The Office of the National Coordinator for
Health Information Technology

The U.S. Department of Health and Humans Services' Office of the National Coordinator for Health Information Technology (ONC)

Data Provenance (DPROV)

Initiative Summary—September 2017

Table of Contents

Background

Provenance Defined

The term "provenance" in the context of Health Information Technology (HIT) 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 generate, recognize, validate, and appropriately utilize the provenance information. This in turn can lead to uses such as "chain of trust" and "chain of custody" and other business requirements/applications (i.e., records management, evidentiary support and clinical decision support).

Challenge

While there are efforts to address data provenance, and there are several authoritative specification, standard, and models for provenance, these have not been widely adopted to-date within the context of HIT. These include the W3C Data Provenance specifications (PROV Document Family)[1], HL7 EHR Lifecycle Events[2], and ISO/TR 21089:2004 Trusted End-to-End Information Flows[3], which is an intra- and inter-HIT system functional interoperability model. However, while all are foundational, none were designed to support implementation of interoperable, protocol specific provenance supportive standards.

This is exemplified by the variance in how health information exchanges (HIEs), electronic health records (EHRs), and personal health records (PHRs) currently capture, retain, and display provenance. This

[1] W3C. An Overview of the PROV Family of Documents (2017, September 27). Retrieved from http://www.w3.org/TR/2012/WD-prov-overview-20121211/.

[2] ONC Tech Lab Standards Coordination Home: DPROV Home. DPROV Artifacts and Deliverables (2017, September 27). Retrieved from https://oncprojecttracking.healthit.gov/wiki/display/TechLabSC/DPROV+Home.

ONC-SI: Provenance Lifecycle Event Standards	System Functional and Lifecycle Model Presentations	Download Document
Trusted End-to-End Information Flows	ISO 21089 Presentation regarding Trusted End-to-End Information Flows	Download Document
ONC-SI: Provenance-Record Lifecycle	EHR Functional Model Presentation from June 26, 2014	Download Document
ONC-SI: Provenance-Record Lifecycle Events Across SI Initiatives	EHR Record Lifecycle Events and Data Provenance Across ONC initiatives from July 1, 2014	Download Document

[3] ISO. ISO/TR 21089:2004 Health informatics -- Trusted end-to-end information flows (2017, September 27). Retrieved from https://www.iso.org/standard/35645.html. Offers a guide to trusted end-to-end information flow for health(care) records and to the key trace points and audit events in the electronic entity/act record lifecycle (from point of record origination to each ultimate point of record access/use). It also offers recommendations regarding the trace/audit detail relevant to each. It offers recommendations of best practice for healthcare providers, health record stewards, software developers and vendors, end users and other stakeholders, including patients.

variability is problematic for the interoperable exchange (system interoperation), integration, and interpretation of health data.

Until the publication of the HL7 Standard for Trial Use CDA Data Provenance Implementation Guide (HL7 DPROV CDA IG), current health information standards, such as the CDA, lacked guidance for handling data provenance despite being fully equipped to handle provenance of clinical documentation. From its inception, the HL7 FHIR specifications have followed the CDA DPROV IG suit and developed a FHIR specific Provenance Resource[4] based on W3C PROV to support provenance tracking of FHIR Resources. Additionally, the receipt and integration of provenance information has until now been variable and dependent upon system capabilities where neither of the above standards have been implemented. Further challenges are presented if one system can share detailed provenance data but those receiving it cannot process the level of detail exchanged.

Data Provenance (DPROV)

In April 2014, the Data Provenance (DPROV) initiative was launched to establish a standardized way of capturing provenance (including inbound, system generated, and outbound provenance), retaining and exchanging the provenance of health information.

Goals of DPROV

The goals for the DPROV initiative include the following:

- 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.

Health Information Technology Standards Committee (HITSC): Data Provenance Task Force

During the HITSC (a committee established in accordance with the federal advisory committee act—FACA) meeting on November 18, 2014, it was recommended that a Data Provenance Task Force be formed to address the specific question (charge) and three supporting questions from the Office of the National Coordinator for Health Information Technology (ONC):

- Given the community-developed S&I Data Provenance Use Case, what first step in the area of data provenance standardization would be the most broadly applicable and immediately useful to the industry?

