Instrument Integration to Knowledge Graph Development journey - *To enable seamless flow of data in Laboratory Space with Allotrope*

Fall 2024 Allotrope Connect


Amanda Murar


Sai Guttikonda


James Yurcho

Amanda Murar, Sai Guttikonda , James Yurcho - Merck & Co., Inc., Rahway, NJ, USA

(Nov/19/2024)

CURRENT

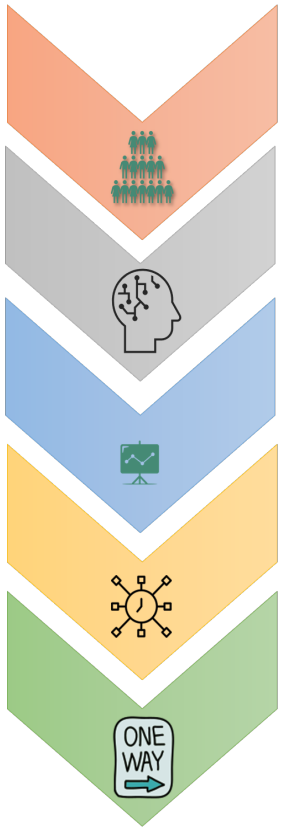
Manual and **People Dependent** Processes**,** Supporting Teams have **Varying** levels of **Technology Capabilities**

Multiple Bespoke Systems and **Highly Complex Processes**

Instrument Result **Data Flows** to LIMS through **Point-to-Point** solutions for digital record ingestion

Data is **Manually** Refined, **Difficult** to Access and Different by Site / Function

Every Site / Function Works Differently



FUTURE

Data-driven, Digitized Processes Supporting an Upskilled Workforce Across the Network

Simplified Systems and **Processes** are Robust, Intuitive, Integrated and Compliant

Increase **Data Accessibility/Availability** Through Holistic Design Patterns, for Multiple Customer Needs/Consumption

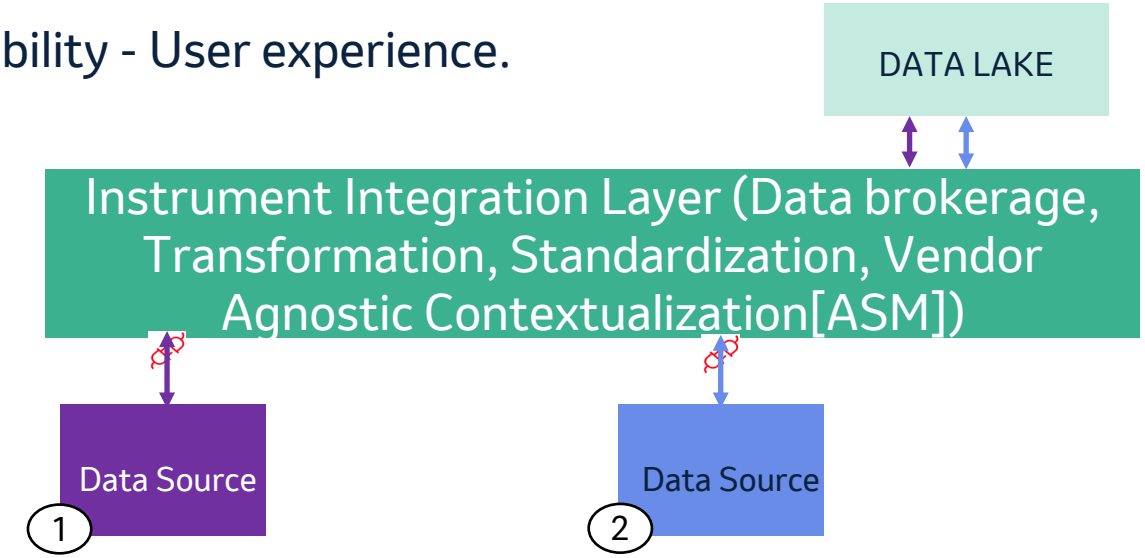
Data is **Accessible, Standardized** and **Available** in **Real Time**

Everyone follows **One Way** – Optionality is Eliminated

Instrument Integration

Challenge Statement: System Integration is designed around the user with no manual transcriptions between systems. User spends MORE time executing the process and LESS time documenting the process.

