Comply Advantage

RoPython Meetup
20 June 2019

ComplyAdvantage.com

About ComplyAdvantage

ComplyAdvantage embraces a data-driven, technological approach to fighting crime as a way to **empower compliance**, helping firms to understand who they are doing business with, and ultimately make decisions faster. Practically, that approach means offering **easy access to a constantly-updated, multifunctional suite of data**, and the technology needed to exploit it effectively.

Our platform includes sanctions lists, transaction and adverse media monitoring tools, global AML databases and more, **integrated seamlessly** with existing tech infrastructure, and navigable with intuitive, user-friendly APIs. We focus on regulatory **efficiency, flexibility, and usability**, as a way to demystify the compliance function, and unlock its potential for every type of financial services firm, from fledgeling start-ups, to established international banking firms.

Comply Advantage

Do you trust your Data?

Comply Advantage

Why?

*We have entered the **Age of Required Knowledge**. With all the data available to us, internally and externally, employees and executives are expected to **know.** There is a **cost of knowing**. Getting caught not knowing could lead to sensational news headlines accompanied by loss of shareholder value, loss of customers, and even industry fines.*

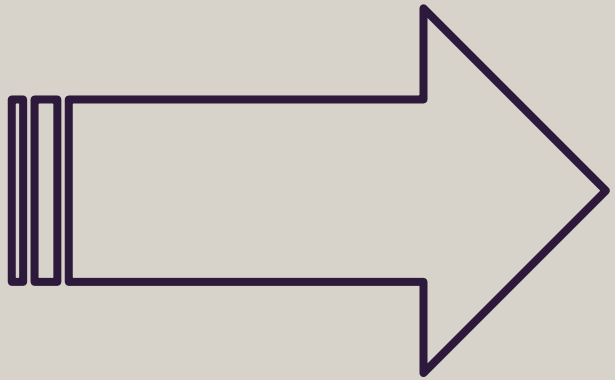
Comply Advantage

Stats

- *By 2020, there will be around 40 trillion gigabytes of data* ***- EMC***
- *90% of all data has been created in the last two years -* ***IBM***
- *Internet users generate about 2.5 quintillion bytes of data each day -* ***Data Never Sleeps 5.0***
- *By 2020, every person will generate 1.7 megabytes in just a second -* ***Domo***
- *In 2012, only 0.5 of all data was analysed -* ***The Guardian***

Data Characteristics

- Accuracy
- Currency & Timeliness
- Correctness
- Consistency
- Usability
- Security & Privacy
- Completeness
- Accessibility
- Accountability
- Scalability



- Authenticity
- Trustability
- Traceability
- Confidentiality
- Integrity
- Availability

Comply Advantage

The Core

Audit

- Exists to provide step by step documented history of a transaction/operation
- Is used to detect when a system is not working properly
- Focused mainly on operations, transactions, events

