Realising the full value of your data with an Enterprise Knowledge Graph

Data Management Summit, London

21 March 2019
About the speakers

Jacobus Geluk

CTO and Founder

Semantic Technology Architect and visionary, specialising in Enterprise Knowledge Graphs and Enterprise level Data Unification.

At BNY Mellon, led the team which delivered the first Enterprise Knowledge Graph platform at scale in production in the financial industry.

Jeremy Posner

Principal

Data and technology specialist with 25 years’ experience, mainly within capital markets.

Formerly an Executive Director at Morgan Stanley and leading consulting practices within FS, he focuses on data management, strategy and architecture at some of the world’s largest banks.
Introduction & Contents

● Our data challenges - now and future
● The value of semantic technologies
● Introducing the Enterprise Knowledge Graph
● How do we get there?
Data Challenges Now
Uncomfortable truths

1. Your data needs are getting more complex and diverse. Your data platform is weighing down your business, not enabling it to deal with new business models.

2. Your efforts to govern and catalogue enterprise data aren’t moving fast enough. Regulators are asking for more and more detail, and you are struggling.

3. There’s huge value in your data, but mainly untapped. Unless you release its potential, your profits face further erosion by challengers, fintechs and tech giants.
The consequence: “insoluble” use-cases

We’ve spent decades trying to solve enterprise-wide problems

- **Customer**: KYC, CRM, agreements, transactions, positions, ...
- **Product/Service**: products, channels, markets, services, ...
- **Organisation**: people, process, data, technology, ...
- **Control**: risk, compliance, legal, entitlements, fraud, ...
- **Finance**: cost, revenue, profit, ...

These use cases are complex, related and overlaid...

...so they resist being solved using current technology
What’s blocking us?

Familiar obstacles:

- A huge application portfolio with functional duplication
- Silos of data with few standards
- Increasing complexity;
  massive change management problems
- High RTB:CTB ratio, prioritising “regulation-first”

Plus a more fundamental obstacle, which we’ll discuss later
The result

Authoritative Source Systems

Cottage Industry of Data Management Systems

Army of Excel Ninjas

Decision Makers

Knowledge

Information

Data

Risk
($ unknown)

Business budget
($ unaccounted for)

Technology budget
($ accounted for)

50%

50%
Data Challenges Future
Our looming future

The **knowledge worker** is being replaced with the **robot**

**Robotic Process Automation (RPA)**

**Intelligent Agents**

Simple

More Complex
Where we are going

2010 2020 2030
2030: “Hello Siri, this is Alexa…”

- Knowledge Workers and Intelligent Agents communicate freely with each other
- RPAs execute and report back
- MySiri meets YourAlexa
But *how* will they communicate?
For knowledge workers, context is everything

Knowledge workers:

- use data across sources
- ask questions
- have different viewpoints
- connect with each other
- share data, explain context
- to make decisions
  … or ask more questions
Can AI & humans explain context to each other?

AI & Humans will also need to:

- **connect** with each other
- share data, explain context (HOW?)

There will never be **ONE** data model.

There will **ALWAYS** be different viewpoints.

Humans and AI must access the **SAME** data network.
Back to the pyramid...

- **Data**
  - Authoritative Source Systems
- **Information**
  - Cottage Industry of Data Management Systems
  - Army of Excel Ninjas
- **Knowledge**
  - Decision Makers

{Decisions}
- Viewpoints
- Context
- Questions
- Data
So to recap...
To succeed, we need to...

Untangle data **today**

Win the AI race **tomorrow**

FORTUNATELY ... THE SAME SOLUTION
So what *is* the solution?
Value: data as an asset

- We have all heard about the importance of being a **data-centric** organisation
- But **how** do we get most value from our data?
Value: of data connectedness

connections!

dwelling, ownership, demographics

background

communication history

agreements transactions, profitability, loyalty

workplace, organisation, history, position

gender

marriage, partner, family

sentiment

hobbies, interests

buying habits

customer name

Typical Customer 360
The world’s most data-savvy companies use **data connectedness** to derive deep insights:

- knowledge graph
- social graph
- connection graph
- Alexa and product graphs
Is “just a graph” enough?

Simply connecting isn’t enough.

- What does the **data** mean?
- What does the **connection** mean?
- How do we **define** those things using **standards**?
- How do we deal with different **viewpoints**?

“Just a graph” won’t **deal with this**:
Semantics: “Things, not Strings”
Semantics are mature and standardised

- Fortunately there are **well-defined standards**
- Standards that built the internet
- Standards that are **mature**
  (Web: 1989, Semantic Web: 2001)
- Standards that allow **machines and humans**
  to understand and communicate meaning
Writing in 1999:

A "Semantic Web" has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize.
Now, in 2019… the standards are mature, the technology is available.
So what is the Semantic Web?

<table>
<thead>
<tr>
<th><strong>Linked Data</strong></th>
<th>publishing structured data so that it can be interlinked and become more useful through semantic queries over the internet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontologies</strong></td>
<td>define the vocabularies/concepts and relationships used to describe and represent an area of concern and its metadata in a machine-readable manner</td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td>technologies and protocols that can programmatically interact with data from the Semantic Web</td>
</tr>
<tr>
<td><strong>Inference</strong></td>
<td>discovering new relationships based on a set of rules and data</td>
</tr>
</tbody>
</table>
And how do we make this real?
Introducing the Enterprise Knowledge Graph
An Enterprise Knowledge Graph...

- uses semantic technologies to connect data across the enterprise
- links both internal and external data
- promotes a true data re-use, so it doesn’t become another silo
- supports multiple viewpoints
- provides data context and meaning
- enables deep insight and decision-making by humans and AI
Why does this work?
What made the WWW work?

