LDBC Consortium Fourth Technical User Community (TUC) meeting

“Customer experiences in implementing SKOS-based vocabulary management systems, and other Semantic-Technology-Driven Systems ....”

Ralph Hodgson, TopQuadrant

CWI, Amsterdam, April 3, 2014
Content

- Introductions
- Overview of TopBraid
- Case Study Overviews
- Deeper Dive on One Case Study: EPIM
- Concluding Remarks
Introductions

Me - Ralph Hodgson
- co-founder and CTO of TopQuadrant, Inc., a US-headquartered company that specializes in semantic solutions, consulting, training, and platforms;
- NASA QUDT Ontologies and Handbook Lead
- EPIM Lead Semantic Applications Architect and Ontology Modeler

TopQuadrant
- Semantic Technology Products, Solutions and Tools, training and consulting, headquartered in the US with offices in Europe
- Innovator of SPARQL-based technologies: SPIN, SPARQLMotion, SWP, SWA
The TopBraid Technology Stack

Applications

Data Processing Pipelines
- SPARQLMotion

Templates, UI Components
- SPARQL Web Pages (SWP), SWA

Inference Engines, Data Mapping Rules
- TopSPIN, BackSPIN, SPINMap

SPARQL-based Templates, Functions, Rules, Constraints
- SPARQL Inferencing Notation (SPIN)

Query Engine
- SPARQL

Model: Classes, Properties
- RDFS, OWL

Change Management
- Teamworks Framework

Data Layer: Base Infrastructure, Databases
- RDF

© Copyright 2014 TopQuadrant Inc.
TopBraid Services

**PRESENTATION SERVICES**
- Provide user interfaces for rendering Semantic Web data
- Provide Entity Services
- Enable querying via the SPARQL language, including SPARQL Endpoints

**RULES SERVICES**
- Define rules and constraints on information and data
- Execute rules and constraints
- Resource-centric web-services

**MODEL SERVICES**
- Manage consolidation of graphs, temporary and persistent triple stores
- Manage Core concept models to which all concepts are mapped
- Provide services for structured vocabularies for disambiguation of terms
- Manage Schema, Dataset, Transformation, and Proxy Models
- Provide services for mapping between concepts and locations

**FEDERATION SERVICES**
- Perform federated queries across multiple databases
- Manage data source locations
- Provide on-demand data access

**DATA SERVICES**
- Query via the SPARQL language
- Access data via connectors for RDFa, RSS/Atom, Email, RDBMS, CMS, Excel, HTML/CalaisText, CSV, UML, XML
- Access data via SPARQL EndPoints including LinkedOpenData
- Extract indices on external data into local triple store
- Redirect from persistent to actual location

**APPLICATION INTEROPERABILITY SERVICES**
- Scripting of data processing pipelines
- Integrate with enterprise applications, e.g., workflow engines, search engines
- Extract transform load and map into local triple stores
- Provision customised web services
Illustrative Enterprise Customers

Digital Media
- OECD
- Thomson Reuters
- Alcatel-Lucent
- U.S. Air Force
- Pearson
- The Church of Jesus Christ of Latter-day Saints

Life Sciences
- AstraZeneca
- Vistakon
- Johnson & Johnson
- P&G
- Lilly
- Syngenta

Other
- EPIM
- Teachers' Pension Plan
- NASA
- EURIWARE
- J.P. Morgan
- Lockheed Martin
These customers use TopBraid to:

- **Mayo Clinic:** “re-integrate and enhance access to knowledge across research, education and clinical practice”

- **Syngenta:** “help scientists to develop insights into research data using databases and information sources – both internal and external”

- **EPIM:** “establish a standards-based knowledge platform for data exchange – receiving, validating, storing, analyzing and transmitting reports”

- **OTPP:** “enable data to be searched without a PhD in SQL”
Case Study: Mayo Clinic

Went Live 5 January 2014 with over 5.6 million page views per day!

Enhance Value of Mayo’s Knowledge Initiative: Knowledge Content Management System (KCMS)

- Enhanced Search
- Taxonomy management
- Run-time terminology services
KCMS Conceptual Architecture

- **TopBraid EVN (single server)**
  - Terminology governance and content tagging

- **TopBraid Live (multiple servers)**
  - Serving production terminologies in real time for enhancing search and content enrichment.
Typical EVN SKOS Customers

- Don’t necessarily or always use SKOS – meta and master data management is an important area for our customers and SKOS doesn’t fit well
- Use between 5 and 100 vocabularies
- Have vocabularies of varying sizes, on average, less than 10 M triples overall
- Has complex requirements for management of changes, promotions of changes, comparison, user permissions, rollback, data quality rules, constraints, validations
  - As a result, edit time user interfaces can rarely be delivered by a single query
- Often uses a combination of EVN (for editing) and TopBraid Live (for read only web services access)
EVN is more than SKOS: an example configured for an ontology of oil and gas facility asset management
EVEN is an extensible model-driven application: Example read service (SKOS)