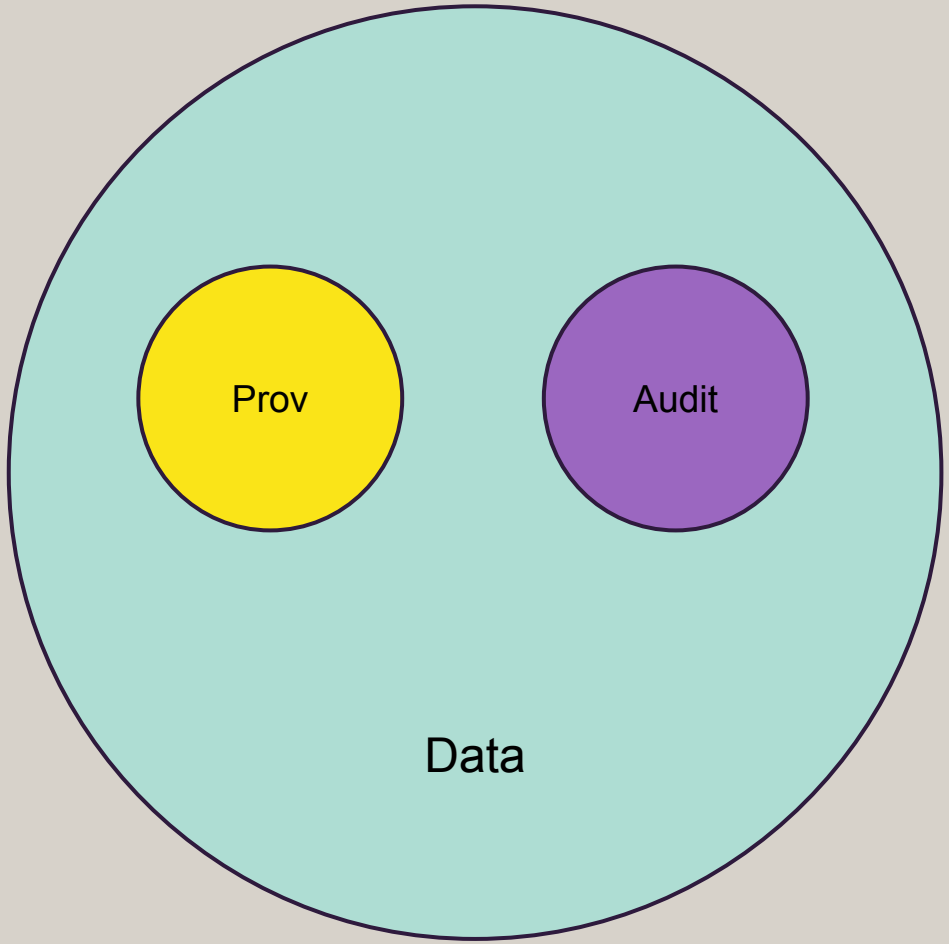
Comply Advantage

Provenance

- Is information about entities, activities and people involved in producing a piece o data
- Is used for assessments about data quality, reliability or trustworthiness
- Exists to provide evidence that a system is working properly
- Focused mainly on relations between Entities, Agents and Activities

