



The Office of the National Coordinator for
Health Information Technology



ONC Privacy & Security Update: Data Provenance (DPROV) S&I Initiative

**NCVHS Committee Meeting
September 22, 2014**

Julie Anne Chua, PMP, CAP, CISSP
Information Security Specialist
Office of the Chief Privacy Officer (OCPO)

Putting the **I** in Health **IT** 
www.HealthIT.gov

Outline

- Definition of Provenance
- Issues and Challenges
- Initiative Goals, Intended Outcomes, and Approach
- Initiative Progress To-Date

What is Data Provenance?

The term “provenance” refers to attributes about the origin/source of health information at the time it is first created and tracks the uses and alterations of the health information over its lifecycle.

Without knowing the provenance of data:

- Healthcare providers may not rely on, or use, information received
- The nation may not achieve the benefits of interoperability and health information exchange

** These address the HIE three-part aim: better care, affordable care, and healthy people and communities*

Why do we need data provenance standards?

- Health care providers need confidence in the authenticity and integrity of health data they review, access, or receive
- The ever expanding role of individuals to contribute data toward their health and care through the use of health IT
- Trends are moving away from documents and toward “atomizing” data

Challenges

No authoritative specification, standard, or model for provenance within HIT

Variability in how HIEs, EHRs, and PHRs currently capture, retain, and display provenance

- Specific health interoperability initiatives guide the design and development of a fully integrated and connected health information system.
- An **S&I Initiative** focuses on a single challenge with a set of value-creating goals and outcomes, and the development of content, technical specifications and reusable tools and services.
- **Call for Participation:** The overall success of the S&I Framework is dependent upon volunteer experts from the healthcare industry



Data Provenance S&I Initiative Goals

- Improve the visibility of the source of, and alterations to, health information
- Improve the confidence healthcare stakeholders have in the authenticity, reliability, and trustworthiness of shared data
- Establish a standardized way for capturing, retaining, and exchanging the provenance of health information

Data Provenance Initiative Intended Outcomes

To achieve these goals, the community will create:

- Technical specifications to standardize data provenance:
 - At creation (i.e., point of origin);
 - When it is 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
- A minimum set of provenance data elements and vocabulary

Data Provenance Initiative Approach: Phase 1

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?*

Data Provenance S&I Initiative Progress

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

Data Provenance

S&I Initiative Progress To-Date

- Initiative Launch – April 2014
- Achieved Consensus on Charter – June 2014
- Working on Use Cases
 - End-to-End Review Comment Period – ended on 9/18/14
- Supported the Data Provenance project in HL7:
 - HL7 Implementation Guide for CDA[®] Release 2: Data Provenance, Release 1 – September 2014 Ballot
- Worked with other HL7 workgroups on vocabulary harmonization

Data Provenance Participation and Wiki

- The ONC Data Provenance Initiative is open for anyone to join.
- We use Wiki pages to facilitate discussion.
 - Information on how to join the Community can be found on the Data Provenance Sign Up Wiki:
<http://wiki.siframework.org/Data+Provenance+Join+the+Initiative>
 - Information on the Project Charter, Use Case, materials, meeting updates, and announcements can be found on the main initiative wiki page:
<http://wiki.siframework.org/Data+Provenance+Initiative>
- In order to ensure the success of our initiative and the subsequent pilot(s), we encourage broad and diverse participation from the community.
- Join us at all-hands community meetings **Every Thursday, from 2:30pm – 3:30pm ET**

BACK UP SLIDES

Initiative Activities: Use Case Scenarios (DRAFT)




