Tracking BioSamples on Hyperledger Blockchain

Carlos Castro-Iragorri
Alexander García
Richard Shute

www.pistoiaalliance.org

Hacking team

BioHackathon Paris, November 2018

Federico Lopez (EMBL-EBI)
Claus Weiland (Senckenberg)
Isabelle Perseil (INSERM)
Carlos Castro-Iragorri (Urosario)

The growing need for provenance in the supply chain of biomedical “samples”

- Genetic samples
- Screening samples
- Reagents
- Tissues & Cells
- Animals
- Active Pharmaceutical Ingredient (API)
- API Precursors
- Clinical Trial Samples
- Packaged Medicines
- Samples from Patients
- Etc.

The Medicines’ Value Chain

BioSamples as surrogate for other biomedical samples

- BioSamples stores and supplies descriptions and metadata about biological samples.
- Industry and academic
- Search, submit, and curate
- Useful, relevant source of data on some key biomedical sample types
  - Mainly tissues, cell lines etc. for genomic research
  - Surrogate for other types of biomedical sample

BioSamples : Opportunity

- Use blockchain technology to support provenance and sample supply chain management
- Room to augment functionality
  - Enable better sample tracking and provenance: Through a workflow or supply chain
  - Provide additional layers of security to the identifier
- Include the protocols used to produce the samples.

Blockchain

Blockchain is a collection of technologies (cryptographic security, decentralization, digital registry, smart contracts, rules and incentives to collaborate among institutions with different levels of trust) .

OpenScience requires provenance, transparency, and availability.

Blockchain delivers the trust layer in OpenScience.

Blockchain approach:

- BioSamples already provides a standard for submitting samples and links to the information; this is a digital asset.
- Well defined protocols used to produce the samples can be seen as operations that spawn new samples; transactions acting on the digital assets.
- Blockchain(s) that registers interactions on the sample repository.
- Provide secure access to samples and protocols through permissioned channels across the network

Hacking project objectives

1. Identify the added value of Blockchain technologies on BioSamples and similar sample management tools.
2. Create a prototype on Hyperledger Composer of a Business Network (BNA) that includes the data standards and current functionalities (workflows) of BioSamples.
3. Smart contract to track provenance in the curation process.
4. Suggest possible extension in order to implement different layers of curation and generate different states of the samples.

Tasks (day 1)

Blockchain technologies and Biosamples (data and workflows)

1. Tutorial on Blockchain technologies (9-11)
2. Tutorial on Hyperledger Composer (11-12, 13-15).
3. Sample use cases: Blockchain4openscience and Digital Diplomas.
    1. Skills: JS.
    2. Output: learn Composer-Playground.
4. Understanding BioSamples: data management and workflows (15-19).
    1. Skills: BioSamples.
    2. Output: Identify data structures and workflows.
5. How would BioSamples benefit from blockchain technologies: **track provenance in curation.**

Tasks (day 2)

Supply chain management in BioSamples (9-12)