TopBraid EVEN ships with > 30 template services, e.g.:

```
CREATEVIEW ConceptsAndBroaderConceptsForResourceAndProperty {
  SELECT DISTINCT ?concept
  WHERE {
    ?concept a ?type .
    ?type rdfs:subClassOf ?skos:Concept .
  }
}
```

Creating a new service can take as little as an hour (often less) – customers implement their own custom services based on SPARQL templates.
Case Study: Syngenta

- Many disconnected and diverse data sources
- Need to solve problems by using insights built from the components of meaning - questions, concepts, perspectives

Topbraid Insight offers a layer of connection and meaning to the user and protects them from the distracting mechanics of data access.
Model-driven Data Virtualization: TopBraid Insight provides a unified query interface over multiple data sources

- Real time, on-demand datamarts
- Documents/IP
- Physical Data Warehouse
- Sample Test Results
- Any source
- LDBC SIDER
- Chembl
- Chebi
- DrugBank
Queries may need to be ‘staged’ and/or ‘nested’

- Multiple queries may be involved by exploring links
- Federated data sources are often relational
  - some may be SPARQL endpoints and other sources
Case Study: EPIM

EPIM is the instrument for the operators on the Norwegian Continental Shelf to secure efficient information sharing among all relevant stakeholders by providing cost effective and user friendly common digital solutions based on international standards.

- EPIM ReportingHub (ERH)
- License2Share (L2S)
- EqHub
- EnvironmentWeb
Supporting EPIM’s Vision for Oil & Gas Solutions

• Build a shared suite of knowledge based-applications using Semantic technology and industry-standard domain concepts
  – i.e. a **Semantic Ecosystem** for the Oil and Gas Industry on the Norwegian Continental Shelf
EPIM ReportingHub (ERH) Architecture – in production for nearly two years now

Operators on the NCS

Semantic Processing

Authorities

License Partners

ISO 15926 Inside
Scale of ERH Solution

- 300 million triples in the next 4 years
  - Potential for many more if data sources added or historical data imported
- 40+ concurrent users
  - User interfaces and Service interfaces required
  - Potential for many more if data sources added
- Phase 1 XML Schemas have 2000-ish elements
  - resulting ontologies have 900-ish classes and 900-ish properties
- Delivered on an SaaS basis with high availability Service Level Agreement
  - Secondary app server running at all times
  - RDF database replication and warm failover
Co-existing in the XML World

1. Make OWL Schemas from XSD Schemas
   - Import XSDs
   - Transform
   - Validate

2. Use the OWL Schemas to make RDF from XML Messages
   - Import XML
   - Transform
   - Validate

LDBC
### EPIM EnvironmentHub (EEH) Example User Interface

#### Reporting Status

<table>
<thead>
<tr>
<th>Field</th>
<th>Error</th>
<th>Structure types (Inactive wellbores not included)</th>
<th>Inactive Wellbores reported</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alve</td>
<td>0</td>
<td>1/1, 119/119, 147/147</td>
<td>267/267</td>
<td>1</td>
</tr>
<tr>
<td>Brage</td>
<td>2</td>
<td>0/1, 112/119, 137/147</td>
<td>249/267</td>
<td>2</td>
</tr>
</tbody>
</table>

#### Facilities not complete

<table>
<thead>
<tr>
<th>Facility</th>
<th>Applicable Structure types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brage A</td>
<td>Oily Water Discharge Jetting</td>
</tr>
<tr>
<td>Brage B</td>
<td>Oily Water Discharge Jetting, Accute pollution to Sea</td>
</tr>
<tr>
<td>Brage C</td>
<td>Oily Water Discharge Jetting, Combustion</td>
</tr>
<tr>
<td>Brage D</td>
<td>Oily Water Discharge Jetting</td>
</tr>
<tr>
<td>Brage E</td>
<td>Oily Water Discharge Jetting</td>
</tr>
</tbody>
</table>

