Data Provenance as a Service

Prepared by Craig Burton
September 12, 2016

craigburton.com
@craigburton

Executive Summary

The JLINC Labs Data Provenance as a Service is an open system platform for managing the virtual chain of custody for digital objects. Data Provenance as a Service goes beyond the lineage of data to include origin, ownership, authentication and governance.

Data Provenance as a Service includes the following services in the platform:

- ***The JLINC Link Contract Service***—manages the signing and countersigning of platform transaction.
- ***The JLINC Data Provenance Service***—manages the processes of the platform to ensure the functionality of the system.
- ***The JLINC Ledger***—a blockchain-based ledger that archives information about each transaction in the provenance management of data objects.

Introduction

We are well into the coming of the Information Age. As can be expected, the currency of the Information Age economy is information. The amount of information that has further ensued with the advent of the internet, is overwhelming. As of mid-March of 2016, it is estimated that there some 4.66 billion web pages. This calculation only includes the searchable web and does not include the Deep Web.

Every second, there are 6,000 tweets tweeted; more than 40,000 Google queries are searched; and more than 2 million emails are sent. That's right, every second.

Even more telling is when the communication capacity of the Internet is measured. According to Cisco's Visual Networking Index initiative, the Internet is now in the "zettabyte era." A zettabyte is one sextillion bytes. It is estimated that by the end of the year, the Internet will reach 1.1 zettabytes per year, and that by 2019, global traffic is expected to hit 2 zettabytes a year. To give you some idea how big this is, one zettabyte is the equivalent of 36,000 years of high-definition video.

Needless to say, information and its growth in this era is big, and valuable. Of course, ownership, usage, and the process of managing this process has become an acute problem. As a result, the idea—and need—of a Data Provenance as a Service has emerged. It is critical that this service be independent, simple and usable by both humans and programs.

The JLINC Data Provenance Services meet these requirements and are a complete system for managing the provenance of information objects of any type and size.

Data Provenance Service Architecture

The Architecture Detail

JLINC is a complete Data Provenance Service. It can be used by customers in the cloud, on premise, or as a hybrid of the two. Figure 1 shows a typical scenario using the architecture.

The JLINC Data Provenance platform consists of three core services:

- The JLINC Link Contract Service
- The JLINC Data Provenance Service
- The JLINC Ledger

