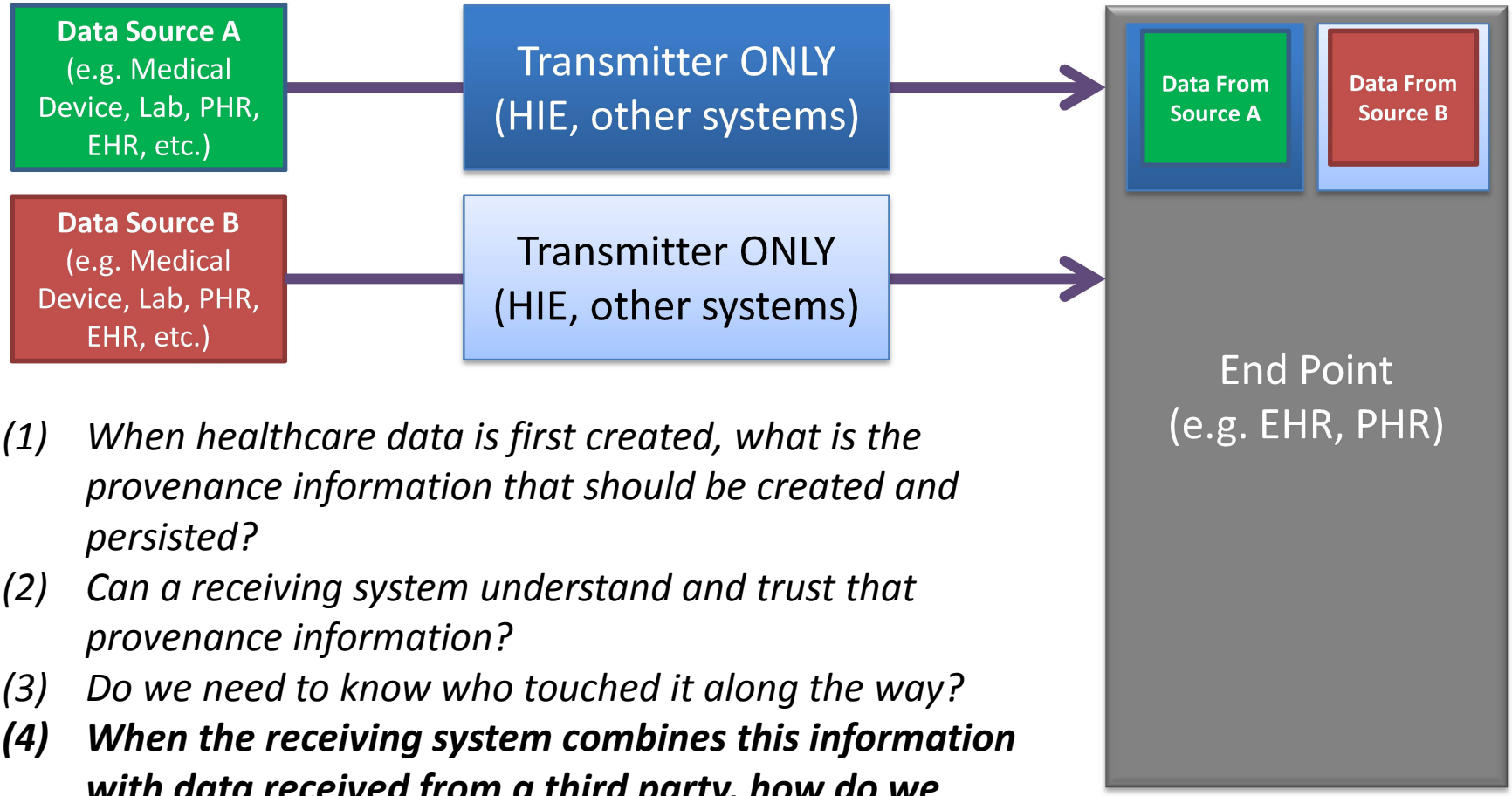
(2) Can a receiving system understand and trust that provenance information?

(3) Do we need to know who touched it along the way?

Initiative Activities: Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information



(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

(2) Can a receiving system understand and trust that provenance information?

