Knowledge Graphs – A Strategic NoSQL Approach for Data Management and Governance

IRENE POLIKOFF, CEO, TOPQUADRANT
RALPH HODGSON, CTO, TOPQUADRANT

March 19, 2019
The history of knowledge graphs in 90 seconds
What are today’s key challenges in data governance
What is a knowledge graph, how does it work
How do knowledge graphs support data governance
Demonstrating the use of knowledge graphs for data governance
WHO WE ARE

FOUNDATION
- TopQuadrant was founded in 2001
- Strong commitment to standards-based approaches to data semantics

MISSION
- Empower people and drive results — by making enterprise information meaningful

FOCUS
- Provide comprehensive data governance solutions

CUSTOMERS
- Over 120 active customer organizations
KNOWLEDGE GRAPHS
2001
Tim Berners-Lee

KNOWLEDGE GRAPHS
KNOWLEDGE GRAPHS

2001
Tim Berners-Lee

2012
Google

Ada Lovelace
Mathematician

Born: December 10, 1815, London, United Kingdom
Died: November 27, 1852, Maida Vale, United Kingdom
Spouse: William King Neil, 1st Earl of Lovelace (pr. 1816–1852)
Children: Anne Blair, 15th Baroness Volunteers, MORE
Parents: Lady Byron, Lady Byron
Known For: Mathematics, Computing

People also search for
View 15+ more

London
Capital of England

London, the capital of England and the United Kingdom, is a world-class city with a history stretching back to Roman times. It is centrally located between the major cities of the UK, with its own distinctive culture and history. The city's famous landmarks include the Tower Bridge, the London Eye, and the British Museum. The city is home to two of the most famous universities in the world, Oxford and Cambridge. London is also a major financial center, with the City of London being one of the leading financial centers in the world.
2001
Tim Berners-Lee

2012
Google

2019
Knowledge Graphs in the news

> Google on AI and knowledge graphs in ZDNET

> Data governance 2.0 on Dataversity

> Conferences & workshops

> Technology trends 2019

> Forbes article by Kurt Cagle
> Data represent value
> Usage of data is subject to regulation
> Bookkeeping of data as assets is necessary

“... organizing people, processes, and information technology required to create a consistent and proper handling of an organization's data across the business enterprise.” … Wikipedia
HOW IS IT DONE?

- Ad hoc
- Documents, spreadsheets, proprietary systems
- Examples:
  - GDPR, CCPA
  - Data quality
  - Data lineage
The challenges of data governance

> Galaxies of data
> Diversity of perspectives
  > Business
  > Technical
  > Regulatory
> Diversity of representation

Solution

> Create a knowledge graph representing data sources
> Link to other relevant enterprise information e.g., systems, policies, infrastructure, activities
> Enrich, discover, connect
> Use to guide business decisions
WHAT IS A KNOWLEDGE GRAPH?

- A Knowledge Graph represents a knowledge domain
- It represents knowledge as a graph
  - A network of nodes and links
  - Not tables of rows and columns
- It represents facts (data) and models (metadata) in the same way
  - Rich rules and inferencing
- It is based on open standards, from top to bottom
  - Readily connects to knowledge in private and public clouds
mi6:JamesBond mi6:worksFor mi6:MI6 .
mi6:JamesBond rdf:type schema:Person .
mi6:JamesBond mi6:worksFor mi6:MI6 .
mi6:JamesBond rdf:type schema:Person .
mi6:JamesBond \texttt{worksFor} mi6:MI6 .

mi6:JamesBond \texttt{rdf:type} schema:Person .

mi6:MI6 \texttt{rdf:type} schema:Organisation .
mi6:JamesBond rdf:type schema:Person

mi6:worksFor

mi6:Mi6 rdf:type schema:Organisation

DATA (INSTANCES) GRAPH
KG RULE:
{?x dbpedia:birthPlace ?place.
  ?place dbpedia:locatedInArea+ ?area.
  "?x example:bornIn example:UK"
}

EXECUTE INFERENCING:
If your birth place is in Scotland, you are born in the UK

QUERY:
Select all persons working for MI6 that are born in the UK
ENTERPRISE DATA GOVERNANCE

ENTERPRISE KNOWLEDGE GRAPH

- Real time access
- Model ingestion
- Feature extraction
- Reusable models
POLICY ATTRIBUTION

SITUATION CAPTURE

EVENT

Right Instances

Obligation Instances

Action Instances

RESPONSE DELIVERY

SYSTEM X

SYSTEM Y

GDPR

ENTERPRISE KNOWLEDGE GRAPH

RECO

Regulatory Compliance Ontology
Providing common search for all types of stakeholders
Connecting business terms to data elements
Supporting regulatory compliance by tracing data lineage
Representing regulations as knowledge graphs – generally (RECO) and specifically (GDPR)
Inferring rules from data in knowledge graphs
Connecting public and private knowledge graphs
Creating enriched knowledge resources
APIs for knowledge graphs – benefits of GraphQL
Welcome to TopBraid EDG

TopBraid Enterprise Data Governance™ (TopBraid EDG™) is an agile data governance solution for today's dynamic enterprises. With EDG, you can choose one or combine two or more data governance packages to support a comprehensive but staged approach to data governance.

