Industrial Knowledge Graph at Siemens
Powered by metaphactory and Amazon Neptune

CERN Openlab Technical Workshop, Geneva
Dr. Alexey Fishkin, Siemens AG, Corporate Technology
Outline

Motivation
• Challenges
• Why graphs?

What we do
• Our vision: Integrated intelligence
• A day of use cases

Technology
• Required functionalities
• Current infrastructure

Towards Graphs-as-a-Service
• Why go managed?
• First experiences with Neptune
## Challenges in our industry

### Isolated data silos
- By owner (Siemens divisions, customers, …)
- By subject (operating data, maintenance data, error information, customer data, …)
- By media type (time series, images, PDFs, …)

### Data inaccessibility
- Access paths are too complicated for domain experts leading to high costs for data access
- No integrated view of data
- No or limited search functionalities

### Inefficient workflows
- Long delay from information needs to data access
- Data provisioning demands big capacities of IT experts
- Heterogeneous storage lead to complex data control

### Low data quality
- Outdated
- Duplicated
- Incorrect or contradictory
Why (Knowledge) Graphs?

Benefits of using knowledge graphs for data representation

- The world is entities and relations!
- Intelligible domain model instead of complex (physical) data model
- Schema-on-read instead of complex schema migration for extensions
- Easy integration of multiple data sources (schemas) and types (structured, unstructured, …)
- Formal semantic representation enables inference and machine processing
Our answer – Industrial Knowledge Graphs for capturing Siemens Domain Knowledge

1. Isolated Data Silos with hand-crafted expert systems
2. Domain-specific Knowledge Graphs generated from DBs
3. Connected Knowledge Graph via automated structure and link discovery
4. Learning Memories extract expert knowledge from observations

Degree of automated knowledge digitalization

Knowledge

From isolated data silos to learning memories

Collected data  Digitalized Knowledge (via reasoning and learning)
Vision – Learning Memories for Integrated Intelligence

Knowledge Graph

Semantic Memory
(know)

Episodic Memory
(remember)

Working Memory
(integrate)

Decision Making
(act)

Perception
(understand)

AI Algorithm

Declarative

...LIONS ARE WILD AND DANGEROUS

...
One day in the life of an artificial assistant @ Siemens

Industrial Knowledge Graph

09:00 – Analyze
Turbine data hub

11:00 – Configure
Configure turbine

12:00 – Maintain
Master Data Mgmt.

13:00 – Mitigate
Financial Risk Analysis

15:00 – Connect
Experts & Communities

18:00 - Guide
Rules & Regulations

AI Algorithms
Challenge
- Required data is distributed across multiple databases
- Source systems have highly complex schemas
- Need to include unstructured information into analyses
- Reactivity and efficiency needs call for end-user access to data

Solution
- NLP to make information from documents accessible for analytics
- Physical and virtual data integration to provide unified view
- Access using a domain ontology and intelligent query construction support
- Connectors providing data to existing tools in legacy format

Value Generation
- Unified Data Hub: All information accessible from one system, independent of source and type
- Empower domain experts: Subject matter experts can use domain language to access data
- Enabler for analytics: Foundation for fleet-level analytics

Industrial Knowledge Graph @ Work – Flexible self-service data access for domain experts

What’s the MTTF distribution across turbines with coating loss last FY?

I will compile the relevant data and visualize it in your favorite BI tool.

Data-as-a-service making anyone a data scientist

Unrestricted © Siemens AG 2017
Page 8 November 2017

SIEMENS
Ingenuity for life

CT RDA BAM
Challenge

- Product configuration information is scattered across spreadsheets, inconsistent and redundant
- Missing transparency on technology interactions
- Information only on HOW to design, but not on WHY to do so

Solution

- Use an Industrial Knowledge Graph to store product configuration knowledge with rich semantics
- For new order, create constraint system on the fly using knowledge graph information & solve for feasible solutions
- Use Industrial Knowledge Graph technology to browse solutions

Value Generation

- Introduce knowledge management into turbine configuration
- Integrated design process across all components and technologies
- Semantics allow explain the WHY behind a design decision
- Speed-up due to automation

How to configure the new turbine as to meet customer requirements?
I will evaluate all constraints and provide a list of possibilities.
Industrial Knowledge Graph @ Work – Building a single source of truth for product relation data