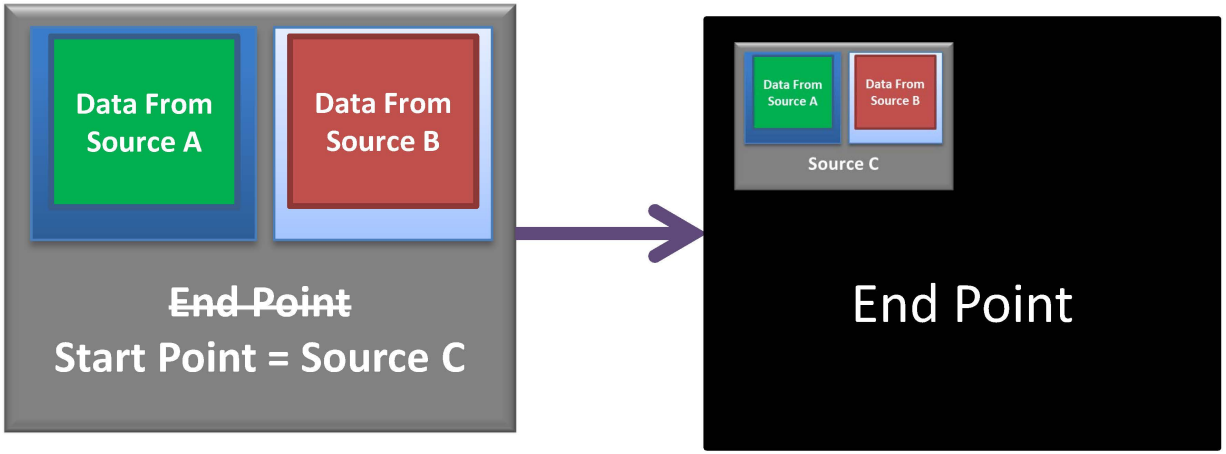
(3) Do we need to know who touched it along the way?

(4) When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?

Initiative Activities: Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information



(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

(2) Can a receiving system understand and trust that provenance information?

(3) Do we need to know who touched it along the way?

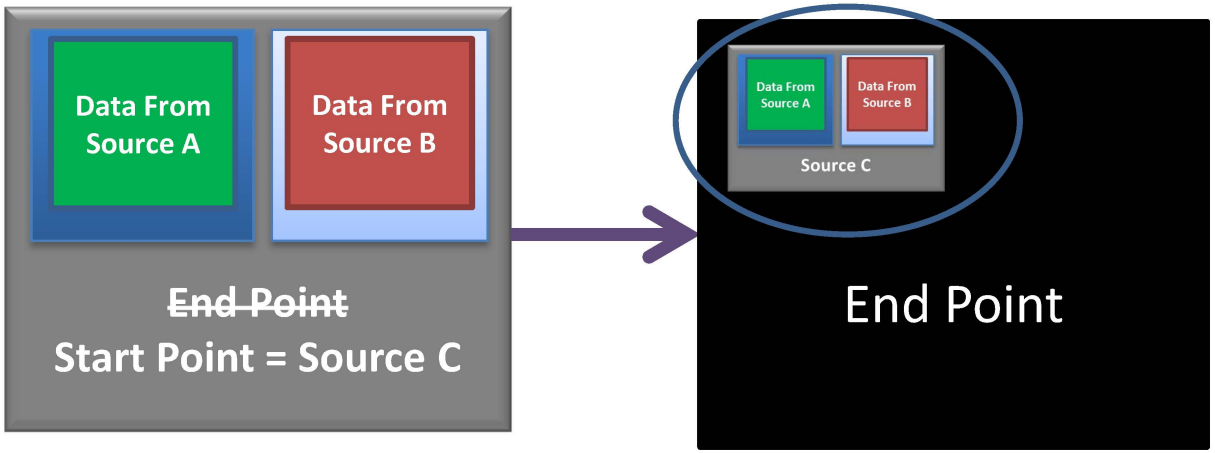
(4) When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?

(5) When multi-sourced data is assembled and sent to another system, how do we convey the provenance of the multiple data sources as well as for the system doing the assembly?