Search the EDG

My Tasks
You have no tasks

Recently Viewed Collections
- FRY-9C Reporting Lineage
- GDPR Regulation
- RECO
- Custom Glossary Model
- Governance Model
Auto-Derived Facets – Based on the Search Results

Business Customer
(Service Strategy) A recipient of a product or a Service from the Business. For example:
if the Business is a car manufacturer then the Business Customer is someone who buys a car.

Client
A generic term that means a Customer, the Business or a Business Customer. For example Client Manager
may be used as a synonym for Account Manager.

The term client is also used in computing to mean:
- A computer that is used directly by a User, for example a PC, Handheld Computer, or Workstation.
- The part of a Client-Server Application that the User directly interfaces with. For example an email
  Client.
NAVIGATING ACROSS CONNECTED GLOSSARIES TO DATA ELEMENTS

Investment services client from the SEC Glossary

General term “client” from the ITIL Glossary

Same term name, but different identity. Each term is uniquely identified by its own URI.
### CUSTOMER_ID (TOPBANKCORP.LOAN_SECURITY) (Database Column)

<table>
<thead>
<tr>
<th>Identifiers Metadata</th>
<th>Data Element Properties</th>
<th>Database Column Metadata</th>
<th>Data Protection Properties</th>
<th>Traceability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>type:</strong></td>
<td>Database Column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>label:</strong></td>
<td>CUSTOMER_ID (TOPBANKCORP.LOAN_SECURITY)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>asset name:</strong></td>
<td>CUSTOMER_ID</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**DRILLING INTO THE DATA ELEMENT INFORMATION**

### CUSTOMER_ID (TOPBANKCORP.LOAN_SECURITY) (Database Column)

<table>
<thead>
<tr>
<th>Identifiers Metadata</th>
<th>Data Element Properties</th>
<th>Database Column Metadata</th>
<th>Data Protection Properties</th>
<th>Traceability</th>
<th>Other Properties</th>
</tr>
</thead>
</table>

- **maps to term:** Client

Mapped to the “client” term from the SEC Glossary
## DRILLING INTO THE DATA ELEMENT INFORMATION

### CUSTOMER_ID (TOPBANKCORP.LOAN_SECURITY) (Database Column)

<table>
<thead>
<tr>
<th>Identifiers Metadata</th>
<th>Data Element Properties</th>
<th>Database Column Metadata</th>
<th>Data Protection Properties</th>
<th>Traceability</th>
<th>Other Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>column of:</strong></td>
<td>LOAN_SECURITY (TOPBANKCORP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>is primary key:</strong></td>
<td>true</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>physical datatype:</strong></td>
<td>NUMBER (ORACLE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifiers Metadata</td>
<td>Data Element Properties</td>
<td>Database Column Metadata</td>
<td>Data Protection Properties</td>
<td>Traceability</td>
<td>Other Properties</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>--------------</td>
<td>------------------</td>
</tr>
<tr>
<td>personally Identifiable Information (PII):</td>
<td>true</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>personal data category:</td>
<td>General Personal Data Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>personal data use consent:</td>
<td>Consent Implied By Contract</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>obfuscated by:</td>
<td>Encryption Method Triple DES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
USE CASE SPECIFIC GRAPH EXPLORATION – TO UNDERSTAND LINEAGE OF DATA IN A REGULATORY REQUIRED REPORT
MORE THAN ONE TYPE OF LINEAGE CAN BE TRACED AND PRESENTED:

**LINK TYPES IN THE LINEAGE DIAGRAMS**

<table>
<thead>
<tr>
<th>Icon</th>
<th>Label</th>
<th>Depicts</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Activity Icon" /></td>
<td>Activity Link Icon</td>
<td>Process-activity</td>
<td>Associates a process with a process activity.</td>
</tr>
<tr>
<td><img src="image" alt="Business Area Icon" /></td>
<td>BusinessArea-activity</td>
<td>Associates a business activity with a business area.</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Data Input Icon" /></td>
<td>SoftwareExecutable-input</td>
<td>Software executables may have inputs that are data elements.</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Information Derived From Link" /></td>
<td>InformationAsset-wasDerivedFrom</td>
<td>Documents a dependency between information assets using the provenance ontology 'prov:wasDerivedFrom' property.</td>
<td></td>
</tr>
</tbody>
</table>

**LineageGram Inferred Link Icons**

Inferred links between resources are created using inference logic over the EDG Knowledge Graphs.

