



The Office of the National Coordinator for
Health Information Technology



S&I Data Provenance Initiative

Presentation to the HITSC on Data Provenance

September 10, 2014

Putting the **I** in Health **IT**
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.

- 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.

- 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.

- 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)



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

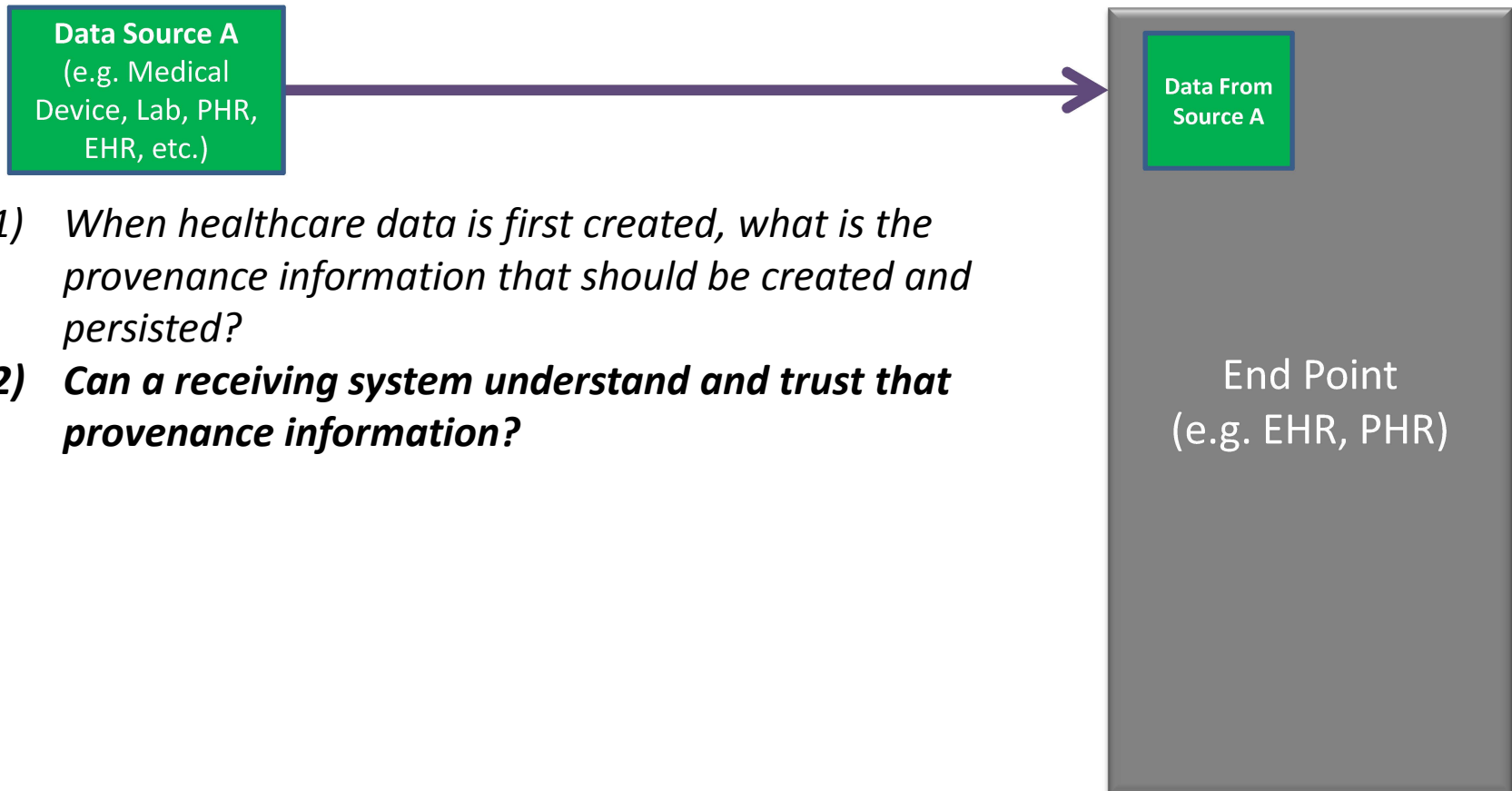
Data Source A
(e.g. Medical
Device, Lab, PHR,
EHR, etc.)