Supporting Questions:

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?

[4] FHIR. FHIR Current Build (2017, September 27). Retrieved from http://build.fhir.org/provenance.html.

2. The Use Case is broad and spans a lot of challenges. Where in the Use Case should the Initiative 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

The Data Provenance Task Force convened three meetings to review and discuss the Initiative Use Case and Executive Summary, listen to stakeholder presentations, and produce recommendations to address the Task Force Charge given by ONC. During these meetings, the Task Force heard from different stakeholder perspectives including the Centers for Medicare & Medicaid Services (CMS) and the Electronic Submission of Medical Documentation (esMD) S&I Initiative, records management, patient privacy rights, and the HL7 EHR Functional Model project.

The final Task Force recommendations were presented to the HITSC on January 27, 2015 and can be accessed here.

Methodology

Scope

In order to establish a standardized way of capturing provenance, and standards for retention and interoperable exchange of health information provenance, it is important 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 such as the application of digital signatures to CDA artifacts as specified by the HL7 Implementation Guide for CDA® Release 2: Digital Signatures and Delegation of Rights, Release 1[5].
- Standard metadata tags for data provenance as specified by HL7 normative vocabulary for SecurityIntegrityProvenanceObservationValue value set, which includes codes that differentiate patient from provider asserted/reported information,[6] and syntax provided by the HL7 Privacy and Security Classification System (HC)[7].
- 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.

[5] HL7. Product Brief (2017, September 27). Retrieved from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=375.

[6] SecurityIntegrityProvenanceObservationValue (2.16.840.1.113883.1.11.20485)

[7] HL7. Product Brief (2017, September 27). Retrieved from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=345.

Requirements and Design

The Data Provenance Use Case aims to capture provenance requirements to improve trust in healthcare data and its applications.

This Use Case describes:

- User stories which link the functional capabilities with the business goals and needs.
- The operational context for the data exchange.
- The stakeholders with an interest in the Use Case.
- The information flows that must be supported by the data exchange.
- The types of data and their specifications required in the data exchange.

The Use Case is the foundation for identifying and specifying the standards required to support the data exchange and developing reference implementations and tools to ensure consistent and reliable adoption of the data exchange standards.

To determine the information interchange (i.e., transport layer) and system (i.e., sending and receiving application) requirements as they relate to exchanging and conveying provenance, two sub-workgroups were created and guided by recommendations provided by the Data Provenance Task Force.

These sub-workgroups developed the functional requirements document that provides an in-depth summary of the work that the Information Interchange and System Requirements. Specifically, this document outlines the functional requirements that systems must meet in order to exchange provenance information, provides a short list of the most common health IT standards that support and/or convey the provenance concepts, and addresses the privacy, security, and policy considerations that factor into the methods used to support the provenance requirements. This document is intended to guide pilot project organizations on the functional requirements of data provenance.

Standards Development Support and Standards Development Organization (SDO) Engagement

Candidate Standards List

During the Pre-Discovery efforts, a list of potential standards was identified and listed on the Candidate Standards List for DPROV. The purpose of this exercise was two-fold. First, the effort was a thought exercise to help community members brainstorm on the potential tools that could be brought to bear on the problem DPROV was solving. Second, this effort allowed the DPROV team to analyze the likely SDO stakeholders for DPROV and to construct a communications plan and an SDO engagement strategy that encompassed this group(s).

The sub-workgroups developed a list of standards that are needed to adequately support the functional and data requirements of provenance. Ultimately, pilot organizations will make the final decisions on what standards are used in their scenarios and how they are implemented based on their own organizational needs. The Data Provenance Task Force offered its recommendations, but pilot organizations should not feel overly restricted or constrained in their participation by these recommendations (e.g., focusing solely on EHR to EHR exchanges).

HL7 Engagement

HL7 International Working Group Meetings (WGM) are held three times per year at varying locations. The purpose of these meetings is to give the HL7 WG's a chance to meet face-to-face to work on the standards as well as the opportunity to network with industry leaders from around the world and to provide an invaluable educational resource for the healthcare IT community.