Challenge
- Consolidate product relation data scattered across multiple systems
- Use intelligible rules to derive and quality-check product relations
- Provide high-quality data on intra-product relations (successor, etc) for customer-facing applications

Solution
- Link relevant product information in the Industrial Knowledge Graph
- Data integrity dashboard with expert-defined rules (SPARQL) to identify data quality issues
- Use cases access required information subsets using specific APIs

Value Generation
- Increase revenue by cross- and up-selling (driven by richer information)
- Facilitate knowledge management by product experts with increased transparency and data integration
- Guaranteed consistency of information provided across tools
- Reduced efforts for product data management

How can I make sure that users see consistent data across all apps?

I have integrated all product information and compiled a data integrity dashboard.

Challenge
- Consolidate product relation data scattered across multiple systems
- Use intelligible rules to derive and quality-check product relations
- Provide high-quality data on intra-product relations (successor, etc) for customer-facing applications

Solution
- Link relevant product information in the Industrial Knowledge Graph
- Data integrity dashboard with expert-defined rules (SPARQL) to identify data quality issues
- Use cases access required information subsets using specific APIs
Industrial Knowledge Graph @ Work – Understanding and mitigating risks in financing

Which investments we have are directly or indirectly affected by this storm?

I have browsed all investments and company relations and put into a report.

**Challenge**
- Siemens bank has a **wide range of investments** across industries
- Complex **networks of company relations** (own, partners, competitors)
- **Limited transparency on risks** due to external events, fraud, partner and competitor activities, etc

**Solution**
- **Combine internal and purchased information** on companies and projects in Knowledge Graph
- Highly **flexible query interface** to support arbitrary queries (structured and natural language search)
- Interfaces to **support analytics** on top of integrated data

**Value Generation**
- Highly agile analysis of risks caused e.g. by unforeseen events
- **Improved transparency** over partners and competitors
- Identification of **potentially fraudulent behavior patterns**
Industrial Knowledge Graph @ Work – Cross-hierarchy community building and expert search

Challenge
- Foster community building across organizational boundaries
- Finding experts within Siemens, utilizing personal networks

Solution
- Industrial Knowledge Graph integrates information on people, projects, and organizations
- Tapping into corporate data silos to provide an integrated view
- Utilize public sources for skill hierarchies to improve search
- Possibility to integrate relevant external sources (career platforms)

Value Generation
- Transparency over informal communities as well as formal organizational hierarchies
- Finding experts made easy
- Utilize FOAF-graph to facilitate support-seeking

Do we have experts near my office who can help me with ontology design?

Stephan G. should be able to help. You might ask Thomas H. to make contact.

Flexible collaboration across organization boundaries
Industrial Knowledge Graph @ Work –
Prescriptive advice from complex rule frameworks

Challenge
- Large organizations have huge bodies of rules & regulations
- Numerous facets of rule scope (country, site, organization, …) complicate finding applicable rules
- Rules tell WHAT to do, not HOW
- Communication is done via PDF-based circular documents

Solution
- Utilize NLP technology to extract subject, roles, scope, and activity information from circulars
- Industrial Knowledge graph to integrate with organizational knowledge, giving contextualized descriptive guidance (e.g. whom to call)

Value Generation
- Reduce time spent on understanding processes and following them
- Reduce non-conformance cost due to process violations
- Increase employee satisfaction by simplifying processes

Help employees follow regulations easily and quickly

I need this order done. Whom need I contact to get it approved?

Based on locally applicable rules, you first need to talk to John Doe from ECC.

15:00
18:00
13:00
12:00
11:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00

Knowledge as a Service – Consuming knowledge and analytics should be as easy as shopping

AI Algorithms
Knowledge Graph

Unrelated Data Silos

Service and Maintenance Data
Usage Data of Siemens Software
Design and Engineering Data
Supply Chain Data
Customer and Sales Data
Operational Sensor Data

Siemens API

Data Platform (e.g. Mindsphere)

Connectivity
Functionalities for an Industrial Knowledge Graph Platform

**Data integration** (reconcile, …)
- Query endpoint (SPARQL)
- Query endpoint (other)
- REST APIs
- Graph analytics
- Semantic search
- Planning & optimization