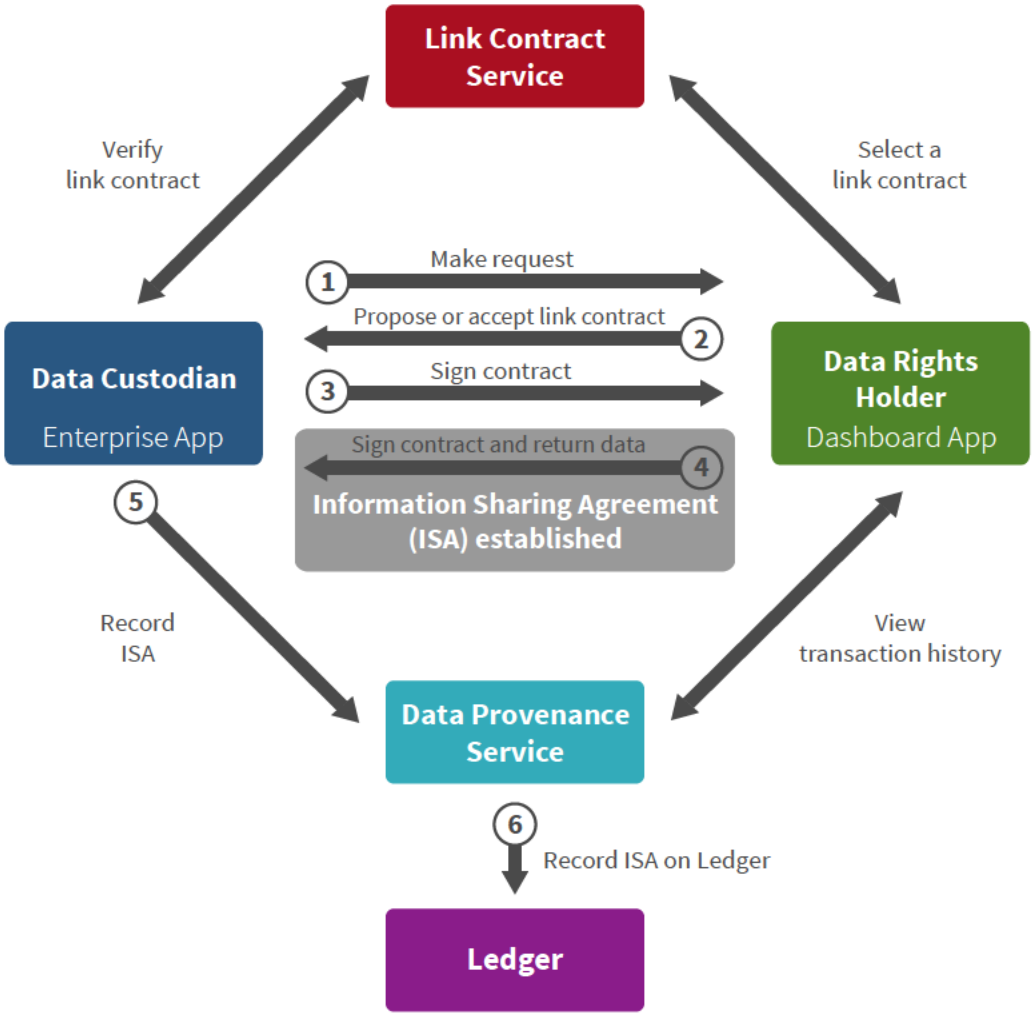

Figure 1—The JLINC Data Provenance Architecture

The DPaaS Transaction

The objective of the JLINC DPaaS is to facilitate private data transactions between two or more parties, typically a data requestor (The Data Custodian) and a transaction responder (The Data Rights Holder).

The transaction process concludes with a signed—and counter-signed—agreement between the parties. This completed agreement is referred to as the Information Sharing Agreement (ISA). (See Figure 1)

The JLINC Ledger

The JLINC Ledger is a blockchain-based immutable ledger. JLINC DPaaS creates a permanent record of every transaction on the ledger. The transaction record includes a cryptographic hash of the ISA signed with the cryptographic public keys of all parties to the transaction.

This design provides an irrefutable transaction record. The irrefutable record can be used to settle any dispute or question that arises between the parties in the transaction. This process makes it very difficult—if not impossible—to change or claim differences from the original ISA between the Custodian and the Data Rights Holder.

Although the record is "public", all the public can actually see is that an agreement was signed with the public keys belonging to the parties to the agreement at a certain time. Only the parties themselves can use the hash recorded on the Ledger to prove that the ISA they (or their Provenance Service) hold is valid and was in fact signed by the other party (or parties).

JLINC currently uses the Stellar Consensus Protocol, a permissioned blockchain technology, as the basis of its Ledger service. Stellar—like JLINC—is an open source project. Although JLINC DPaaS is currently based on the Stellar Ledger, the unique JLINC design allows for other blockchain-based systems to be used as needed by the customer.

The JLINC Link Contract Service

There are most commonly two participants in the Data Provenance Platform process:

- **The Data Rights Holder**—the entity that created or possesses author's rights to the data included in the transaction.

- **The Data Custodian**—an entity that receives data after having entered into an agreement (an ISA) as to the confidentiality and usage of that data.

The Data Rights Holder

The Data Rights Holder is the entity that has the rights and privileges to the data being included in the transaction. The Data Rights Holder is responsible for making available the terms under which it will share data, and for asserting that it does have author's or creator's rights over the data.

The Data Custodian

The Data custodian is the entity being granted certain rights and privileges to the data being originated by the Data Rights Holder. The Custodian is responsible for handling data as specified under the ISA. The ISA governs the data defined in the transaction.

The JLINC Data Provenance Service

The JLINC DPaaS platform keeps a record of completed ISAs between the two parties. There is an additional need for Data Custodians to be able to view the provenance of the data they receive (i.e. "where did this data come from?").

At the same time, the Data Rights Holder needs the ability to see the chain of custody of the data (i.e. "where has my data gone?").