During this step, the DPROV Support team worked with the pilot community and the participating standards development organization (SDO) to develop draft implementation specifications to meet the technical requirements. The HL7 CDA® R2 Implementation Guide: Data Provenance, Release 1 - US Realm was developed as an effort to identify existing opportunities within CDA R2 where basic provenance information about clinical (and other care related information), who created it, when was it created, where was it created, how it was created, and why it was created, can be conveyed in a consistent and interoperable manner. Also conveyed is what action was taken - resulting in (documented by) the information captured, and optimally attested to by the authoring entity, which might be a provider or a HIT system with assembler/composer capabilities. In particular, this IG builds upon the provenance preserving constraints in the Consolidated CDA and Data Segmentation for Privacy IGs, as well as reusing the CDA Consent Directive IG. This IG provides guidance to any CDA R2 implementer on the use of CDA templates to represent data provenance. These templates may also be used as building blocks for conveying provenance using other information exchange Standards. In essence, the CDA DPROV IG created a template of "overlays" to any CDA implementation to assist in specifying the provenance constraints appropriate to the developer's use case and business requirements. This supported a key takeaway for the ONC and HL7 Data Provenance projects: every implementation is workflow specific, and agility is an absolute key for gaining adoption. The CDA DPROV IG offered that agility to CDA implementers. The follow on HL7 Security Work Group's management of the FHIR Provenance Resource seeks to continue that flexibility.

Pilot Activity

In order to provide experience with actual implementations, the DPROV IG was tested or piloted by multiple organizations. The benefits of pilots allowed for the following:

- The ability to leverage initiative resources. Build on the expertise, tools and any open-source code developed through the Data Provenance Initiative to create a better, faster, and higher quality implementation.
- Contribute to the community. Each pilot would have significant visibility with Government agencies, ONC grantees, and within the community of volunteers that support health information exchange.
- Be recognized as an early adopter.
- Use participation in this important national initiative to heighten your organization's name recognition.

The Louisiana Public Health Institute (LPHI) and RAIN Live Oak Network both signed on as pilots to represent the role of the Transmitter; however, due to unforeseen circumstances, both had to withdraw as pilot participants.

As a result, ONC launched the "Oh the Places Data Goes: Health Data Provenance" Challenge. The purpose of this Challenge is to identify the current capabilities and methods used by industry to convey the provenance of health data as it is used to support clinical care. Under this challenge, organizations will be provided the opportunity to win funding to develop and test solutions related to the provenance of health data.

The challenge is conducted in two phases. For Phase 1, participants were required to submit white papers that describe their current capabilities and methods used to demonstrate provenance of health data. In Phase 2, participants are tasked with prototyping, test performances of their solutions, and providing lessons learned.

The winners of Phase 1 are as follows.

- **Hyper E-Health**: This proposal builds on existing standards (e.g., the Sequoia project, the NwHIN), and innovates them using blockchain technology for security and immutability of records.
- **RAIN Live Oak Technology**: This project will demonstrate consistent data provenance in multiple environments into a cloud-based health record, then to a local clinical Electronic Health Record (EHR) via Health Information Exchange (HIE) as well as Patient Generated Health Data (PGHD).
- **1UpHealth**: 1upHealth plans to pilot the use of its partner's provider application to surface provenance information and help providers find aggregated data from various sources using FHIR along with proposed improvements afforded by smart contracts on the blockchain's public ledger.
- **Emrify**: This solution will use a blockchain application stack to enhance data provenance via the use of PHR smart contracts on an Ethereum blockchain.

To learn more, please visit the ONC Challenge website.

Summary

Value of DPROV

Through the development of the HL7 CDA® R2 Implementation Guide: Data Provenance, Release 1 - US Realm, DPROV has improved 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. Benefits to healthcare professionals, healthcare organizations, researchers and research organizations include but are not limited to the following:

- Improve the visibility of permutations of health information from creation to exchange, integration and use across multiple health information systems.
- Improve the confidence healthcare stakeholders have in the authenticity, reliability, and trustworthiness of shared data. Identifying the set of modular components and industry standards

that could be assembled together as valid combinations to promote interoperability for the various business requirements of the community.

Lessons Learned

- In order to produce consensus based implementation guides and standards refinement it is necessary to work in collaboration with SDOs (i.e., HL7). Participation with SDOs is resource and time consuming. Without work effort sponsorship and committed resources, work is unlikely to succeed in producing timely, effective, and adopted interoperability specifications.

