TopQuadrant™

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

http://www.ldbc.eu:8090/display/TUC/Fourth+TUC+Meeting,+April+2014

v5

TopQuadrant™

Content

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

LDBC

Introductions

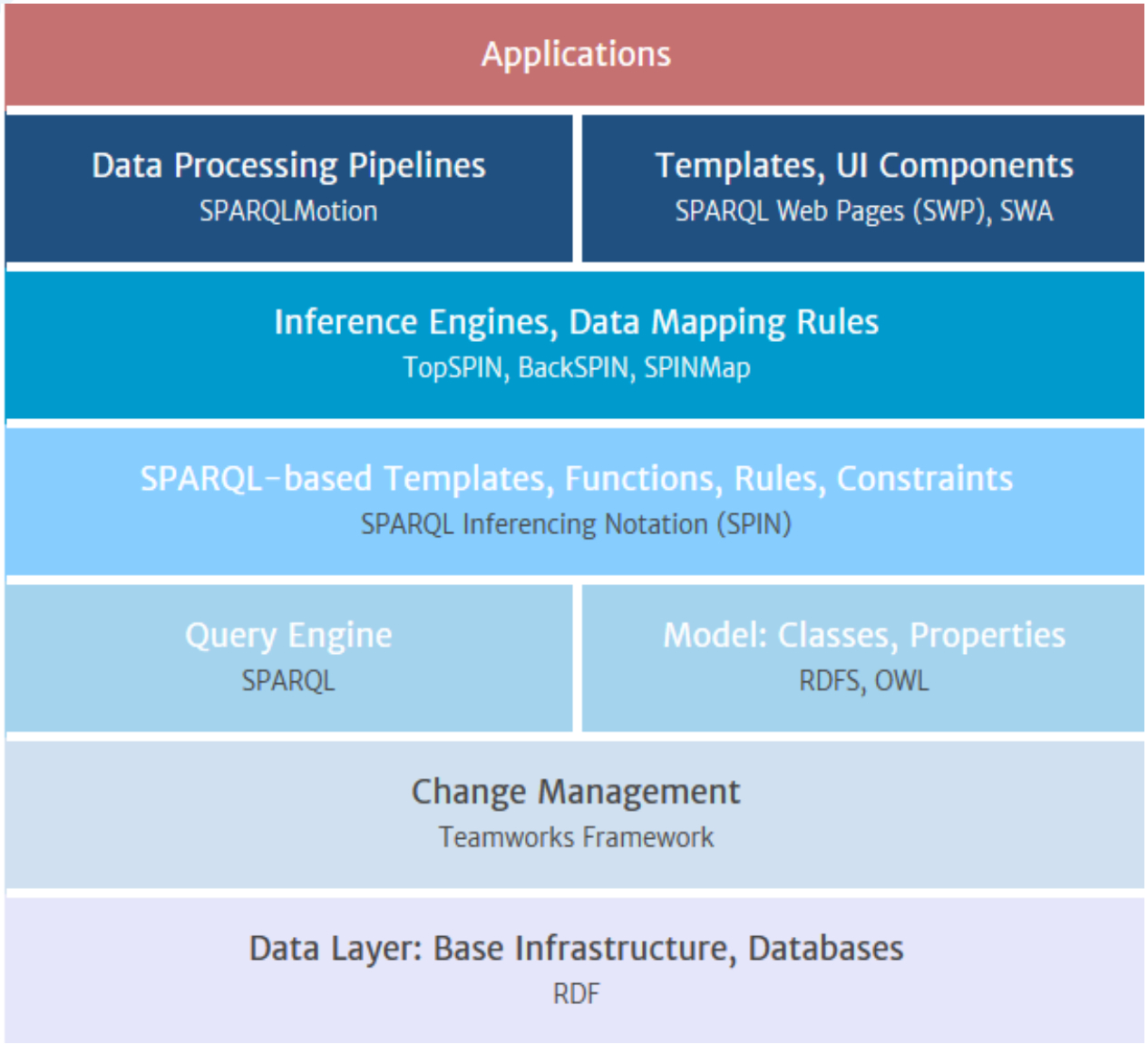
Me - Ralph Hodgson

- *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*