Comply Advantage

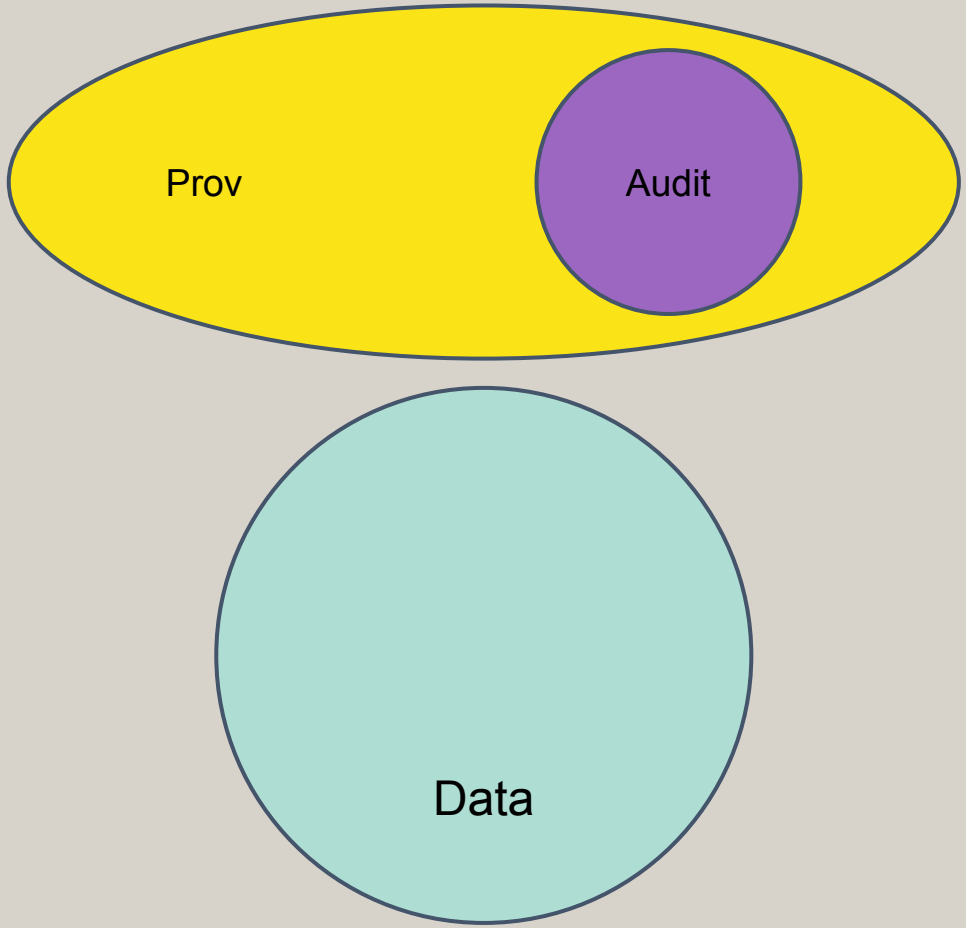
Reality



- Prov and audit data partially or totally embedded into data

Comply Advantage

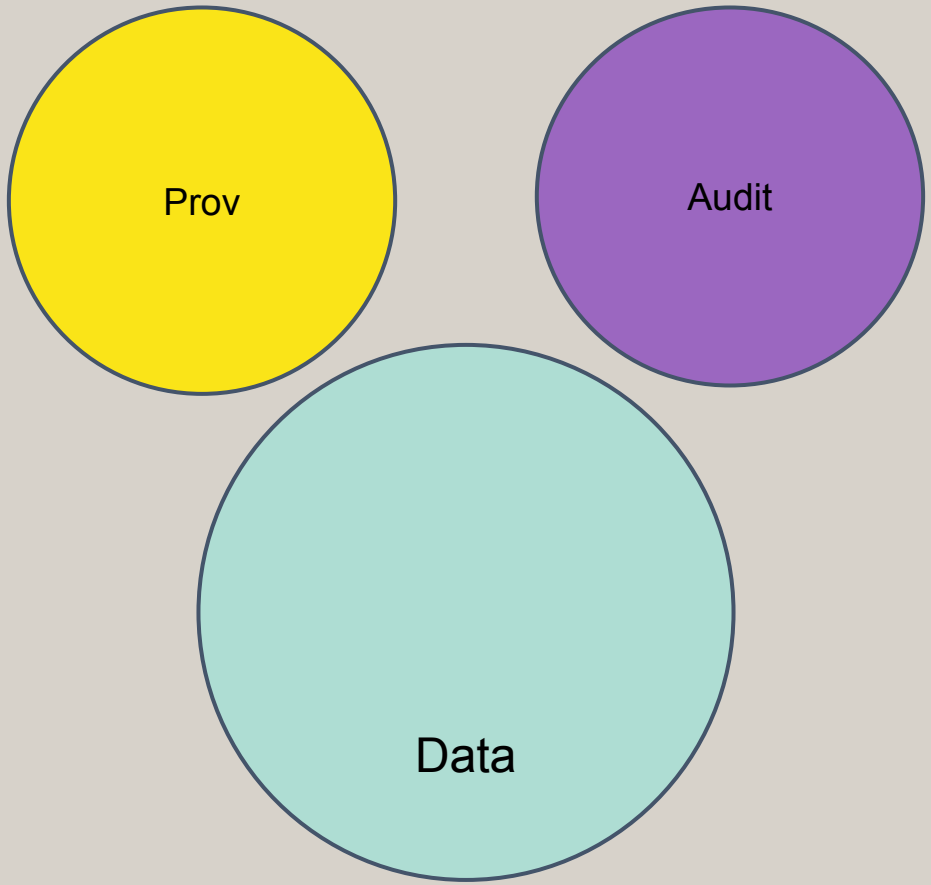
W3C Theory



- Prov and audit data decoupled and managed by the same solution

Comply Advantage

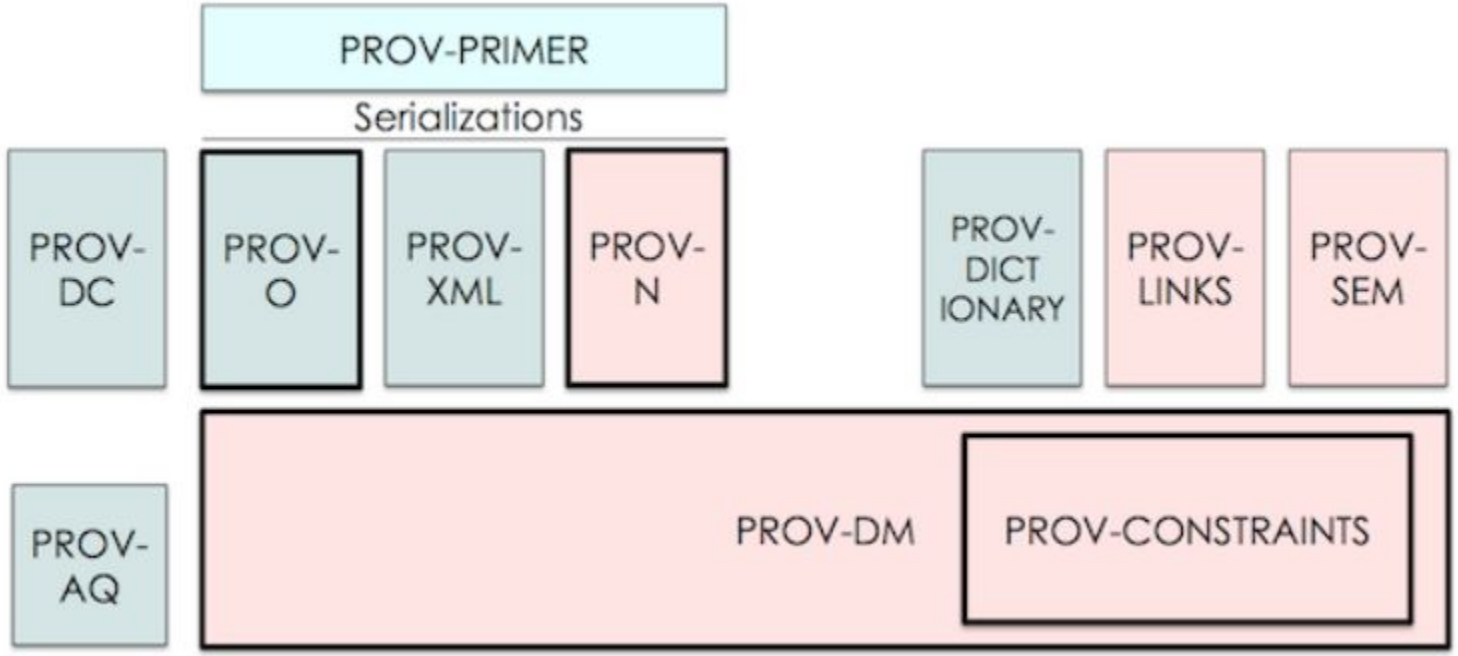
