Information Integration using a Linked Data approach

OSLC Tool-chain architecture for data-intensive computations

Jawad Munir
KTH, Royal Institute of Technology
Scania AB
ISO/IEC 15288 SE Process V-Model

Many Relationships & Overlaps between these Models

Models-based Systems Engineering Development

System Architecture
Requirements
Computer Aided Design (CAD) (e.g. 3D models, 2D drawings)
Management
Simulation Environment

Many More ...
Problem: Rollover Risk of Buses

• ECE R66.02 Valid from 2017, regulates the strength of the assembly in buses and the requirement of the roll-over test.
The strength of bus Superstructure

<<Requirement>>
Roof Crush Resistance
Bus superstructure section consist of four bays including rear entrance door, emergency door and rear end. Residual Space envelope requirement, in case of structural deformation in rollover test: residual space need to be 38.0 mm

<<Test Case>>
Accuracy test for bus superstructure model, quasi static load test

The bus superstructure (left) & CAD model of rear bus body section with 3D residual space envelop (right). The Models developed using CATIA software

<<verify>>
System Engineer define the test

Mechanical Engineer designs the bus geometrical model
Dynamic complete computer rollover simulation test for bus body

<<requirement>>
Vehicle in motion rollover test
--------------------------
The superstructure of the vehicle shall be of sufficient Strong so that in the event of rollover it doesn’t intrude inside the Residual space, for it, dynamic simulation should verify it at speed of 40km/h

<<testCase>>
Vehicle in motion rollover simulation

<<verify>>
System engineer defines simulation testCase

<<testCase>>
Vehicle in motion rollover simulation

Mechanical Engineer performs simulation with dynamic system model
Link between Residual Space parameter of Geometric Model and Simulation Model

Residual Space in Geometric Model  +  Computer simulation: inertia  →  Residual Space in simulation model
Relationships between Engineering Data

Requirements

- Roof Crush Resistance
- Vehicle in motion rollover test

Simulation Model

- Geometric Model

testCase

- Accuracy test: bus superstructure model, quasi static load test
- Vehicle in motion rollover simulation
Reality in Complex System Design: (Too Many) Relationships between Engineering Data
Network of Relationships
Impact Analysis

Change here
Queries

Query: Which elements are related to this requirement?
Overview of Development Process

Good Overview => Better Decisions
Scania – A case study

• 40000 employee
• Company in VM Group (Heavy Vehicles)

• In-house development
• Number of tools (in-house + commercial): 1000+

• The product: a truck in many versions and configurations
• Continuous Development and integration
Tradional Tool and Data Integration

• Point-to-Point connections
  • Don’t scale
  • limited coverage

• Tight Coupling
  • Dependence on internal structures and meanings
  • Inability to identify crucial entities (components, locations etc.)

• Monocultures
  • Vendor lock-in
  • Processes and methods needs to follow the tool
  • Incomplete w.r.t domains
  • What to do with in-house tools? – results in Data duplication

• There is no (can not be a) single Enterprise Information Model
Relationships on the Web

Linked Web Pages
(Unstructured Data)

Linked Data
(Structured Data)

Facebook
Open Graph
Example Linked Data: Dbpedia Query

**Dbpedia Query:** Give me name and date of establishment of universities in Sweden?

```
SELECT DISTINCT ?name ?established
WHERE {
  [] dbpprop:country dbpedia:Sweden;
    rdf:type dbpedia-owl:University;
    dbpprop:nativeName ?name;
    dbpprop:established ?established.
}
ORDER BY ?established
LIMIT 50
```
Tim Berners-Lee, Linked Data principles

- Tim berners-lee, Linked Data Principles
  1. Use **URIs** to identify the “things” in your data
  2. Use **http:// URIs** so people (and machines) can look them up on the web
  3. When a URI is **looked up**, **return** a **description** of the thing (in RDF format)
  4. Include **links to related things**
Open Services for Lifecycle Collaboration

• OSLC = Reusing the Web for tool integration
• Based on Web standards (Linked Data and RESTful Web Services)
• Initiated by IBM, It is OPEN-initiative
• Adopted by many tool vendors like Ericsson, IBM, Airbus etc ...
• Many tools have OSLC support like Jira, Jenkins, IBM rational tools etc ...
• Managed by OASIS
The OSLC Technology Stack

Domain Specifications
- Requirements Management
- Quality Management
- Change Management
- ...