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

Initiative Activities: Use Case Scenarios (DRAFT)



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

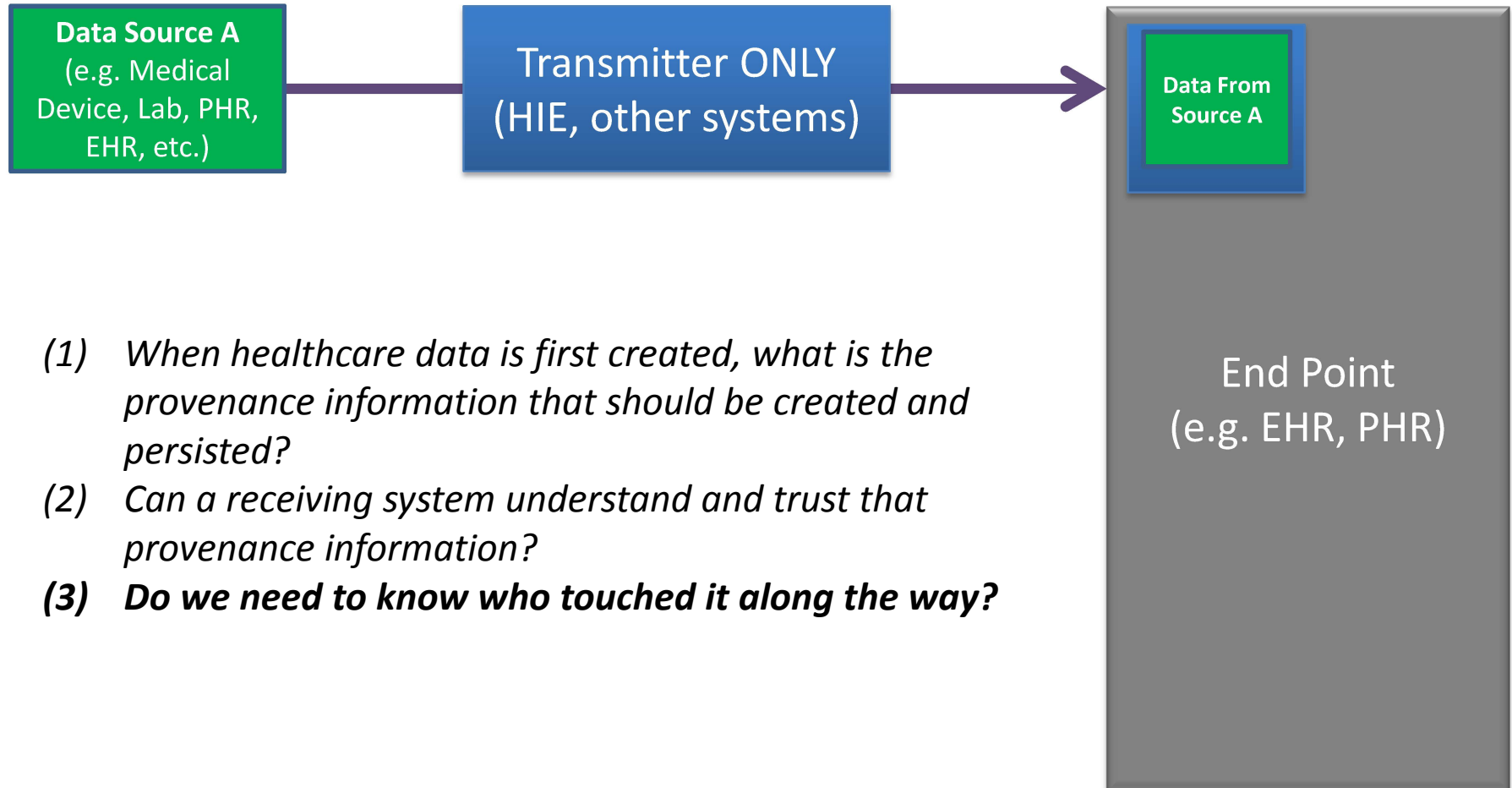


- (1) *When healthcare data is first created, what is the provenance information that should be created and persisted?*
- (2) ***Can a receiving system understand and trust that provenance information?***

Initiative Activities: Use Case Scenarios (DRAFT)