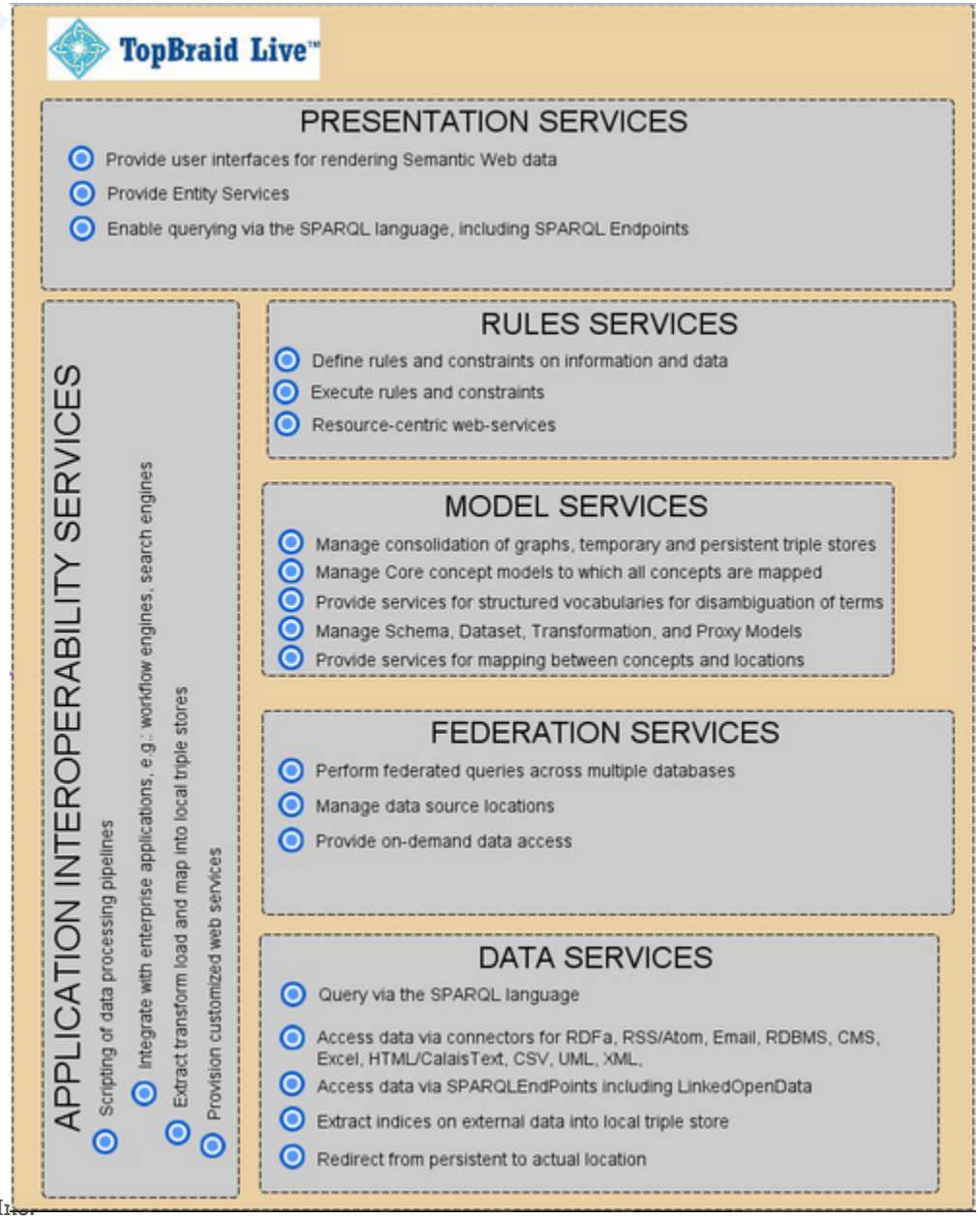
TopQuadrant

- *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



TopBraid Services



Illustrative Enterprise Customers

Digital Media

- OECD — BETTER POLICIES FOR BETTER LIVES
- THOMSON REUTERS
- Alcatel·Lucent
- U.S. AIR FORCE
- PEARSON
- MAYO CLINIC
- THE CHURCH OF JESUS CHRIST OF LATTER-DAY SAINTS