CA



- Prov and audit data decoupled and managed by different solutions

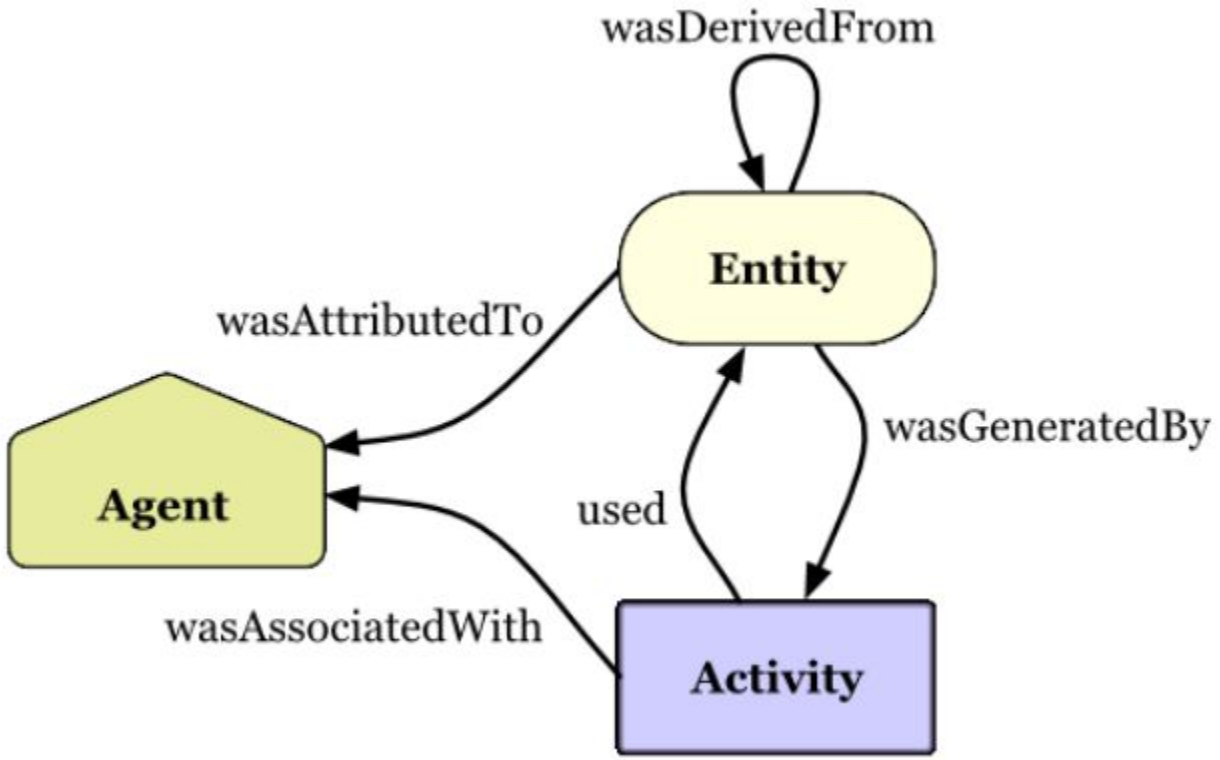
Comply Advantage

W3C PROV



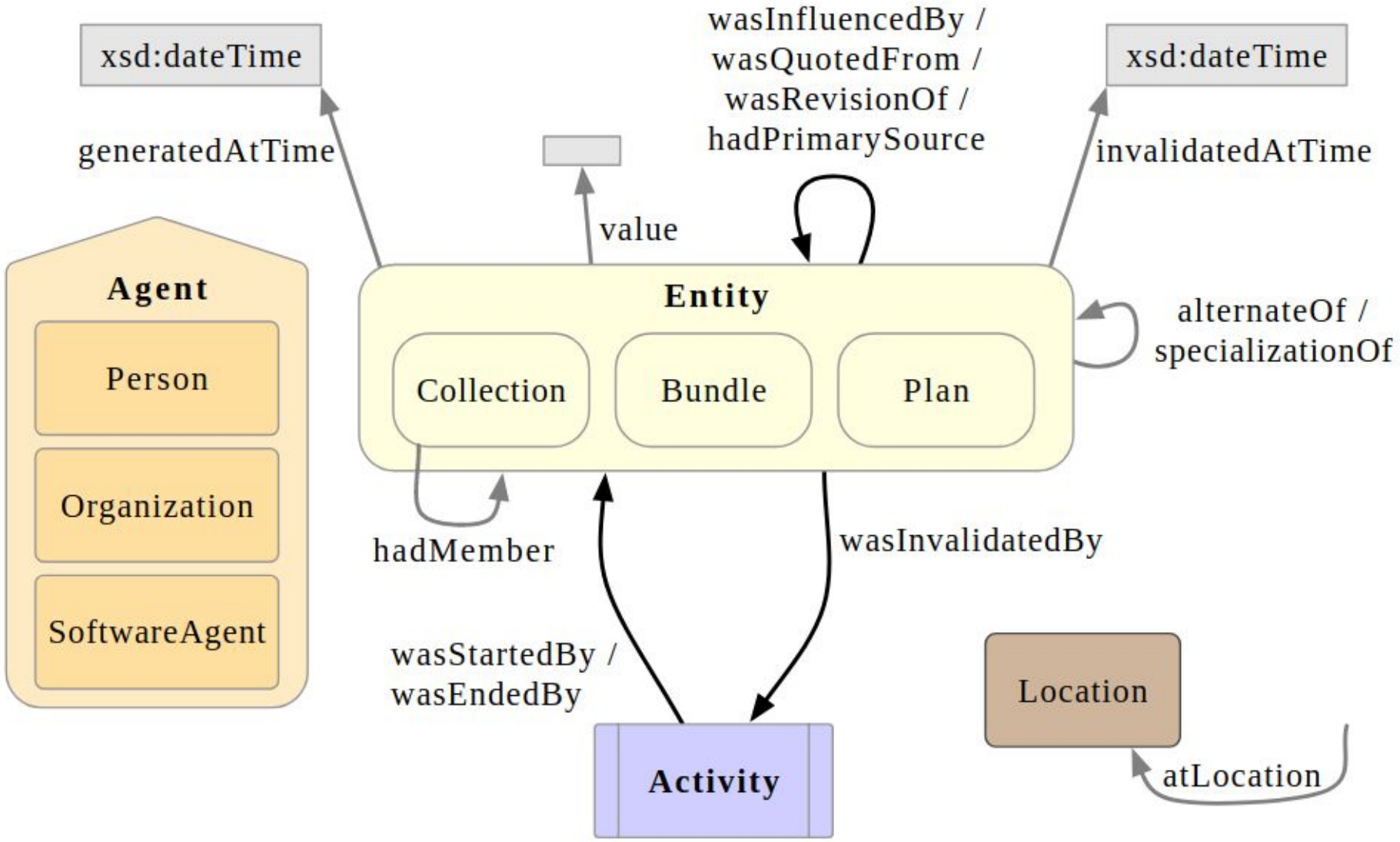
Comply Advantage

W3C PROV



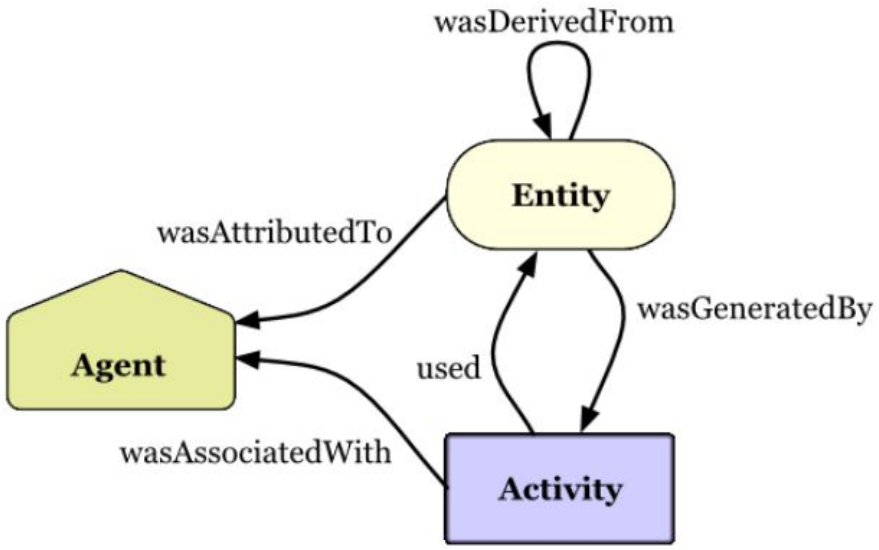
Comply Advantage

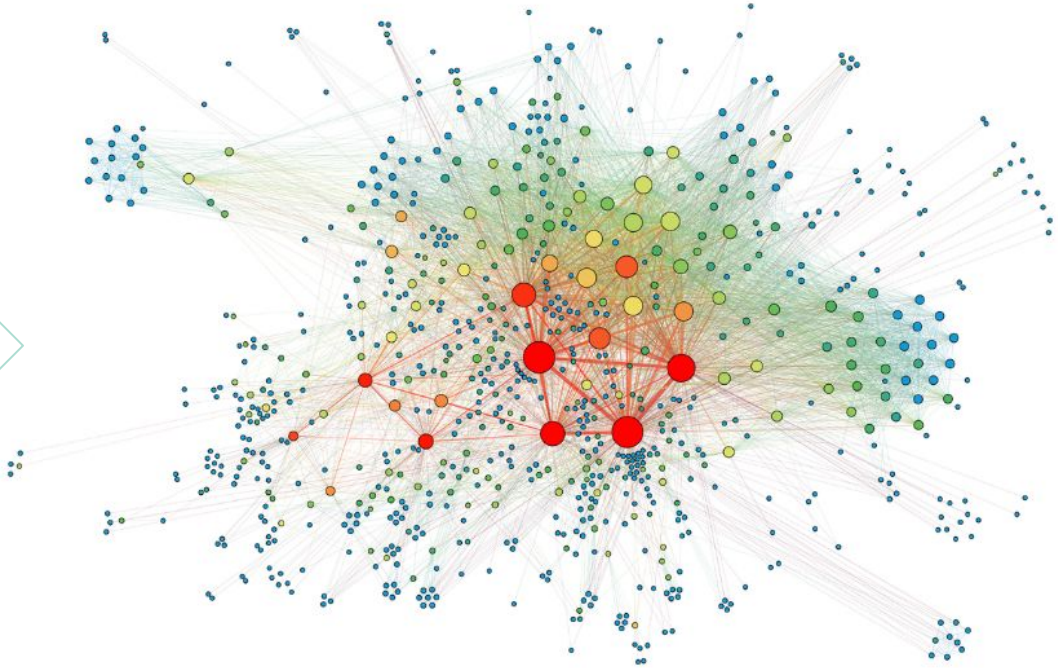
W3C PROV