DPROV Milestones

- Initiative Launch (April 2014)
- Project Charter Completed (June 2014)
- Use Case Completed (August 2014)
- HL7 Implementation Guide for CDA® Release 2: Data Provenance, Release 1 – US Realm Published (December 2015)
- Oh, the Places Data Goes: ONC Health Data Provenance Challenge Launch (April 2017)
- DPROV Initiative Closes (September 2017)

Appendix A: DPROV Project Deliverables

Foundational Documents			
Document	**Description**	**Download**	**View**
Data Provenance Charter Document	The Charter document is the first document to propose the need for tracking Data Provenance across Health IT. It describes the challenges, scope, stakeholders for this initiative as well as defining the value of the project and the first potential set of data elements needed to ensure security across a medical record's lifecycle.	Download Document	View PDF
Data Provenance Functional Requirements Document	The Functional Requirements document goes in depth defining all of the data elements needed for a successful pilot demonstration of provenance.	Download Document	View PDF
Use Case	The Use Case document reviews the scenarios in which Data Provenance can be applied.	Download Document	View PDF
DPROV General Reference Materials			
Document	**Description**	**Download**	
Data Provenance Initiative Launch	The first meeting for the Data Provenance Initiative held on April 10, 2014. Click here to view the recording.	Download PowerPoint	
Data Provenance Executive Summary	The Data Provenance Executive Summary is a high level overview made to attract pilots, describing the need for Data Provenance and showcasing the initiative's Use Case set	Download PDF	
Data Provenance Glossary	This page provides a list of important terms to the DPROV Initiative	Visit Wikipage	

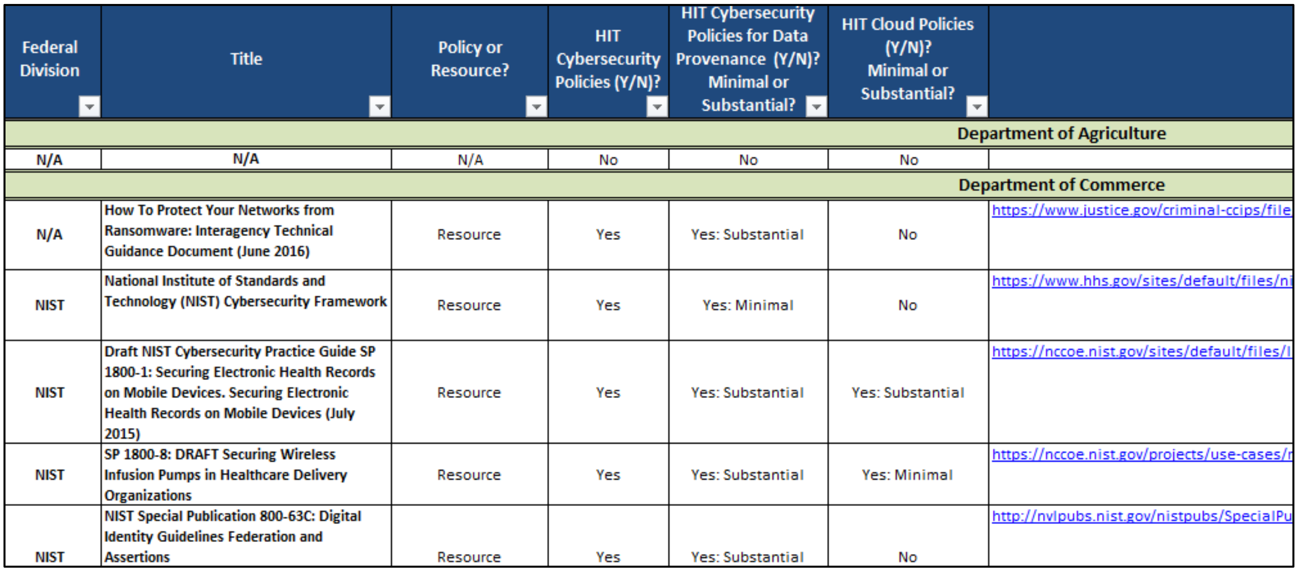
DPROV HL7 Artifacts	This links to the HL7 Data Provenance Wiki page, providing all HL7 artifacts connected to the DPROV Initiative	Visit Wikipage
Health IT Standards Committee: Recommendations to the National Coordinator for Health IT	This page describes the Standards recommendations for ONC	Visit Website
Implementing Interoperable Provenance in Biomedical Research	A document by Vasa Curcin, et al describing how to implement provenance in biomedical domains	Download Document
ONC-SI: Data Provenance - FHIR Provenance Resource	FHIR DSTU Release 1.1 Provenance Resource Presentation from January 16, 2015	Download Document
HL7 Implementation Guide for CDA® Release 2: Data Provenance, Release 1 – US Realm	This IG provides guidance to any CDA R2 implementer on the use of CDA templates to represent the data provenance as well specifying reusable models as building blocks for use in other information exchange standards	Visit HL7 Standards Website
Provenance-FHIR 0.4.0	PDF of the FHIR Provenance Resource	Download PDF
ONC-SI: Provenance Lifecycle Event Standards	System Functional and Lifecycle Model Presentations	Download Document
Trusted End-to-End Information Flows	ISO 21089 Presentation regarding Trusted End-to-End Information Flows	Download Document
ONC-SI: Provenance-Record Lifecycle	EHR Functional Model Presentation from June 26, 2014	Download Document
ONC-SI: Provenance-Record Lifecycle Events Across SI Initiatives	EHR Record Lifecycle Events and Data Provenance Across ONC initiatives from July 1, 2014	Download Document