- Move away from P2P Integrations from source systems to LIMS and to Platform based approach to allow for data availability with context for down stream processing and analytics.
- Integrate only global standards.
- Develop a unified instrument Integration layer for data acquisitions, Transformation and Standardization.
- Right Balance of Capability - User experience.



Allotrope Simple Models: ASM patterns

Manifest Node

```
(:Manifest {
  url: "http://purl.allotrope.org/manifests/balance/REC/2024/09/balance.manifest"
})
```

Device Node

```
(:Device {
  type: "balance",
  identifier: "serial-number",
  serialNumber: "serial-number",
  calibrationTime: "2018-09-05T16:16:03.009Z",
  calibrationCertificate: "file://C:\\data\\calibration.cert"
})
```

Measurement Node

```
(:Measurement {
  id: "413befdd-c7e2-4edd-9e9b-06cf1cb0283f",
  time: "2018-09-05T16:16:03.009Z",
  analyst: "Amgentoaks1"
})
```

Sample Node

```
(:Sample {
  batchId: "batch-number",
  sampleId: "unknown-10"
})
```

ASM patterns define the standard ways to interpret ASM data using the semantics of Allotrope Foundation Ontologies (AFO).

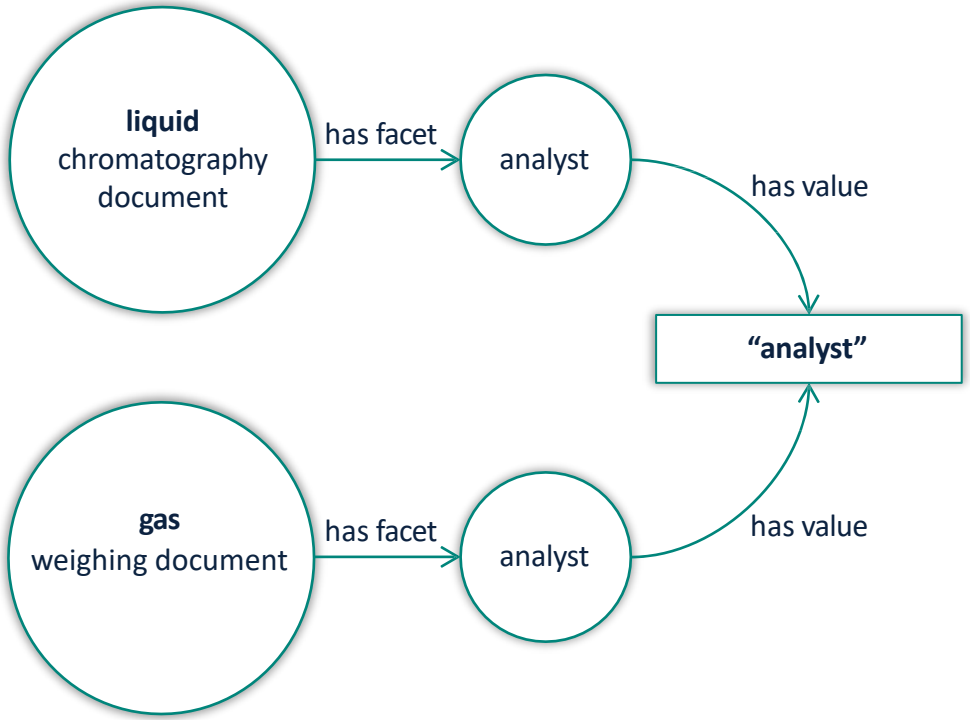
They guide an ASM JSON schema-aware transformation of data from ASM files into Labelled Property Graph (LPG) triples, facilitating the representation of structured data in graph databases.
This transformation enables the data to be stored as nodes and relationships with properties, which can be efficiently queried and analyzed using graph database technologies.