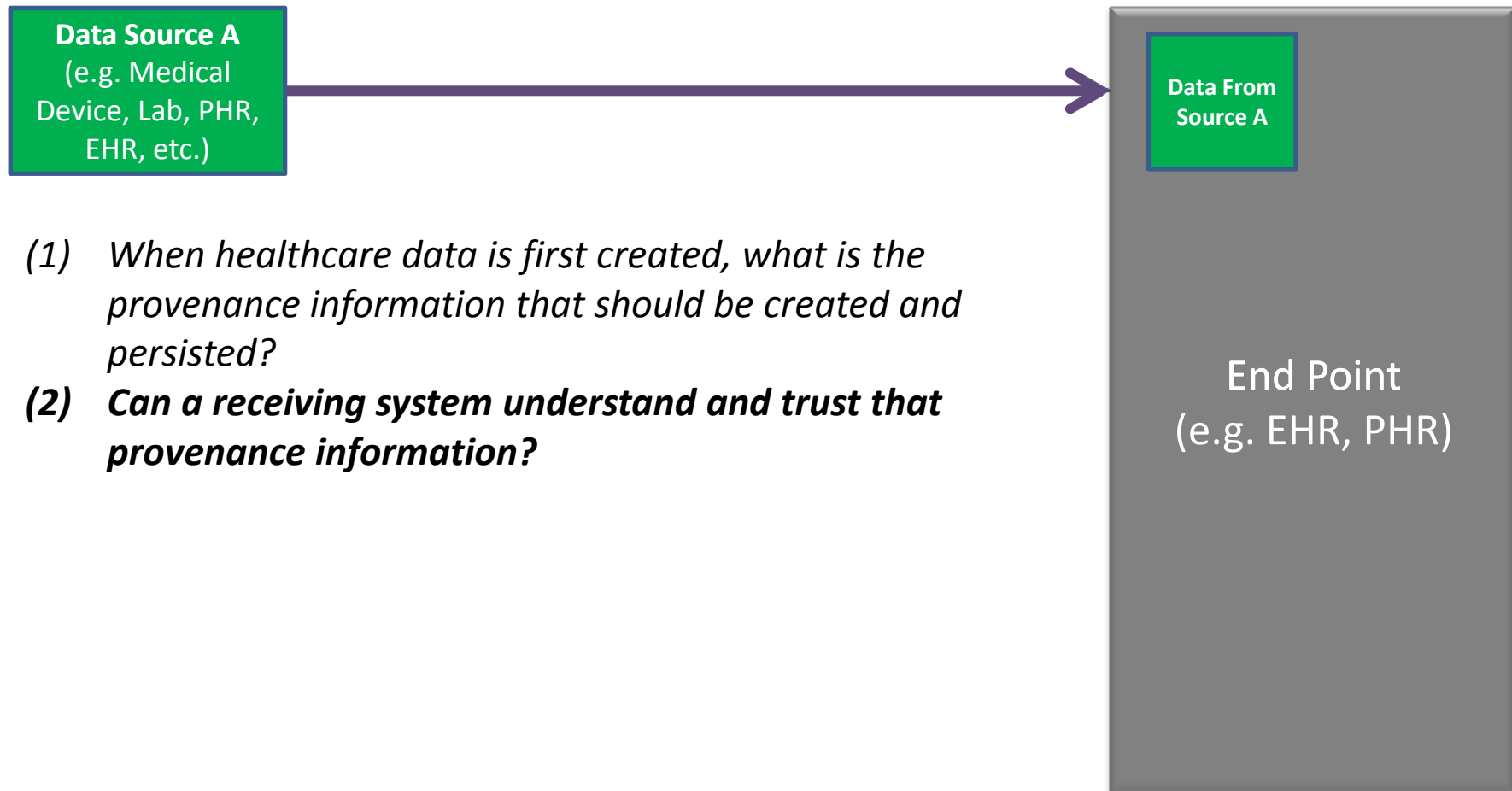
Pre-step : Creation of the data and associated provenance information

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)