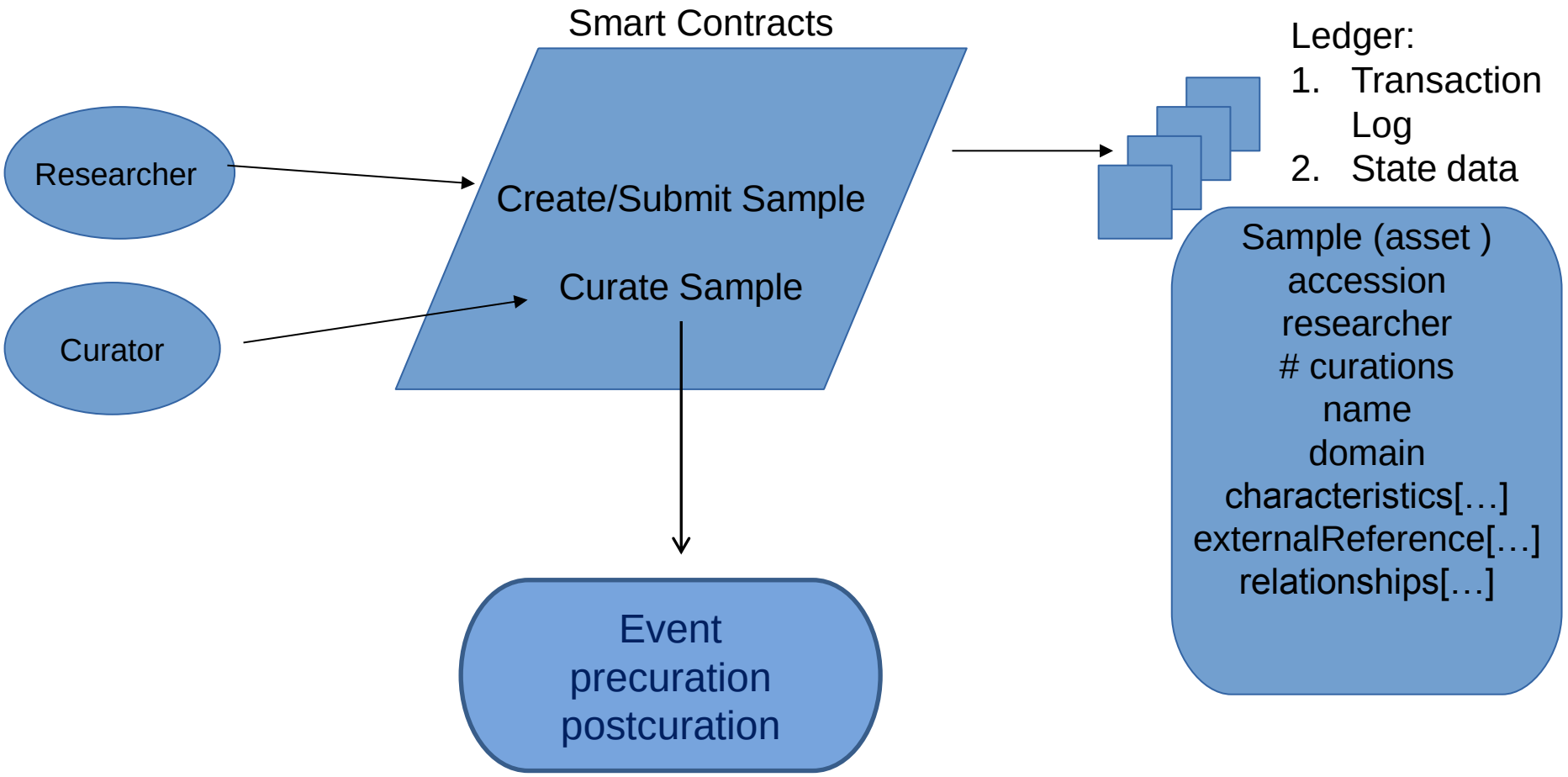
1. Designing and implementing a business network application (biosamples.bna) on Hyperledger Composer for the project:
2. Define assets, participants, transactions and events in model file (biosamples.cto).
    1. Output: BioSamples.bna
    2. Example for sample submission (create asset)
3. Report.

Tasks (day 3 and 4)

Supply chain management in BioSamples

1. Code transactions logic in JS for curation. Update and document bna with new curation functions.
    1. Skills: JS.
    2. Output: BioSamples.bna
2. Extensions.
3. Writing-up a presentation for the project.
4. Documenting (GitHub repo) and roadmap.
5. 1 page report for writethon.