OSLC Core Specification
- CRUD for resources
- Tracking changes
- Common Resources
- ...

OSLC

Linked Data
- Query: SPARQL
- Ontology: OWL
- Rule: RIF
- Data interchange: RDF, XML
- URI / IRI

REST
How OSLC works?

- OSLC defines set of Specifications and best practices for using linked data approach for tool integration and interoperability.

**Operations on OSLC Resources**

- **Publishing** OSLC resources
- **Retrieving** OSLC resources
- **Linking** OSLC resources across tools
- **Tracking changes** to OSLC resources
Thesis Project

• Assumption – Building an OSLC Tool-chain data integration architecture for performing complicated cross domain queries.

• Goal
  - Identify Integration Scenario for tool integration
  - Develop a prototype to support that Integration Scenario (Proof of Concept)
  - Test the prototype with cross domain queries
Integration Scenario – Scope + Tools involved

- We want to link **MSC Drawer** with **Sesamm Tool**
- Both tools have been developed by RESA
OSLC Adaptor for Sesamm Tool

• Sesamm OSLC adaptor publish two resources to the web
  • Components – ECU’s
  • UF
Sesamm Tool OSLC Adaptor

Example: Showing Sesamm Tool Adaptor Resources

R: UF

http://.../sesammuserfunctionCompleteName

http://.../sesammhascomponents

http://./userFunctions/249258

http://w3.../rdf#type

http://www.sesammuserFunction

http://.../sesammComponents/E30

http://.../sesammComponents/E113

http://.../sesammComponents/O1

http://.../sesammComponentCode

http://.../sesammfullname

http://w3.../rdf#type

OSLC LINK

Intra Tool

R: Component

O1

Tachograph

http://www.sesamm.com/sesamm/sesammComponent
OSLC Adaptor for MSC Drawer

• MSC OSLC adaptor publish one resource to web
  • MSC – SCN of an UF
Linking Resources between MSC Drawer and Sesamm Tool – Detail

OSLC Adaptor for MSC-Drawer
Resource: MSC

http://.../mscs/100

http://.../mscdrawerfullname

http://.../mscdrawerhascomponents

http://.../mscdrawertype

http://www.mscdrawer.com/mscdrawermsc

OSLC Link Tool

http://.../sesammComponents/E30

http://.../sesammComponents/T104

http://.../sesammComponents/O1

http://.../sesammComponentCode

Tachograph

OSLC Adaptor for Sesamm Tool
Resource: Component

http://www.sesamm.com/sesammFactory

http://www.sesamm.com/sesammComponent
Tracking changes to OSLC Resources

• Tracked Resource Set (TRS) Specification is used to track changes.
• Part of OSLC core Specification.
• The specification allows a OSLC server to expose an exact set of resources (TRS), track additions to and removals from the set, and track changes to the resources in the set.

<table>
<thead>
<tr>
<th>TrackedResourceSet Properties</th>
<th>Prefix Name</th>
<th>Occurs</th>
<th>Read-only</th>
<th>Value-type</th>
<th>Representation</th>
<th>Range</th>
<th>Description</th>
</tr>
</thead>
</table>
OSLC Tool-Chain Architecture for data integration

Diagram:
- Sesamm Database
- MSCs
- TRS Provider
- Tool Cache
- Data Analyzer & Graphical Interface (RELM)
- Integrated life-cycle Data Cache (Tool-chain cache)
- Data Indexing & Query Engine (LQE)
RELM is an analysis tool, which helps engineering teams to visualize, analyze, and organize engineering data and their relationships.