By mapping ASM JSON schema elements to nodes and relationships in LPG, the data becomes more interconnected and accessible for advanced data analytics, leveraging the strengths of graph-based data modeling.

Weight Nodes

```
(:Weight {
  type: "sample",
  value: 1.2354,
  unit: "g"
})
```

Queries across laboratory instrument data #1



Knowledge graphs are stored in **graph databases** which allow **efficient joins** across multiple sources of data, such as laboratory instrument runs.

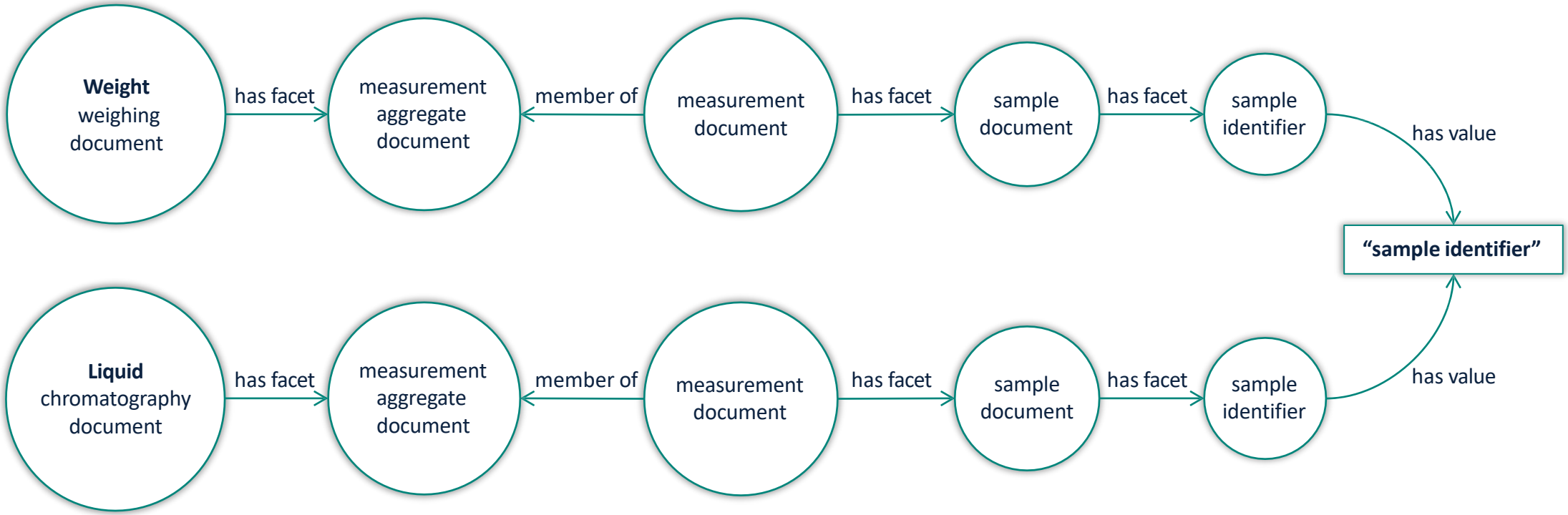
Cypher Query on a Knowledge Graph

```
MATCH (a:Analyst {name: "Kumar"})
RETURN a
```