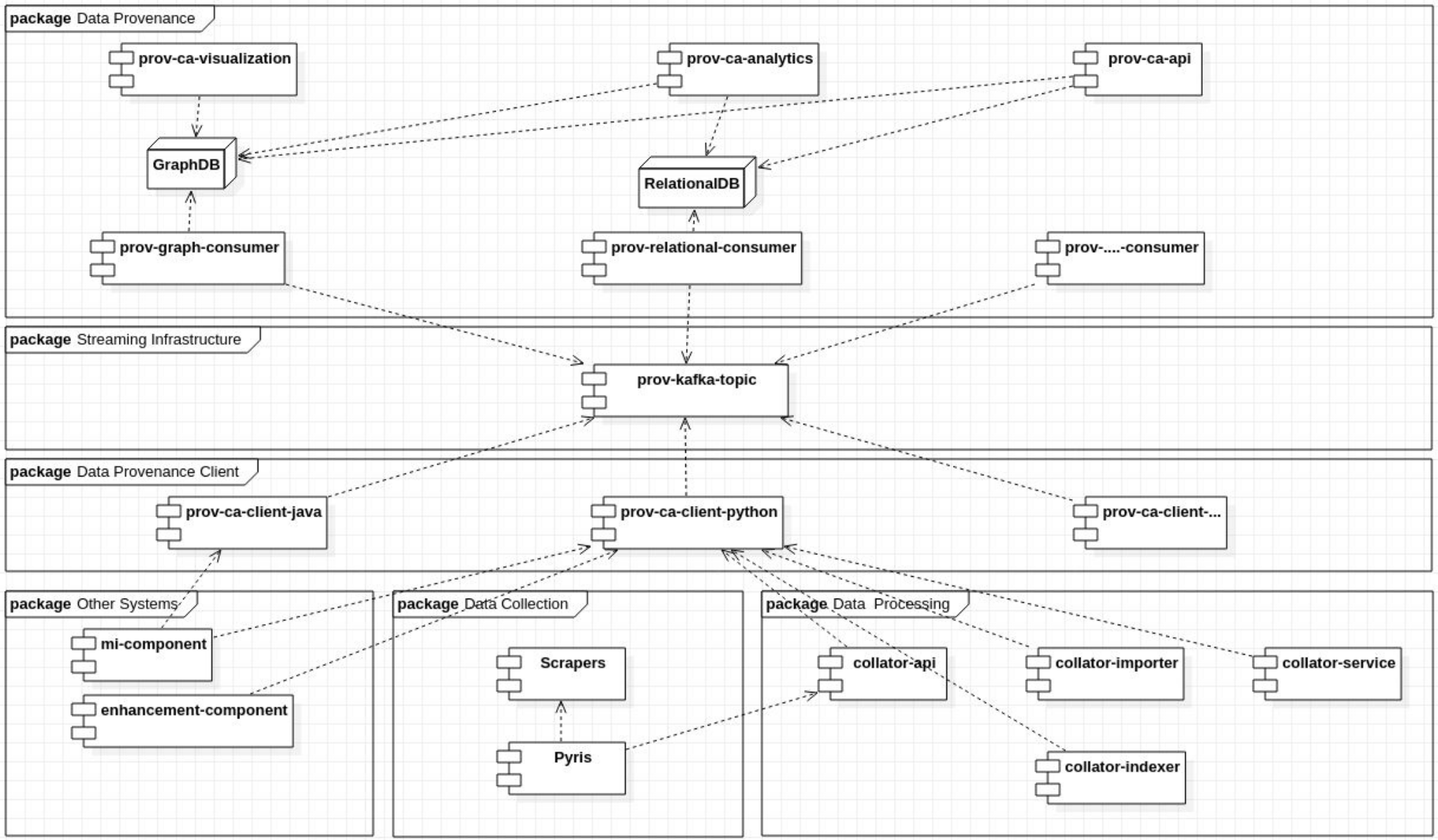
Comply Advantage

Why?





Comply Advantage



Comply Advantage

Experimental Stack

- Python
- Python prov library
- Python prov-db-connector
- Kafka
- JanusGraph (ThinkerPop, Gremlin - gremlinpyton, Cassandra)
- PostgreSQL
- AWS
- Docker, Kubernetes

Comply Advantage

Challenges & Learnings

- Robust specification - W3C PROV
- Even if the specification can be considered simple the complexity is given by identifying and modeling properly Entities, Agents, Actions and the relationships between them
- Can be challenging to collect and link provenance informations generated by different systems
- Scalability
- Provenance data retention
- Graph databases seems to be more suitable for provenance informations
- Didn’t find mature Gremlin support for prov documents (better support for Neo4j)
- prov-db-connector can be extended to support Gremlin
- decoupling audit and provenance data from business domain data can decrease the complexity of the systems
- decoupling audit and provenance data from business domain data can increase the complexity of the systems
- avoid using provenance data for other reasons, es. provide client services that are using PROV data
- OpenSource python provenance related libraries needs more support

Comply Advantage

Q&A

Comply Advantage