#### Other Fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Error</th>
<th>Structure types</th>
<th>Total</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fram</td>
<td>3</td>
<td>0/1, 17/119, 7/98</td>
<td>24/218</td>
<td>Submitted</td>
</tr>
<tr>
<td>Giltne</td>
<td>0</td>
<td>1/1, 85/85, 105/105</td>
<td>191/191</td>
<td>Submitted</td>
</tr>
<tr>
<td>Grane</td>
<td>0</td>
<td>1/1, 85/85, 140/140</td>
<td>226/226</td>
<td>In progress</td>
</tr>
<tr>
<td>Gullfaks</td>
<td>-</td>
<td>0/1, 0/85, 0/140</td>
<td>0/226</td>
<td>No started</td>
</tr>
<tr>
<td>Heidrun</td>
<td>-</td>
<td>0/1, 0/85, 0/140</td>
<td>0/226</td>
<td>No started</td>
</tr>
<tr>
<td>Heimdal</td>
<td>0</td>
<td>1/1, 119/119, 147/147</td>
<td>267/267</td>
<td>In progress</td>
</tr>
<tr>
<td>Huldra</td>
<td>0</td>
<td>1/1, 85/85, 105/105</td>
<td>191/191</td>
<td>In progress</td>
</tr>
<tr>
<td>Hyme</td>
<td>0</td>
<td>1/1, 85/85, 105/105</td>
<td>191/191</td>
<td>In progress</td>
</tr>
<tr>
<td>Kristin</td>
<td>0</td>
<td>1/1, 85/85, 140/140</td>
<td>226/226</td>
<td>In progress</td>
</tr>
<tr>
<td>Kvitebjørn</td>
<td>4</td>
<td>0/1, 68/85, 140/140</td>
<td>208/226</td>
<td>In progress</td>
</tr>
</tbody>
</table>
3.1 Utslipp av olje

Tabell 3-1 gir en oversikt over utslipp av oljeholdig vann fra feltet i rapporteringsåret. Oljeholdigvann fra jetting måles ikke separat, men er inkludert i mengde produsert vann til rensing og utslipp. Drenasjevannet blir ledet til produsertvannssystemet for rensing før utslipp, det er derfor ingen egne beregninger av korrelasjonsfaktor for drenasjevann.

<table>
<thead>
<tr>
<th>Vanntype</th>
<th>Totalt vannvolum (m³)</th>
<th>Midlere oljeinnhold (mg/l)</th>
<th>Midlere oljedeheng på sand (g/kg)</th>
<th>Olje til sjø (tonn)</th>
<th>Injisert vann (m³)</th>
<th>Vann til sjø (m³)</th>
<th>Eksportert prod. vann (m³)</th>
<th>Importert prod. vann (m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produsert</td>
<td>996 928</td>
<td>6.44</td>
<td></td>
<td>6.42</td>
<td>0</td>
<td>997 582</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fortregning</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drenasje</td>
<td>7 348</td>
<td>6.81</td>
<td></td>
<td>0.06</td>
<td>0</td>
<td>8 708</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Annet</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1 004 275</strong></td>
<td></td>
<td></td>
<td><strong>6.48</strong></td>
<td>0</td>
<td><strong>1 006 289</strong></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
SELECT ?Vanntype ?Totalt_vannvolum_UNITm3_SUM
((IF((?Vann_til_sj_UNITm3_SUM != 0.0), ((?Olje_til_sj_UNITtonn_SUM / ?Vann_til_sj_UNITm3_SUM) * 1e6), ""))
AS ?Midlere_oljeinnhold_UNITmg_per_l)
((IF((1.0 != 0.0), (?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM / ?counter_SUM), 0.0))
AS ?Midlere_oljevedheng_p_sand_UNITg_per_kg)
?Olje_til_sj_UNITtonn_SUM ?Injisert_vann_UNITm3_SUM ?Vann_til_sj_UNITm3_SUM
?Eksportert_prod_vann_UNITm3_SUM ?Importert_prod_vann_UNITm3_SUM
FROM <http://www.environmenthub.no/env/schema/1.0/interface-model>
FROM <http://www.environmenthub.no/data/npd/facts-interface>
FROM <http://eeh.testing/ttl/3.1-Oseberg>
WHERE {
  {
    ?Vanntype spif:split (?Vanntype_LIST "," ) .
  }
  SELECT ?Vanntype
    ((SUM(?Totalt_vannvolum_UNITm3_SUM_TMP)) AS ?Totalt_vannvolum_UNITm3_SUM)
    ((SUM(?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM_TMP)) AS ?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM)
    ((SUM(?counter_TMP)) AS ?counter_SUM)
    ((SUM(?Olje_til_sj_UNITtonn_SUM_TMP)) AS ?Olje_til_sj_UNITtonn_SUM)
    ((SUM(?Injisert_vann_UNITm3_SUM_TMP)) AS ?Injisert_vann_UNITm3_SUM)
    ((SUM(?Vann_til_sj_UNITm3_SUM_TMP)) AS ?Vann_til_sj_UNITm3_SUM)
    ((SUM(?Eksportert_prod_vann_UNITm3_SUM_TMP)) AS ?Eksportert_prod_vann_UNITm3_SUM)
    ((SUM(?Importert_prod_vann_UNITm3_SUM_TMP)) AS ?Importert_prod_vann_UNITm3_SUM)
  WHERE {
    {
      ?Vanntype spif:split (?Vanntype_LIST "," ) .
    }
  ....
EEH Example Query – Generating a Report Table (KLIF 3.1) – 3 of 3

?oilyWaterDischarge a eeh-ui:OilyWaterDischargeProducedDrainageDisplacement .
  ?oilyWaterDischarge eeh-ui:oilyWaterDischargeProducedDrainageDisplacement-WaterType
    ?WaterType_CONSTRAINT .
  ?WaterType_CONSTRAINT rdfs:label ?WaterType_VALUE .
  FILTER eeh-lib:MatchStrings(?WaterType_VALUE, ?Vanntype) .