Initiative Activities: Use Case Scenarios (DRAFT)

Putting the I in HealthIT
www.HealthIT.gov

Pre-step : Creation of the data and associated provenance information



- *Is this considered new data?*

(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

(2) Can a receiving system understand and trust that provenance information?

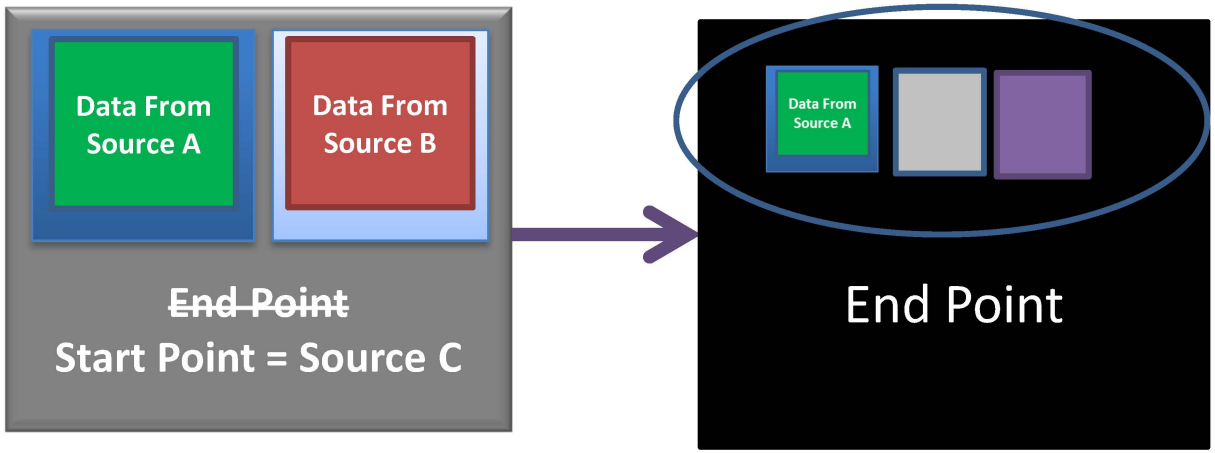
(3) Do we need to know who touched it along the way?

(4) When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?

(5) When multi-sourced data is assembled and sent to another system, how do we convey the provenance of the multiple data sources as well as for the system doing the assembly?

Initiative Activities: Use Case Scenarios (DRAFT)

Pre-step : Creation of the data and associated provenance information



- *Is this considered new data?*
- *What if the assembling system "cherry picks" from multiple sources, or adds some new health information of its own?*

(1) When healthcare data is first created, what is the provenance information that should be created and persisted?

(2) Can a receiving system understand and trust that provenance information?

(3) Do we need to know who touched it along the way?

(4) When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?

(5) When multi-sourced data is assembled and sent to another system, how do we convey the provenance of the multiple data sources as well as for the system doing the assembly?

Initiative Progress

- Achieved Consensus on Charter
- Working on Use Cases
- Formed Tiger Team and proposed the Data Provenance project in HL7:
  - HL7 Implementation Guide for CDA® Release 2: Data Provenance, Release 1
- Worked with other HL7 workgroups on vocabulary harmonization

Initiative Activities: S&I Framework

Putting the I in HealthIT
www.HealthIT.gov

Phase	Planned Activities
Pre-Discovery	• Development of Initiative Synopsis • Development of Initiative Charter • Definition of Goals & Initiative Outcomes
Discovery	• Creation/Validation of Use Cases, User Stories & Functional Requirements • Identification of interoperability gaps, barriers, obstacles and costs • Review of Candidate Standards
Implementation	• Creation of aligned specification • Documentation of relevant specifications and reference implementations such as guides, design documents, etc. • Development of testing tools and reference implementation tools
Pilot	• Validation of aligned specifications, testing tools, and reference implementation tools • Revision of documentation and tools
Evaluation	• Measurement of initiative success against goals and outcomes • Identification of best practices and lessons learned from pilots for wider scale deployment • Identification of hard and soft policy tools that could be considered for wider scale deployments