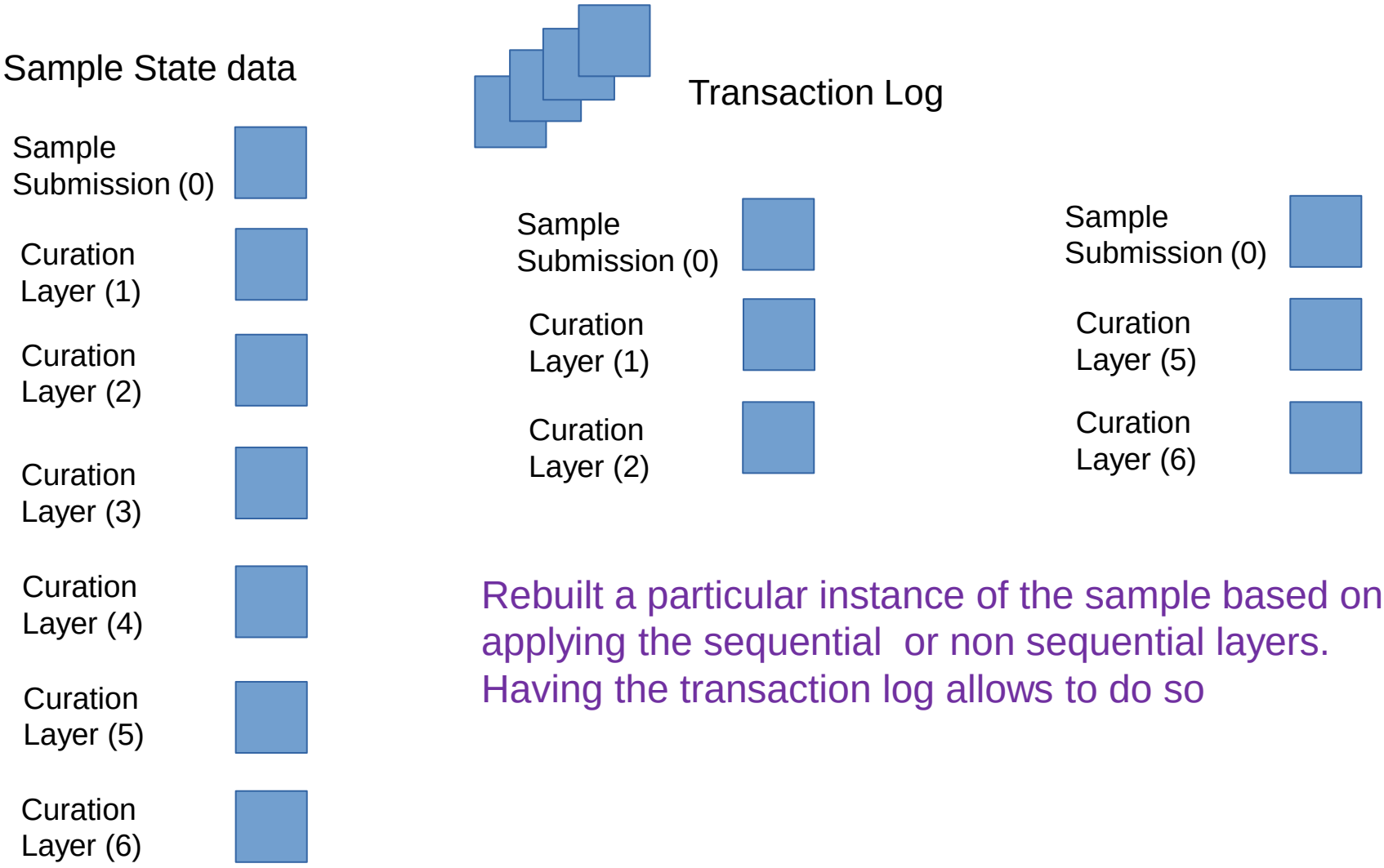
Business Network, biosamples.bna (playground)



BioSamples and Blockchain

- Blockchain provides various provenance features to sample management:
- Track researchers interaction with the samples.
- Relationships between samples.
- Track the curation process (layers) using the two levels of the Ledger: immutable transaction log and state data (last view of the sample).
- Predefined (filters) and dynamic queries.
- No immediate benefit form decentralization.

Extension: Rebuilt a particular state, based on applying a number of curation layers.



Rebuilt a particular instance of the sample based on applying the sequential or non sequential layers. Having the transaction log allows to do so