Traditional SQL Query

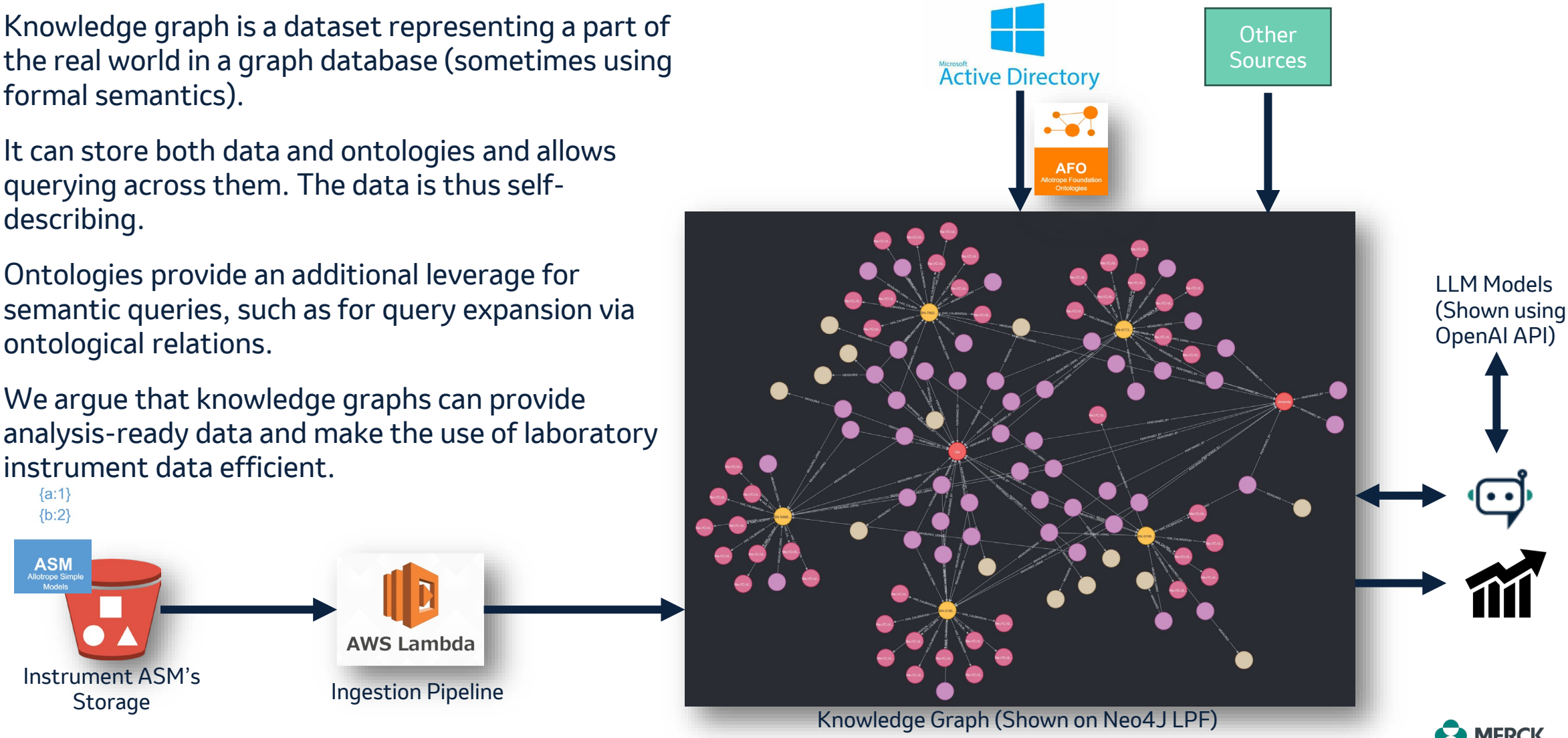
```
SELECT *
FROM liquid chromatography document
WHERE Analyst = 'Kumar'
UNION
SELECT *
FROM gas weighing document
WHERE Analyst = 'Kumar';
```