**Model editing & management**
- Graph DB (triple store)
- Graph DB (property graph)
- Policy enforcement (e.g. GDPR)

**Model bootstrap**
- Machine Learning (connector)
- Advanced Querying (semantic, similarity, …)
- Metadata manager (sourced data)
- Metadata manager (other)

**Reasoning**
- Federation
- SPARQL connector (OBDA)
- SQL connector (via API/DB/file/…)
- Siemens connectors (scripted, from files)
- Batch ETL (XLS, XML, …)
- Self-service ETL (from searchable docs)
- NLP / text processing (from searchable docs)
- OCR (for non-searchable docs)

**Query endpoint**
- (SPARQL)
- (other)

**REST APIs**
- (other)

**Graph analytics**
- (connector)

**Semantic search**
- (sourced data)

**Planning & optimization**
- (other)
Siemens AWS Use Case Architecture
metaphactory Knowledge Graph Platform

Knowledge Graph Management
- SPARQL endpoint UI
- Navigation, exploration, visualization
- Authoring, ontology and instance data management

Knowledge Graph Application Development
- Rapid prototyping of end-user oriented applications
- Web components for end-user oriented data interaction

Knowledge Graph Middleware
- “Queries as a Service”
- Interfaces to third party applications
- Integration with other AWS services
Reasons for Amazon Neptune

- Fully managed service
- Scalability & Performance
- High availability and durability
- Security & Encryption
- Standards compliance
- RDF / SPARQL and property graphs
First Use Case on Neptune: Siemens Product Knowledge Network

Product Relationship Management based on Master Data
- Integrated data from variety of sources
- Central hub for applications to access product data

POC with metaphactory and Neptune
- Graph-based data integration:
  - 1.2 M products
Graph Visualization and Exploration

Variety of graph structures:
- Product metadata and relationships
- Successor, predecessor network
- Taxonomic information
Editing instance data

Semantic forms for authoring product data

- End-user oriented interface
- Auto-suggestions against the knowledge graph
- Constraint validation
Data Quality checks

- Checks against integrated graph populated from many sources
- Rules and constraints defined as graph patterns
- Evaluated as SPARQL queries
- Visualized in interactive data quality dashboard
End-user Oriented Search

Semantic Search

- End-user interface supporting complex information needs
- Visual and interactive query construction
- Faceted result exploration
Queries as a Service: Dynamic REST APIs

Dynamic REST Services

- Declarative data access with SPARQL queries
- Automatically exposed as REST APIs
- Easy application development
- Fine-granular access control

<table>
<thead>
<tr>
<th>REST URL</th>
<th>Query</th>
<th>Enabled?</th>
<th>ACL permission</th>
<th>Response format</th>
<th>Repository</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pkn/poa/delete-instance</td>
<td>delete-instance</td>
<td>yes</td>
<td>spars.execute*</td>
<td>(default)</td>
<td>(default)</td>
<td>admin</td>
</tr>
<tr>
<td>/pkn/poa/new-asset-instance</td>
<td>new-asset-instance</td>
<td>yes</td>
<td>spars.execute*</td>
<td>(default)</td>
<td>(default)</td>
<td>admin</td>
</tr>
<tr>
<td>/pkn/poa/new-asset-type</td>
<td>new-asset-type</td>
<td>yes</td>
<td>spars.execute*</td>
<td>(default)</td>
<td>(default)</td>
<td>admin</td>
</tr>
<tr>
<td>/pkn/poa/retrieve-instance</td>
<td>retrieve-instance</td>
<td>yes</td>
<td>spars.execute*</td>
<td>(default)</td>
<td>test/turtle</td>
<td>admin</td>
</tr>
</tbody>
</table>

Add service
Summary of First Experiences with Neptune

Data scale
- 1.2 million products
- 120 million edges / triples
- Heterogeneous data

Query workload
- Real time queries against the graph for end-user frontend
- Analytical queries for data quality assessments

Standards-compliance
- Easy migration to Neptune via SPARQL 1.1 Protocol
Thanks for your attention! Questions?

Dr. Thomas Hubauer
Portfolio Project Manager

Dr. Alexey Fishkin
Senior Key Expert

Siemens AG
CT RDA BAM SMR-DE

siemens.com