Client Application LQE indexes TRS

Index (Lifecycle Query Engine) → Tracked Resource Set

Analysis Tool

Linked Lifecycle Data

Rational Engineering Lifecycle Manager (RELM)

Visualize
Cross Domain Navigator / Explorer

Analyze
Query, Reporting, and Impact Analysis

Organize
Product and System Definition

Open Lifecycle Integration
LifeCycle Query Engine

- Indexes resources from TRS providers (TRS Client)
- Indexed resources are represented as RDF
- Exposes an `Endpoint` for running queries on that data
  - **Endpoint**: A server that exposes its data via the SPARQL protocol
Test Queries: Cross Domain

Query 1:
“for all the components in one UF, give me all the MSCs where the components exists?”

Query 2:
“for all components in one MSC, give me all the UFs where the components exists?”

Query 3:
“for all components in UFs starting with UF1** give me all the MSCs where the components exists?”
DEMO – OSLC Tool-chain prototype
## Results

<table>
<thead>
<tr>
<th>Sr No.</th>
<th>Queries</th>
<th>Time of Execution (In seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“for all the components in one UF, give me all the MSCs where the components exists?”</td>
<td>&lt;1</td>
</tr>
<tr>
<td>2</td>
<td>“for all components in one MSC give me all the UFs where the components exists?”</td>
<td>&lt;1</td>
</tr>
<tr>
<td>3</td>
<td>“for all components in UFs starting with <strong>UF1</strong> give me all the MSCs where the components exists?”</td>
<td>&lt;10</td>
</tr>
</tbody>
</table>

**Table: Query Processing Time for each cross domain query**
### Results

<table>
<thead>
<tr>
<th>Sr No.</th>
<th>Data Source</th>
<th>Indexing time (hh:mm:ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRS Provider for Sesamm Tool</td>
<td>2:30</td>
</tr>
<tr>
<td>2</td>
<td>TRS Provider for MSC Drawer</td>
<td>1:10</td>
</tr>
</tbody>
</table>

**Table: Data indexing time for each tool**
Conclusions

• Data Consistency and Tracebility across the tool-chain.
• Capability of computation intensive, cross domain queries.
• Caching at individual Tool level and at Tool-Chain level.
• A federated approach to SPARQL query processing: Federated over multiple single repositories.
• LQE is dead-end of the Tool-Chain
• FO, SA or ST were not interested in RELM as an analysis tool.
Conclusions

• **Key advantages**
  • Open Technologies (Using Open Web Standards XML, RDF, SPARQL etc...)
  • Dramatically reduce data integration costs, increase enterprise flexibility.
  • Scalability

• **Cons**
  • No open implementation for TRS Client exists.
  • Patent for TRS Client specification.
  • Installation of IBM RELM can be frustrating
  • No Interface for machines (APIs etc.)
  • SPARQL is still an expensive skill
Future Work

• HTTP PUSH capability in TRS Protocol
  • Tool data don’t change so often
  • Safety critical systems – instant update necesary

• Implantation of TRS Client
  • Interface for humans
  • Interface for machines (APIs etc.)
  • Analysis tool (Graphical tool) – Customizable for Specific needs.

• Further Development of prototype: further develop existing Adaptors + Including other tools
Master Thesis presentation

ends with

says

did

gets

Thanyou for your Attention!

want
Appendix
Alternatives

- **SDshare** – A Specification used for publishing and consuming resource descriptions
- A protocol for tracking changes in semantic data store
- Based on Linked Data principles
- Defines how an RDF system can maintain a synchronized local copy of master data contained in another RDF system
- Not an active community behind the standard

- **Mature Product:**
  - Hafslund Sesam
SESAM is an architectural approach and set of components for integrating, querying, reusing and repurposing enterprise data. SESAM enables the data from many sources to be seamlessly merged together in a schema-less, and scalable data store, that we call the DataHub.

There is, however, more to it than just this.
Sdshare – More Details

• SDshare spec
  • http://www.egovpt.org/fg/CWA_Part_1b

• SDshare issue tracker
  • http://projects.topicmapslab.de/projects/sdshare

• SDshare use cases
  • http://www.garshol.priv.no/blog/215.html
Approaches to Federation

a) Integration in a central repository

b) Federation over multiple single repositories

c) Federation over multiple SPARQL endpoints