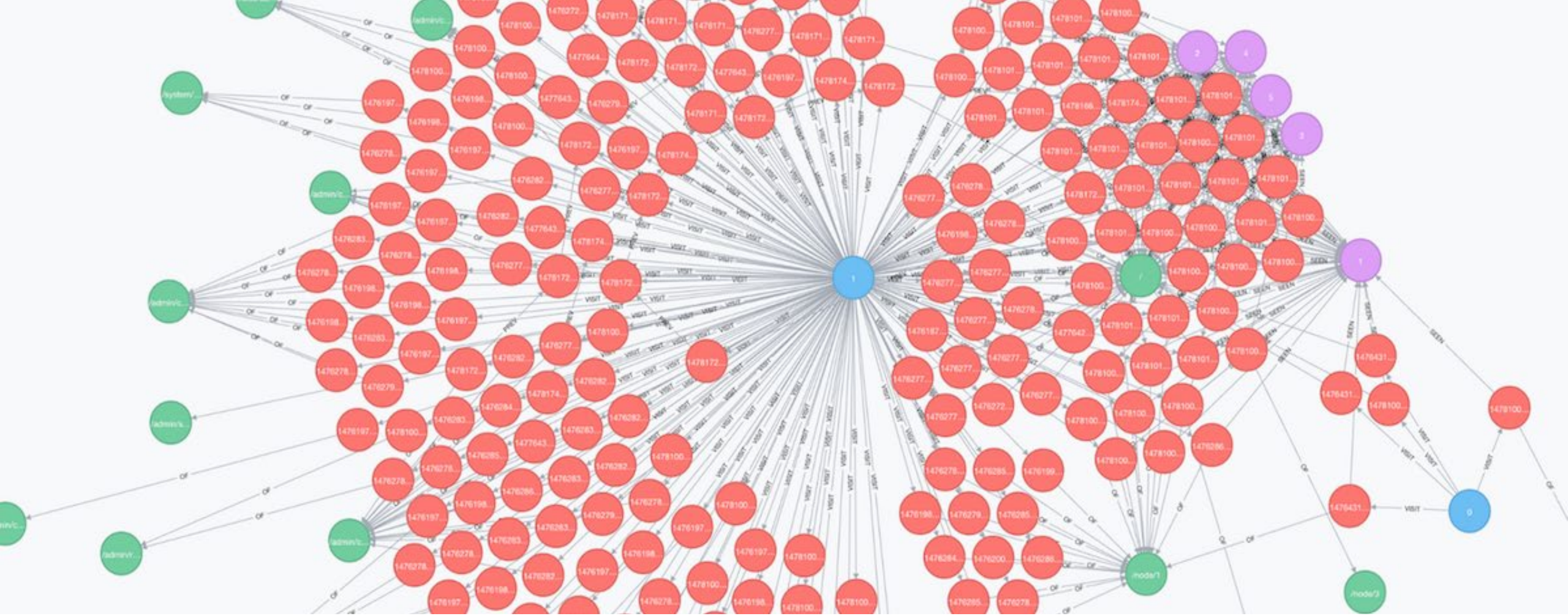
Queries across laboratory instrument data #2