<table>
<thead>
<tr>
<th>Icon</th>
<th>Label</th>
<th>Depicts</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Information Dependency Icon" /></td>
<td>Information Dependency Link</td>
<td>Inferred dependency based on how a software executable requires an information asset that another software executable produces.</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Software Program based dependency Link" /></td>
<td>Software Program based dependency Link</td>
<td>A 'software program dependency' link is an inferred dependency based on how a software program's inputs and outputs are used.</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Transform-based Icon" /></td>
<td>Transform-based Dependency Link</td>
<td>A 'transform dependency' link is an inferred dependency based on how a software function's inputs and outputs are used.</td>
<td></td>
</tr>
</tbody>
</table>
DRILLING IN TO UNDERSTAND DETAILS BEHIND THE INFERRED CONNECTION: FROM A HIGH LEVEL TO A DETAILED VIEW
WHAT IS MEANT BY “CONSENT IMPLIED BY CONTRACT”
KNOWLEDGE GRAPH OF RECO – MODEL FOR DESCRIBING REGULATIONS
FROM DATA SOURCE TO THE OBLIGATION AND ITS MEANING

ENTERPRISE KNOWLEDGE GRAPH

SYSTEM X

SYSTEM Y

GDPR

RECO

Regulatory Compliance Ontology
KNOCKLEDGE GRAPHS IN ACTION

- Common Search for all types of stakeholders
- Connecting business terms to data elements
- How data lineage supports regulatory compliance
- Representing regulations as knowledge graphs
  - generally (RECO) and specifically (GDPR)

→ Inferring rules from data in knowledge graphs
- Connecting public and private knowledge graphs
- Creating enriched knowledge resources
- APIs for knowledge graphs – benefits of GraphQL
FROM: RULES IMPLICIT IN DATA
TO: RULES EXPLICIT IN A KNOWLEDGE GRAPH
MAKING RULES IMPLICITLY PRESENT IN DATA, BE EXPLICITLY REPRESENTED IN A KNOWLEDGE GRAPH

For example, based on the data, what rules can we learn about who can get a loan?

Should have high credit rating, income-to-loan ratio above a certain number, etc.

LEARNED FACTS, EXPRESSED AS RULES, ARE:

• REVIEWED FOR THE INCLUSION IN THE KNOWLEDGE GRAPH
• ONCE INCLUDED, USED IN FURTHER PROCESSING
Common Search for all types of stakeholders
Connecting business terms to data elements
How data lineage supports regulatory compliance
Representing regulations as knowledge graphs
  – generally (RECO) and specifically (GDPR)
Inferring rules from data in knowledge graphs

→ Connecting public and private knowledge graphs

Creating enriched knowledge resources
APIs for knowledge graphs – benefits of GraphQL
STANDARDS BASED KNOWLEDGE GRAPHS CONNECT PUBLIC AND PRIVATE KNOWLEDGE
STANDARDS BASED KNOWLEDGE GRAPHS CONNECT PUBLIC AND PRIVATE KNOWLEDGE
CREATING ENRICHED KNOWLEDGE RESOURCES

New Zealand (Country)

Types: Country

Labels and Description

preferred labels: Neuseeland (de), New Zealand (en), Nouvelle-Zélande (fr), Nowa Zelandia (pl), Nueva Zelandia (es), Новая Зеландия (ru)

type: Country

Geopolitical Characteristics

calling code: 64

capital: Wellington

Location

latitude: -41.28333333333333
longitude: 174.45

Standard Relationships

broader concept: Pacific Ocean
narrower concept: Tokelau, Wellington

Population (inferred): 4855500

Location

latitude: -41.28333333333333
longitude: 174.45

Matching Relationships

wikidata country: New Zealand

Other Properties

ISO ALPHA-2 country code: NZ
ISO ALPHA-3 country code: NZ

Flag (inferred): 🇳🇿

© Copyright 2019 TopQuadrant Inc.
KNOWLEDGE GRAPHS IN ACTION

- Common Search for all types of stakeholders
- Connecting business terms to data elements
- How data lineage supports regulatory compliance
- Representing regulations as knowledge graphs
  - generally (RECO) and specifically (GDPR)
- Inferring rules from data in knowledge graphs
- Connecting public and private knowledge graphs
- Creating enriched knowledge resources

→ APIs for knowledge graphs – benefits of GraphQL
WHAT IS GRAPHQL

- GraphQL is an application layer query language.
  - Meaning that GraphQL is designed to interpret a string from a server or client and return that data in an understandable, stable, and predictable format.
- Developed by Facebook for internal use
- Open sourced in 2016
- Rapidly growing in popularity, replacing REST style APIs

“Describe your data, ask for what you want, get predictable results.”

https://graphql.org/
KNOWLEDGE GRAPH APIs

GRAPHQL ACCESS with INTROSPECTION
GOVERNANCE OF ENTERPRISE DATA ASSETS REQUIRES:

Cataloging
Connecting
Curating

→ Collaborating

Governance Workflows, Roles and Metrics as Knowledge Graphs
BENEFITS OF A KNOWLEDGE GRAPH BASED PLATFORM FOR SUPPORTING DATA GOVERNANCE 2.0

- Flexible and extensible based on standards
- Integrates reasoning and machine learning
- Enables people (UI) and software (APIs/web services) to view, follow and query
- Bridges data and metadata “silos” for a seamless data governance
- Delivers Knowledge-driven data governance
Thank You!

... Questions?

VISIT US AT BOOTH #44