Life Sciences

- AstraZeneca
- VISTAKON — Division of Johnson & Johnson
- P&G
- ucb
- Lilly
- syngenta GLOBAL

Other

- EPIM — E&P Information Management Association
- ONTARIO TEACHERS' PENSION PLAN
- NASA
- EURIWARE
- J.P.Morgan
- LOCKHEED MARTIN — We never forget who we're working for®

LDBC

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)

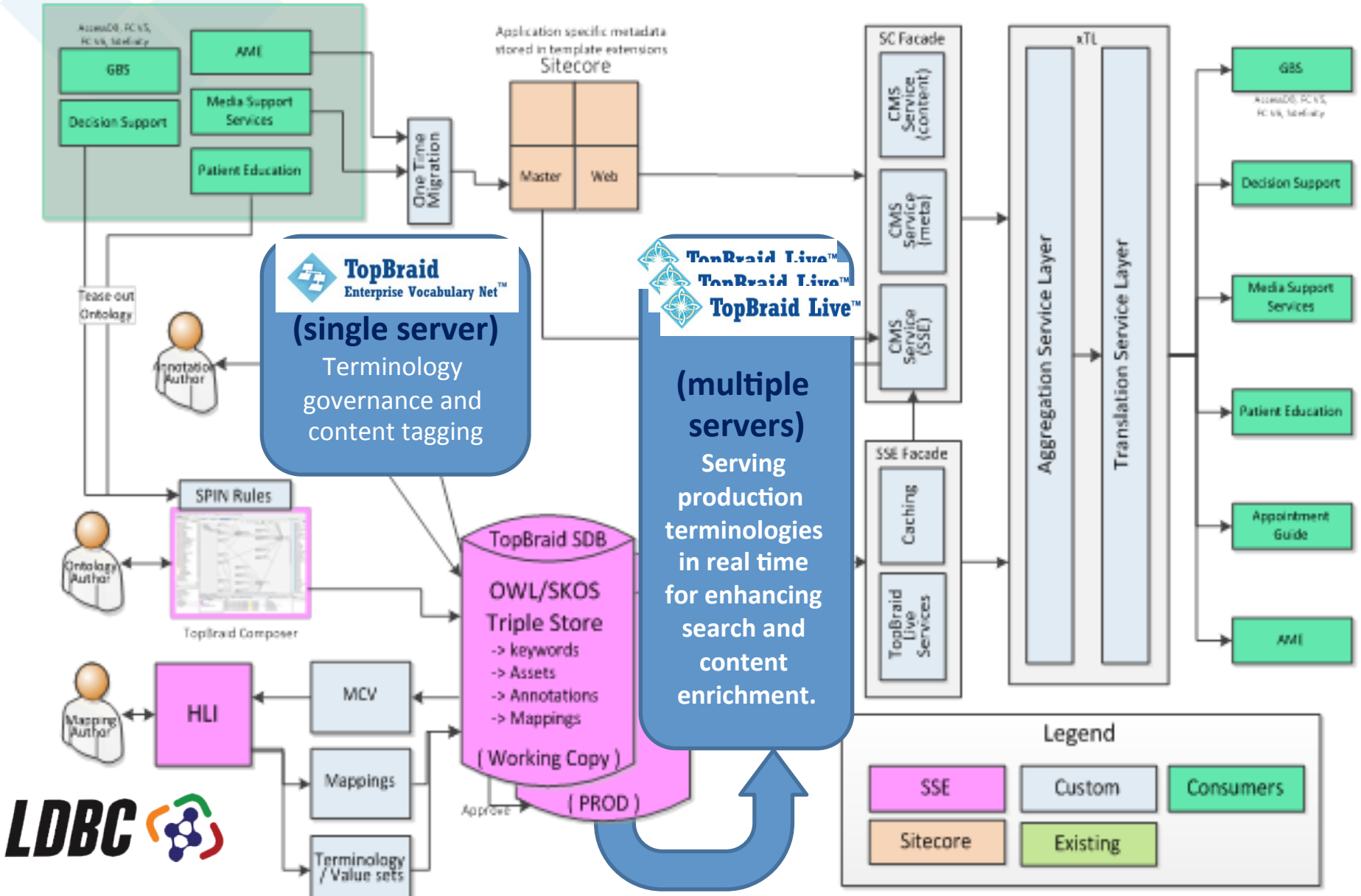
- Enhanced Search
- Taxonomy management
- Run-time terminology services