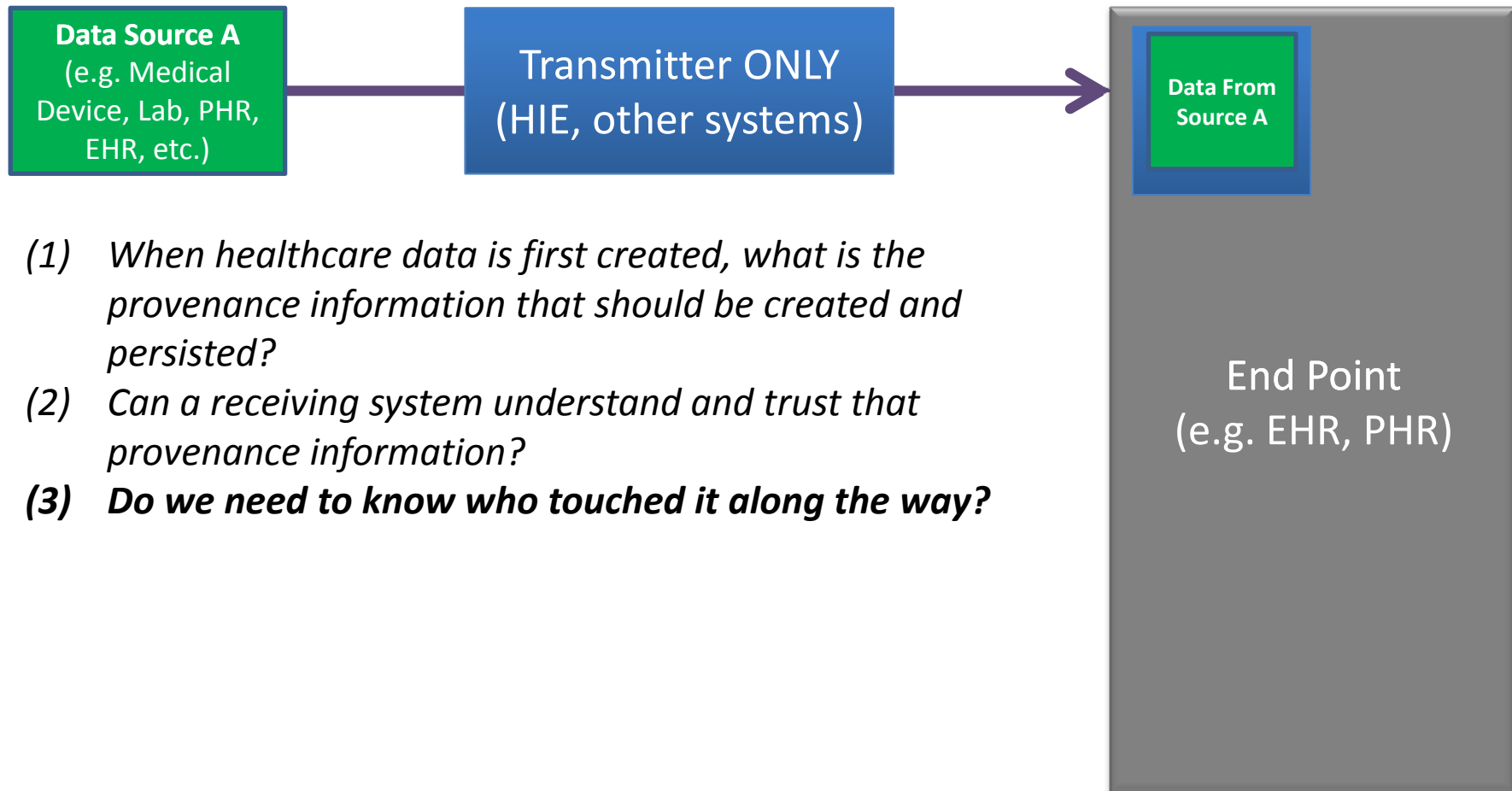
 ***Pre-step : Creation of the data and associated provenance information***