The Data Provenance Service keeps a living record of the data movement under the ISA and makes the information available to the parties involved in the transaction.

The JLINC Apps

There are two core applications in the platform:

- The Enterprise App
- The Dashboard App

The Enterprise App

An App or "plugin" comes pre-configured with the data handling agreements under which data can be accepted. The Enterprise App then interacts as a software agent with data rights holders or their agents to negotiate and complete ISAs and properly record them. JLINC Labs currently offers a Salesforce.com app and plans plugins for other popular CRMs (e.g Oracle Sales Cloud CRM, SugarCRM, SAP, and Microsoft Dynamics CRM).

The Dashboard App

The JLINC Dashboard App provides an easy to use interface for end users to pre-select from among standard terms, set preferences and manage data sharing. The Dashboard App then acts as a software agent to negotiate, sign and record ISAs on behalf of the user.

Dashboard Apps will be offered for the Mac, Windows and Linux desktops, with mobile apps on the iOS and Android operating systems. An introductory version of the Dashboard runs in all standard web browsers.

Example Sequence of a Transaction

Figure 1 shows an example transaction broken down into six distinct events:

1. The data custodian makes a request to the data rights-holder.

2. The rights holder responds with one or more standard terms for personal data sharing. The rights holder indicates which terms are acceptable (in order of preference if more than one agreement is presented).

3. The data custodian cryptographically signs the agreement of choice and sends that signed agreement—along with the custodian's public key—with a request for the data.

4. If the agreement is acceptable, the rights holder counter-signs the agreement and sends the counter-signed agreement—along with the rights-holder's public key and the data—to the data custodian. This step constitutes the completed Information Sharing Agreement (ISA). If needed, the rights holder can encrypt the data included using the custodian's public key.

5. The data custodian sends a copy of the ISA to the data provenance service as designated (in the ISA) by the data rights holder.

6. The data provenance service records the public keys of the rights holder and the data custodian with a cryptographic hash of the ISA as a transaction on the ledger. The data provenance service also allows all parties in a transaction to look up and view the history of the movement of the rights holder's data.

Many other use cases are also possible, among two or more parties, including rights holder initiated transactions, one data custodian pushing data to another data custodian, or to a rights holder, to name a few. JLINC DPaaS transactions currently operate over http, however other transport protocols could also be used. This design makes the programmatic processes and interfaces to the service both simple and ubiquitous.

Conclusion

As the Information Age continues to grow in size at an astronomical rate, the need for automated systems capable of managing the provenance of data becomes paramount. Further, such systems need to be flexible and well-designed enough to meet the needs of organizations now and in the future as things will inevitably evolve.

JLINC Data Provenance as a Service is the first complete programmatic platform designed to manage the chain of custody of any data element. It's unique design fits ideally with organizations that need a cloud, on premise or hybrid design to manage the provenance of data both now and in the future.

Glossary

JSON-LD
The data-graph standard used by JLINC to represent Link Contracts, Data and ISAs. It has been a W3C standard since Jan 2014. https://www.w3.org/TR/json-ld/

Public Key
One of the two keys in a standard asymmetric cryptography system. It can be used to prove with mathematical certainty that a signature was created with the corresponding private key. See https://en.wikipedia.org/wiki/Public-key_cryptography

Link Contract
The part of an ISA containing the link to agreed data sharing terms, metadata about the parties to the agreement and the data being referenced, expressed in JSON-LD.

Information Sharing Agreement
An ISA is a Link Contract along with signatures by all parties, and optionally the actual data itself, all expressed as a JSON-LD document as described in the JLINC protocol specification. https://github.com/JLINClabs/JLINC-spec

Data Rights Holder
The individual or entity that has author's or creator's rights over data to be shared.

Data Custodian
An individual or entity that enters into an agreement (an ISA) to be responsible for the correct handling of data being shared.

Data Sharing Terms
The terms under which a rights holder is willing to share data with a data custodian. Standard terms can be hosted by trusted 3rd parties at known URL's.

Ledger
A publicly readable, immutable, distributed global database, i.e. a permissioned blockchain. JLINC currently uses the Stellar Consensus Protocol. https://stellar.org