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

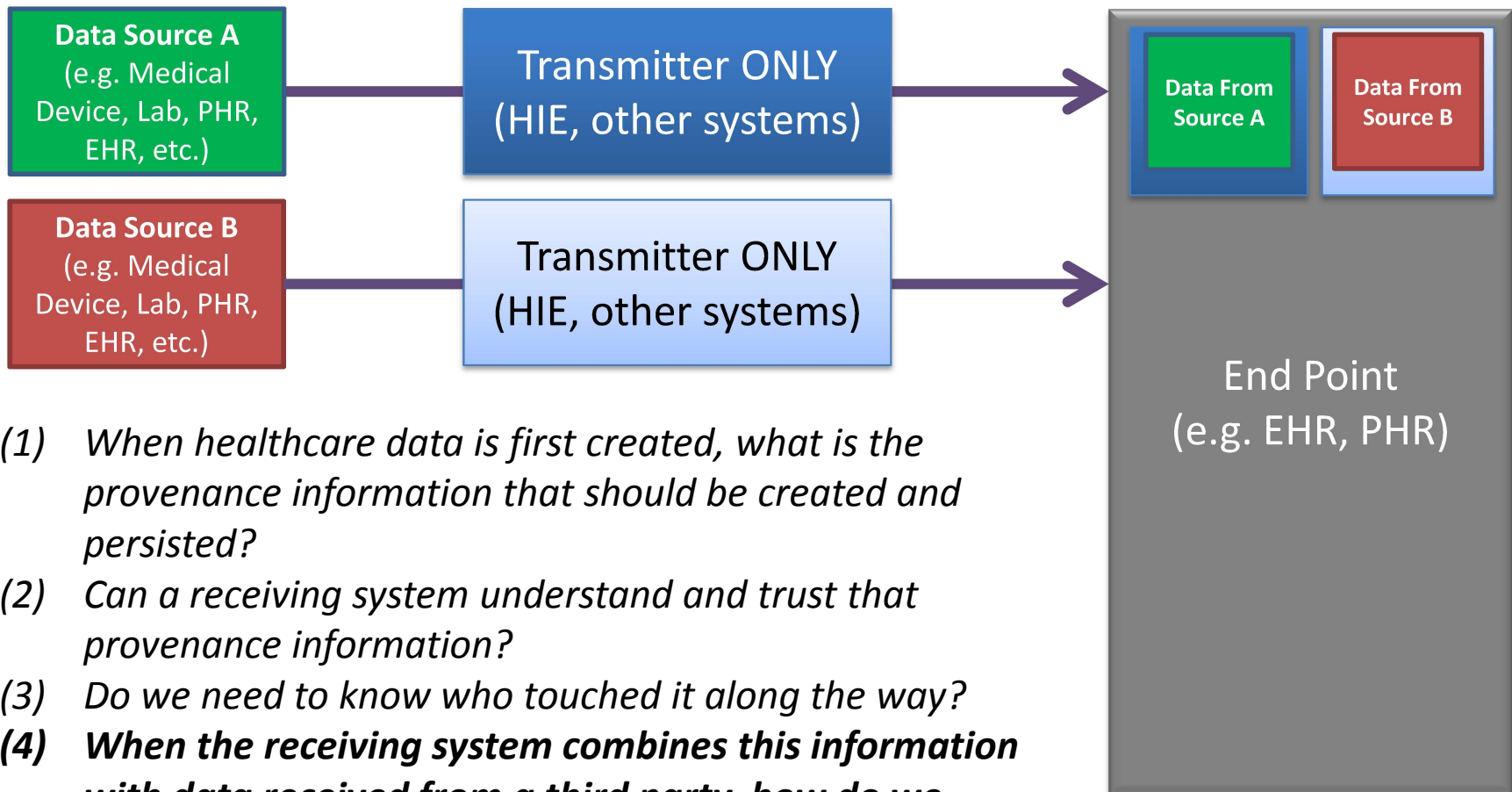


- (1) *When healthcare data is first created, what is the provenance information that should be created and persisted?*
- (2) *Can a receiving system understand and trust that provenance information?*
- (3) *Do we need to know who touched it along the way?*

Initiative Activities: Use Case Scenarios (DRAFT)



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

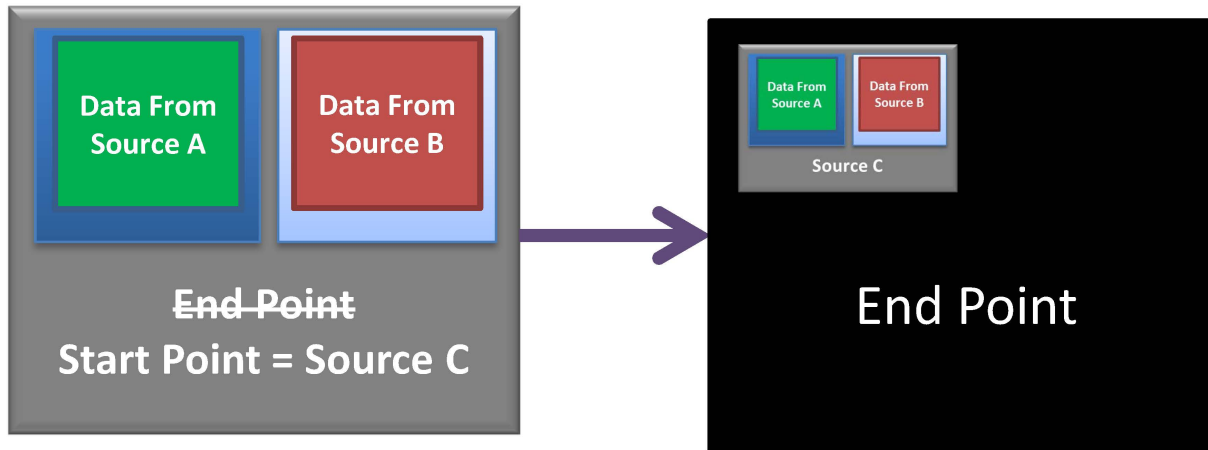


- (1) *When healthcare data is first created, what is the provenance information that should be created and persisted?*
- (2) *Can a receiving system understand and trust that provenance information?*
- (3) *Do we need to know who touched it along the way?*
- (4) ***When the receiving system combines this information with data received from a third party, how do we persist the provenance from multiple sources?***

Initiative Activities: Use Case Scenarios (DRAFT)