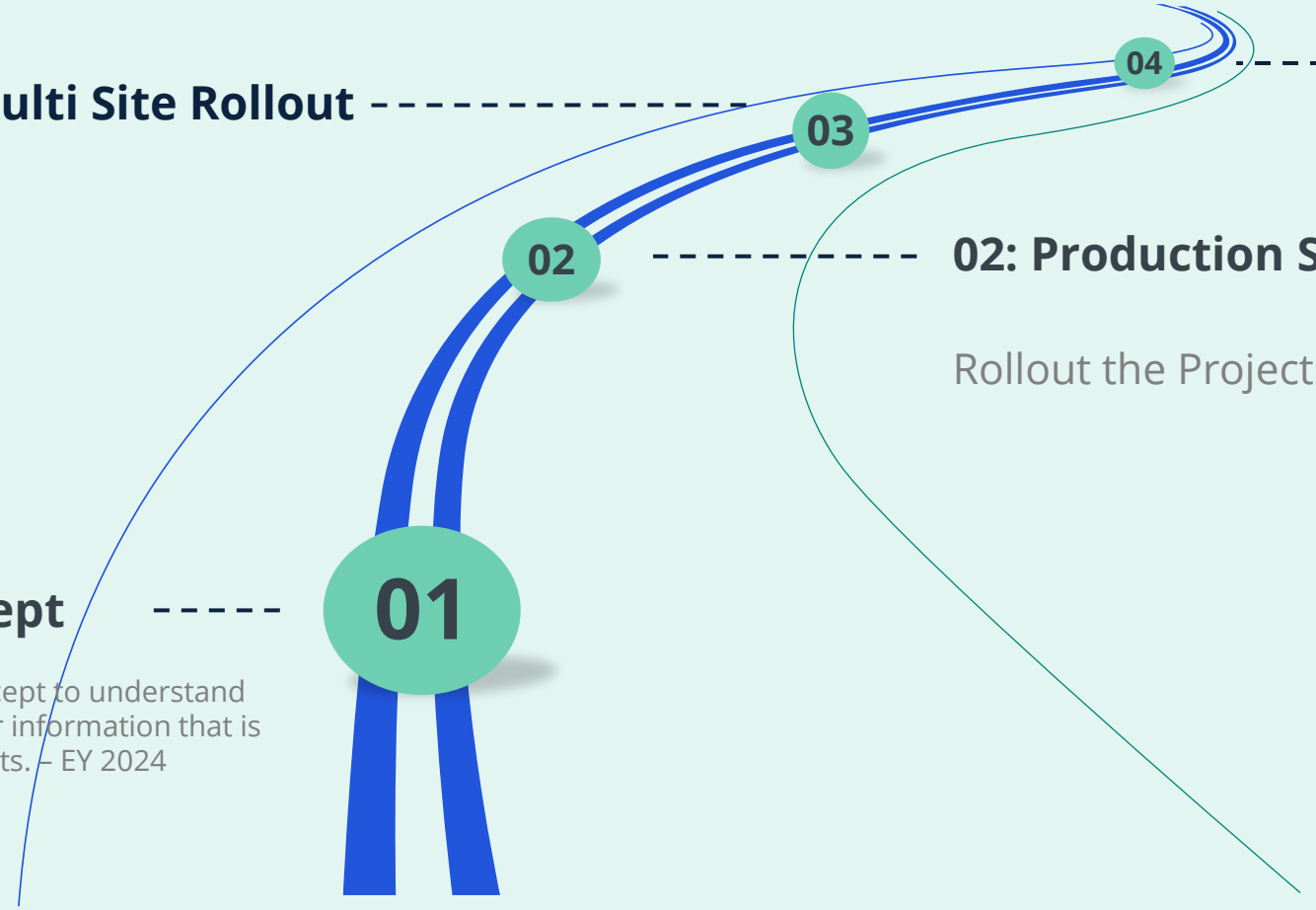
Knowledge Graph

- Knowledge graph is a dataset representing a part of the real world in a graph database (sometimes using formal semantics).
- It can store both data and ontologies and allows querying across them. The data is thus self-describing.
- Ontologies provide an additional leverage for semantic queries, such as for query expansion via ontological relations.
- We argue that knowledge graphs can provide analysis-ready data and make the use of laboratory instrument data efficient.





Live Demonstration



04: Complete Knowledge Graph and Continuous Improvement

03: Multi Site Rollout

02: Production Site

Rollout the Project to Pilot Site

01: Proof Of Concept

Team is running Proof of Concept to understand concept limitations and gather information that is required for production rollouts. – EY 2024 Completion.

- Finalize which is right Framework (RDF vs LPG) for Graph Database.
-Finalize the final place for storing the ASM Data cubes.
- Manufacturers Supporting Native ASM
- Develop Allotrope Standards for all the Instrument Classes that are present in the Road Map.