- (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*

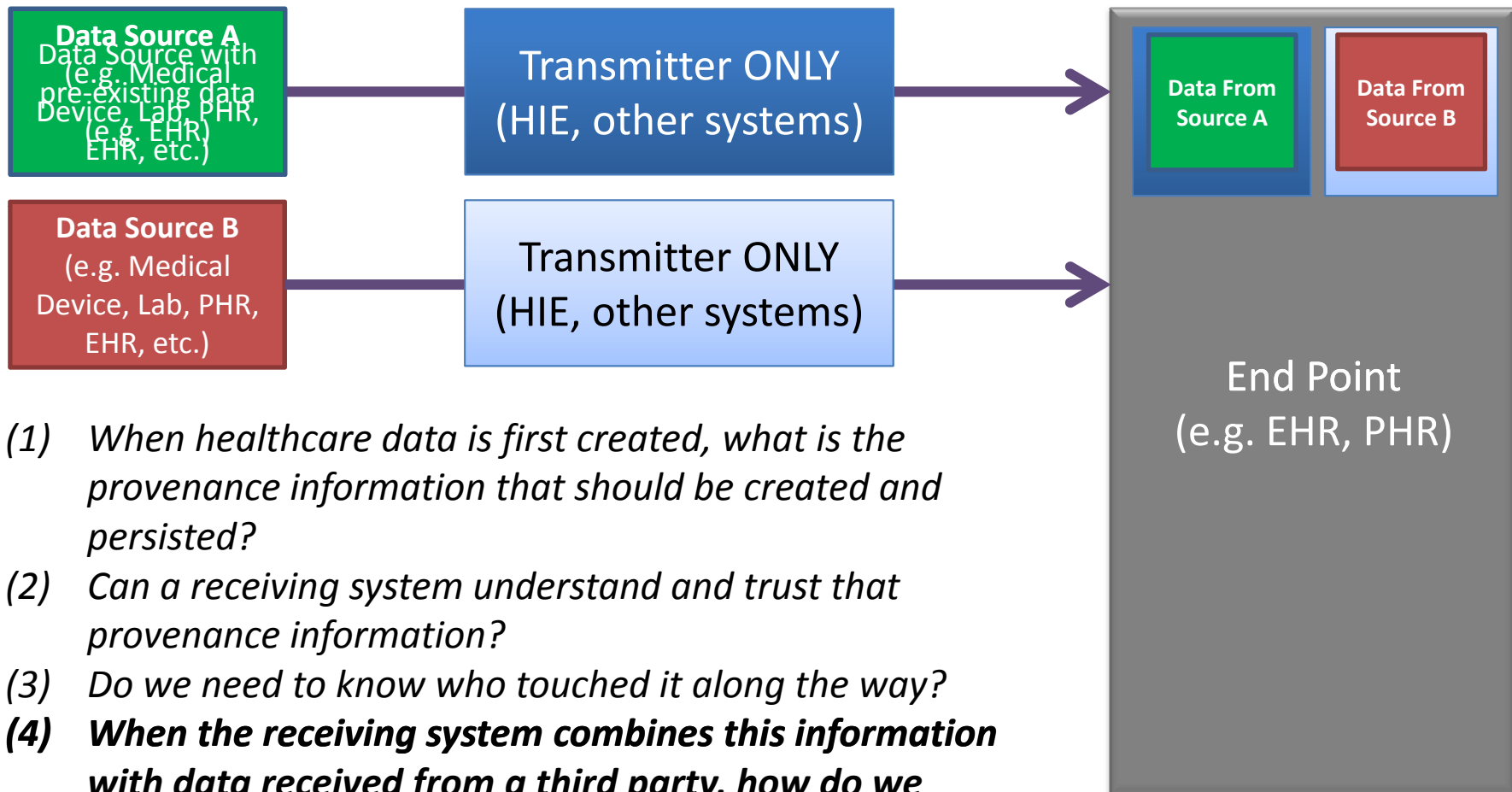


- (1) *When healthcare data is first created, what is the provenance information that should be created and persisted?*
- (2) *Can a receiving system understand and trust that provenance information?*
- (3) *Do we need to know who touched it along the way?*

Initiative Activities: Use Case Scenarios (DRAFT)