- Unifies browse & search
- Enables connected content
- Decentralised and inclusive
- Built on open standards
What makes an EKG work?

- Unifies browse & search
- Enables connected content
- Decentralised and inclusive
- Built on open standards
- Powered by semantics
- It’s a graph!
- Links data, doesn’t move it
- Standards defined by W3C
Modelled on the web

The World Wide Web

Browser

Web Search

Web server

Web server

Web server

The Enterprise Knowledge Graph

Client

EKG Search

KG service

KG service

KG service
“But I already have a graph database…”
Property Graph vs Semantic Graph

**Property Graph**
- No Data or Metadata Standards
- Many Query Standards *
- No Reasoning Standards
- No Ontologies
- Supports one model at a time ("Closed World Assumption")
- Semantic meaning separated from data

A candidate for point solutions, but “Yet Another Silo”

**Semantic Graph**
- Mature Data Standards (RDF)
- Mature Query Standards (SPARQL)
- Mature Reasoning Standards (OWL)
- Mature Open Ontologies (e.g. FIBO)
- Supports many simultaneous models ("Open World Assumption")
- Semantic meaning forms part of data

**Fit for a true enterprise platform**

* as of early 2019
What about multiple viewpoints?
The relational world is 2D

When you think of data as a table:

- Row = “Identity”
- Column = “Meaning”
- Cell = “Value”

and there’s only room for one of each
The real world is more complicated

Not only multiple versions, with different identities, meaning and values over time

But also many sources, with the same or similar meaning but with different values

Source A

Source B

Source C

Source D
...so is there *really* a “single version of the truth”?
Embracing multiple viewpoints

- An EKG can store **multiple versions of the truth**

- **Context** (a kind of metadata) records where each “truth” originated

- The choice between conflicting “truths” is made at query time

- The answer may be different depending on the context of the query
Datapoint 360

- Our definition of a datapoint:

  The business meaning of a concept ("my name", "your account"), with all "fact claims" from all sources, combined into one "datapoint object"

- Gathers all values, identities, versions and semantic definitions from any given source

- Links to all other aspects as shown here
A new way of thinking

Legacy Thinking

- Force everything into a single canonical model
- To find out what a property means, look in a data dictionary
- Match between data sources using heuristics
- Choose a “truth” at ingestion time, baking in undocumented assumptions

EKG Thinking

- Embrace multiple data models
- Store semantic meaning as an integral part of the data
- Match using unambiguous universal IDs
- Choose a “truth” at query time, once the context is known
Enterprise Knowledge Graph Summary

Features:

- Supports multiple models and viewpoints that convey context
- A distributed set of connected data across your whole enterprise
- Based on standards that built the internet
- Machine readable, ready for AI
- Supporting search, query, updates, navigation, provenance, inference

Outcomes:

- Untie the Gordian Knot of hidden decisions
- Enable AI to make business decisions
- Implement “insoluble” use-cases
- Uncover new insights
- Become more data-centric
- Have fewer silos
- Decrease time-to-market for new ideas
How do we get there?
Key tools in your journey towards EKG

A Maturity Model for EKG

- Based on CMMI
- Map your ambitions and plot your journey
- Understand your current position and where you need to focus
- Review tools and solutions objectively

A Use-Case Tree

- Planning tool for the EKG build-out
- Business priority driven
- With a data centric-approach
- Allows incremental build, hence value
- Top-down or bottom up
EKG maturity levels

1. MVP+
2. Platform 1.0
3. Enterprise Ready
4. Driving Better Decisions
5. Pervasive, Self-Improving

“Insoluble” use cases are here

You are (probably) here
EKG maturity levels

1. MVP+
   - The beginnings of a full-fledged EKG architecture.
   - Limited in scope, but serving multiple use cases. Built on solid design principles as a foundation for future evolution.

2. Platform 1.0
   - A first implementation, including Minimal Viable Product (MVP) and subsequent early iterations. Demonstrates the potential

3. Enterprise Ready
   - Serving multiple enterprise use-cases. Demonstrably mature and stable. Able to support mission-critical activities. The default choice for new data projects.

4. Driving better decisions
   - Widely adopted, and fully connected to virtually all enterprise data resources. Decision-makers rely on the EKG for trustworthy, holistic information. AI-driven decisions rolled out in some areas.

5. Pervasive, Self-Improving
   - Existing data silos eliminated where possible, as the EKG becomes authoritative. EKG is closely integrated into all core business operations. Fully connected AI is just another author.
The use case tree

From top down…

one complex target use case decomposes into many others

...or bottom up

start with foundational use case(s) to enable many more complex ones

<table>
<thead>
<tr>
<th>Level 5</th>
<th>Level 4</th>
<th>Level 3</th>
<th>Level 2</th>
<th>Level 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKG</td>
<td>CMM</td>
<td>Level</td>
<td>Level</td>
<td>Level</td>
</tr>
</tbody>
</table>

CUSTOMER 360

KYC

SINGLE VIEW OF CUSTOMER
Maturity and technology

- Reaching the highest maturity levels requires the correct technologies and architecture.
- Switching technologies or architectures is always expensive.
- So, the target maturity level must be considered from the start, to avoid being locked out of the most valuable use cases.

![Diagram of maturity levels and time/Enterprise Scope relationship]
Back to the pyramid...

- Knowledge
- Information
- Data

- Authoritative Source Systems
- Decision Makers

- EKG

- Decisions
- Context
- Viewpoints
- Questions
- Data
1. Formulate a **vision and strategy**, backed by C-Level champion.

2. Obtain **mandate to innovate** and adopt new approaches and **seed investment**.

3. Develop **roadmap and programme** driven by most **impactful use-cases**, broken down into clear regular milestones.

4. Build a **center-of-excellence** with the right capabilities.
Questions?