...
...
}
UNION
{
  ?oilyWaterDischarge a eeh-ui:OilyWaterDischargeJetting .
  EXISTS {
    ?oilyWaterDischarge eeh-ui:oilyWaterDischargeJetting-OilOnSolidParticles-g_per_kg|eeh-ui:oilyWaterDischargeJetting-
      OilToSeaISOmethod-tonnes ?testForValue .
  } .
  OPTIONAL {
    ?oilyWaterDischarge eeh-ui:oilyWaterDischargeJetting-OilOnSolidParticles-g_per_kg
      ?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM_TMP_MAYBE .
  } .
  BIND (COALESCE(?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM_TMP_MAYBE, 0.0) AS
    ?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM_TMP) .
  BIND (IF(bound(?Midlere_oljevedheng_p_sand_UNITg_per_kg_SUM_TMP_MAYBE), 1, 0) AS ?counter) .
  OPTIONAL {
  } .
  BIND (COALESCE(?Olje_til_sj_UNITtonn_TMP_MAYBE, 0.0) AS ?Olje_til_sj_UNITtonn_SUM_TMP) .

}
GROUP BY ?Vanntype
Example of using SPIN Functions

```
WHERE {
    ((eeh-lib:sortStringListRemovingDuplicates(GROUP_CONCAT(?Nuklides; SEPARATOR=','))) AS ?Nuklider)
    ((SUM(?Sendt_til_land_UNITtonn_SUM_TMP)) AS ?Sendt_til_land_UNITtonn_SUM_TMP2)
    ((SUM(?Total_aktivitet_UNITGBq_SUM_TMP)) AS ?Total_aktivitet_UNITGBq_SUM_TMP2)
  WHERE {
    ?hazardousWasteOther a eeh-ui:HazardousWasteOther .
    ?yearData eeh-ui:facilityYearHazardousWasteOther ?hazardousWasteOther .
    OPTIONAL {
    } .
    BIND (COALESCE(?Beskrivelse_MAYBE, "") AS ?Beskrivelse) .
    FILTER fn:starts-with(?Avfallsnummer, "3") .
    OPTIONAL {
    } .
    BIND (COALESCE(?Nuklides_MAYBE, "NULL") AS ?Nuklides) .
    BIND (COALESCE(?Sendt_til_land_UNITtonn_SUM_TMP1, 0) AS ?Sendt_til_land_UNITtonn_SUM_TMP) .
...
  }
  GROUP BY ?Avfallstype ?Beskrivelse ?Avfallsnummer
  ORDER BY (?Avfallsnummer) (?Beskrivelse)
} .
BIND (IF((datatype(?Total_aktivitet_UNITGBq_SUM_TMP2) != xsd:double), "", ?Total_aktivitet_UNITGBq_SUM_TMP2) AS ?Total_aktivitet_UNITGBq_SUM) .
```
Ontology Architecture for Composite Graphs

Simple View Ontologies for UI, reporting and analytics

SPARQL Query-View-Transformation

ISO-15926 4D Ontologies
Transforming is controlled by a SPARQLMotion Web Service (1 of 2)

SPARQL Views: Lifting the ISO 15926 Data Graphs into the Interface Model Representation.

- Iterate over all official field-year graphs
- “Lift” the Integration Model using SPIN transforms
- Bulk Load the transformed triples
- Use of W3C Provenance Ontology for traceability
Transforming is controlled by a SPARQLMotion Web Service (2 of 2)

SPARQL Views: Lifting the ISO 15926 Data Graphs into the Interface Model Representation.

- Total triples in the Integration Model (ISO-15926 Ontologies): ~ 4,000,000
- Total triples in the Interface Model: ~400,000
EPIM Logistics Hub (ELH)

Offshore Supply Chain: CCU Tracking

Hirer/Supplier (H)  CCU Owner (C)

Supply Base (S)  Operator (O)  Offshore Facility
Concluding Remarks

- SPARQL is more than a query language
- Model-driven applications exploit the real power of semantic web technologies
- Different kinds of solution types can be distinguished
- We need more kinds of benchmarks
Thank you!

- rhodgson AT topquadrant.com
- Twitter @ralphtq, @topquadrant
- www.scribd.com/ralphtq
- www.linkedmodel.org