CentriHealth Comments Actors-Roles-Accountability	CentriHealth comments on Actors, Roles, and Accountability from August 10, 2014	Download Document
CentriHealth Comments Assumptions-Pre-Post-Conditions	CentriHealth comments on Assumptions and Pre/Post-Conditions from August 10, 2014	Download Document
CentriHealth Comments User Story Examples	CentriHealth comments on User Story Examples from August 10, 2014	Download Document

Appendix B: DPROV Cybersecurity Task

The increasing demand to provide access to information through the internet has allowed for rapid increase in web connectivity, instant access, and mobility of information. However, the rise of online information delivery has led to numerous vulnerabilities such as cyberattacks, identity theft, data breaches, ransomware, and so forth. The DPROV initiative examined current federal policies and resources surrounding cybersecurity policies to safeguard against online hacks to the U.S. healthcare system.

These policies and resources were captured in a spreadsheet which provide an overview of all the federal agencies and divisions that currently have HIT cybersecurity policies. The spreadsheet indicates whether the information referenced is a policy or resource; indicates the type of policy; states whether data provenance and the cloud are referenced; states whether the reference is substantial or minimal, and lastly, provides a direct web link.



Federal Division	Title	Policy or Resource?	HIT Cybersecurity Policies (Y/N)?	HIT Cybersecurity Policies for Data Provenance (Y/N)? Minimal or Substantial?	HIT Cloud Policies (Y/N)? Minimal or Substantial?	
Department of Agriculture						
N/A	N/A	N/A	No	No	No	
Department of Commerce						
N/A	How To Protect Your Networks from Ransomware: Interagency Technical Guidance Document (June 2016)	Resource	Yes	Yes: Substantial	No	https://www.justice.gov/criminal-ccips/file
NIST	National Institute of Standards and Technology (NIST) Cybersecurity Framework	Resource	Yes	Yes: Minimal	No	https://www.hhs.gov/sites/default/files/ni
NIST	Draft NIST Cybersecurity Practice Guide SP 1800-1: Securing Electronic Health Records on Mobile Devices. Securing Electronic Health Records on Mobile Devices (July 2015)	Resource	Yes	Yes: Substantial	Yes: Substantial	https://nccoe.nist.gov/sites/default/files/l
NIST	SP 1800-8: DRAFT Securing Wireless Infusion Pumps in Healthcare Delivery Organizations	Resource	Yes	Yes: Substantial	Yes: Minimal	https://nccoe.nist.gov/projects/use-cases/r
NIST	NIST Special Publication 800-63C: Digital Identity Guidelines Federation and Assertions	Resource	Yes	Yes: Substantial	No	http://nvlpubs.nist.gov/nistpubs/SpecialPu