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

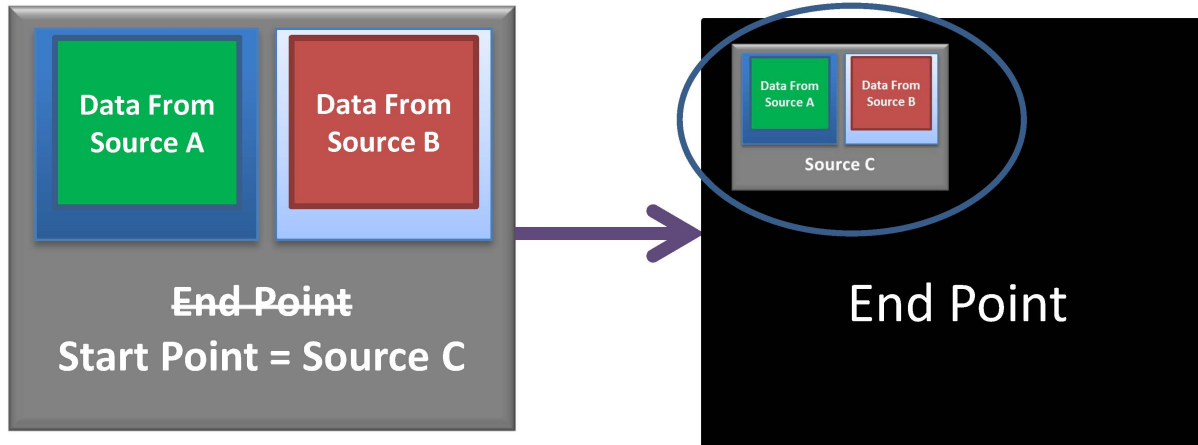


- (1) *When healthcare data is first created, what is the provenance information that should be created and persisted?*
- (2) *Can a receiving system understand and trust that provenance information?*
- (3) *Do we need to know who touched it along the way?*
- (4) *When the receiving system combines this 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)



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



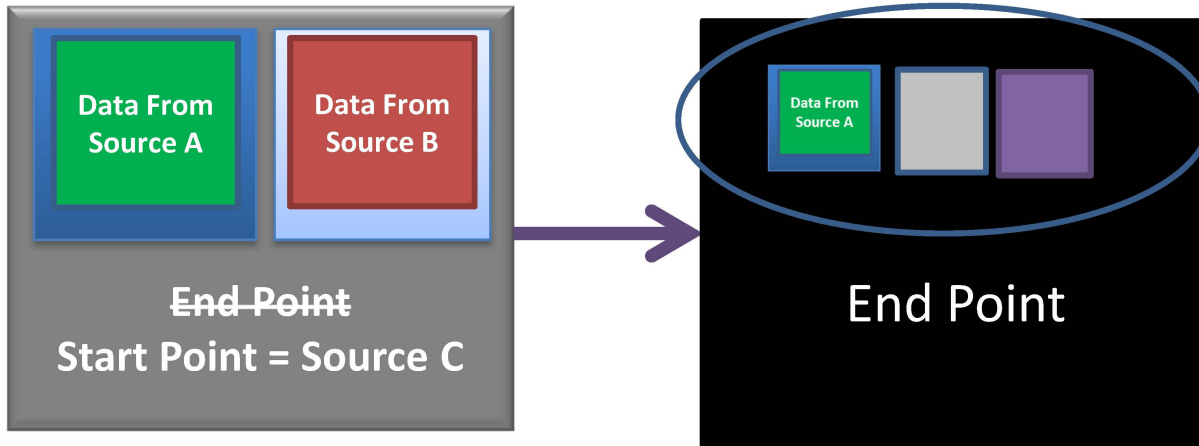
- Is this considered new data?

- (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?***

- 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

Phase	Planned Activities
Pre-Discovery	<ul style="list-style-type: none">• Development of Initiative Synopsis• Development of Initiative Charter• Definition of Goals & Initiative Outcomes
Discovery	<ul style="list-style-type: none">• Creation/Validation of Use Cases, User Stories & Functional Requirements• Identification of interoperability gaps, barriers, obstacles and costs• Review of Candidate Standards
Implementation	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• Validation of aligned specifications, testing tools, and reference implementation tools• Revision of documentation and tools
Evaluation	<ul style="list-style-type: none">• 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