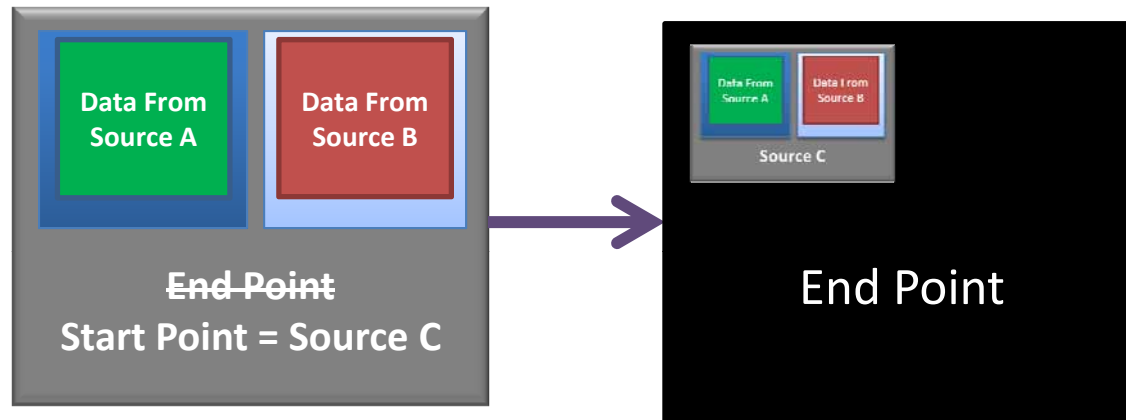
Pre-step : Creation of the data and associated provenance information



- (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)

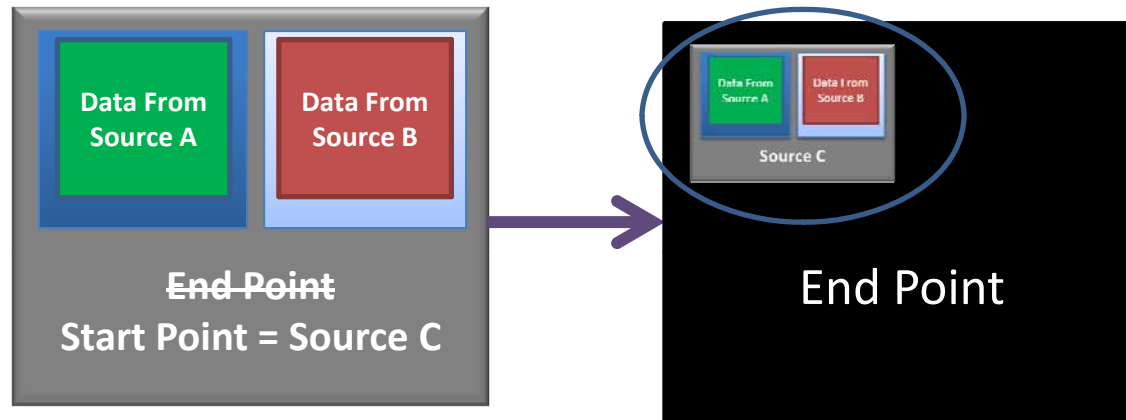
★ *Pre-step : Creation of the data and associated provenance information*



- (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*

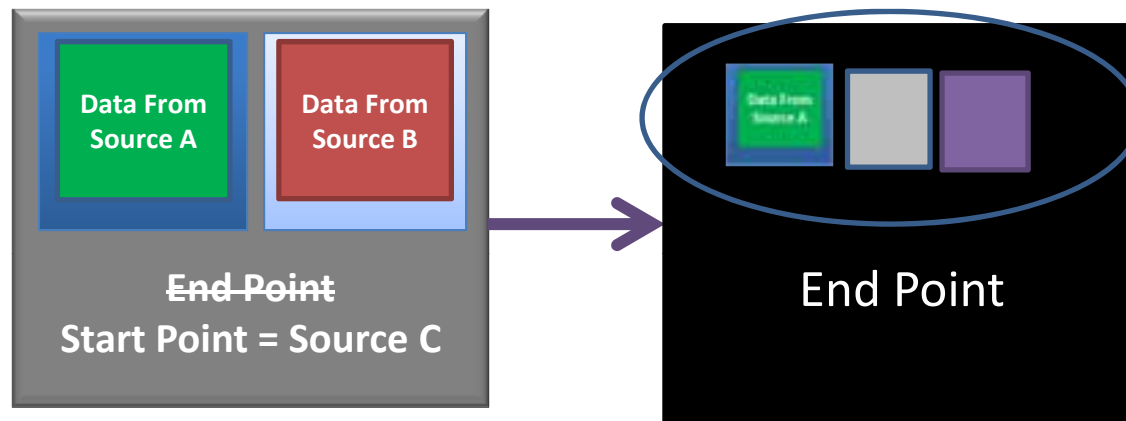


- *Is this considered new data?*

- (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?***