TopBraid Enterprise Vocabulary Net™

TopBraid Live™

LDBC

KCMS Conceptual Architecture



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

TopBraid Enterprise Vocabulary Net™

Navigation Hierarchy

- Unified Enterprise Asset Management
  - Contents
    - Assets
    - Documents
      - Electrical Diagram Document
      - Hydraulic Diagram. MLW.
      - Instruction for Wire Installation
      - Instructions for Installation. MLW.
      - Lubrication Chart. MLW.
      - Painting Specification 400T SHS
    - Equipments
      - Main Lift Winch
    - Function Locations
      - SHS
        - Electrical
        - Emergency
        - Hydraulic
        - Mechanical
          - Kingpost
          - Superstructure
            - Running Tool
            - Slew Motors
            - Winches
              - CGFW
              - Control Winch
              - GWW1
                - Sheave1
                - Spooling Gir
                - Winch
                - Brakes
                - Motors

MLW1 Use case

Special Handling System (Physical asset)

Properties

Name: Special Handling System
Description:
Image:

Function location: SHS
Equipment:
Related documents:
Work processes: SHS Commissioning
Comments:
Parent: Assets

EVN is an extensible model-driven application: Example read service (SKOS)

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

skostemplate:ConceptsAndBroaderConceptsForResourceAndProperty (spin:SelectTemplate)

Given a resource and a property, return all vocabulary concepts, and all concepts broader than the matched concepts, referenced by the resource/property pair.

Arguments

arg:property (rdf:Property): The property.
arg:resource (rdfs:Resource): The URI of the content object.

Service Syntax

```
template/skostemplate/ConceptsAndBroaderConceptsForResourceAndProperty?property=...&resource=...
```

or

```
template?_template=skostemplate:ConceptsAndBroaderConceptsForResourceAndProperty&property=...&resource=...
```

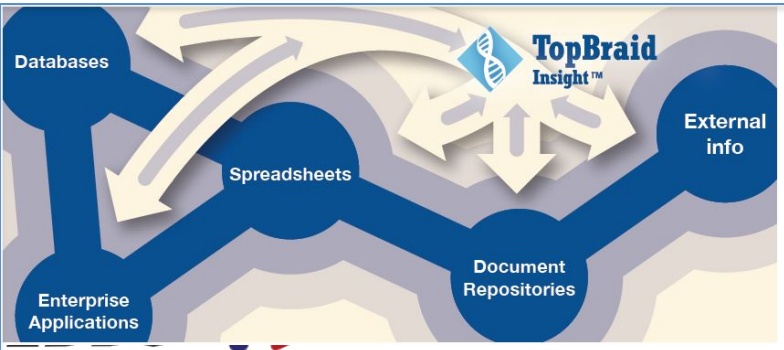
Template Body

```
SELECT DISTINCT ?concept
WHERE {
    ?resource ?property ?broaderConcepts .
    ?broaderConcepts (skos:broader)* ?concept .
    ?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.

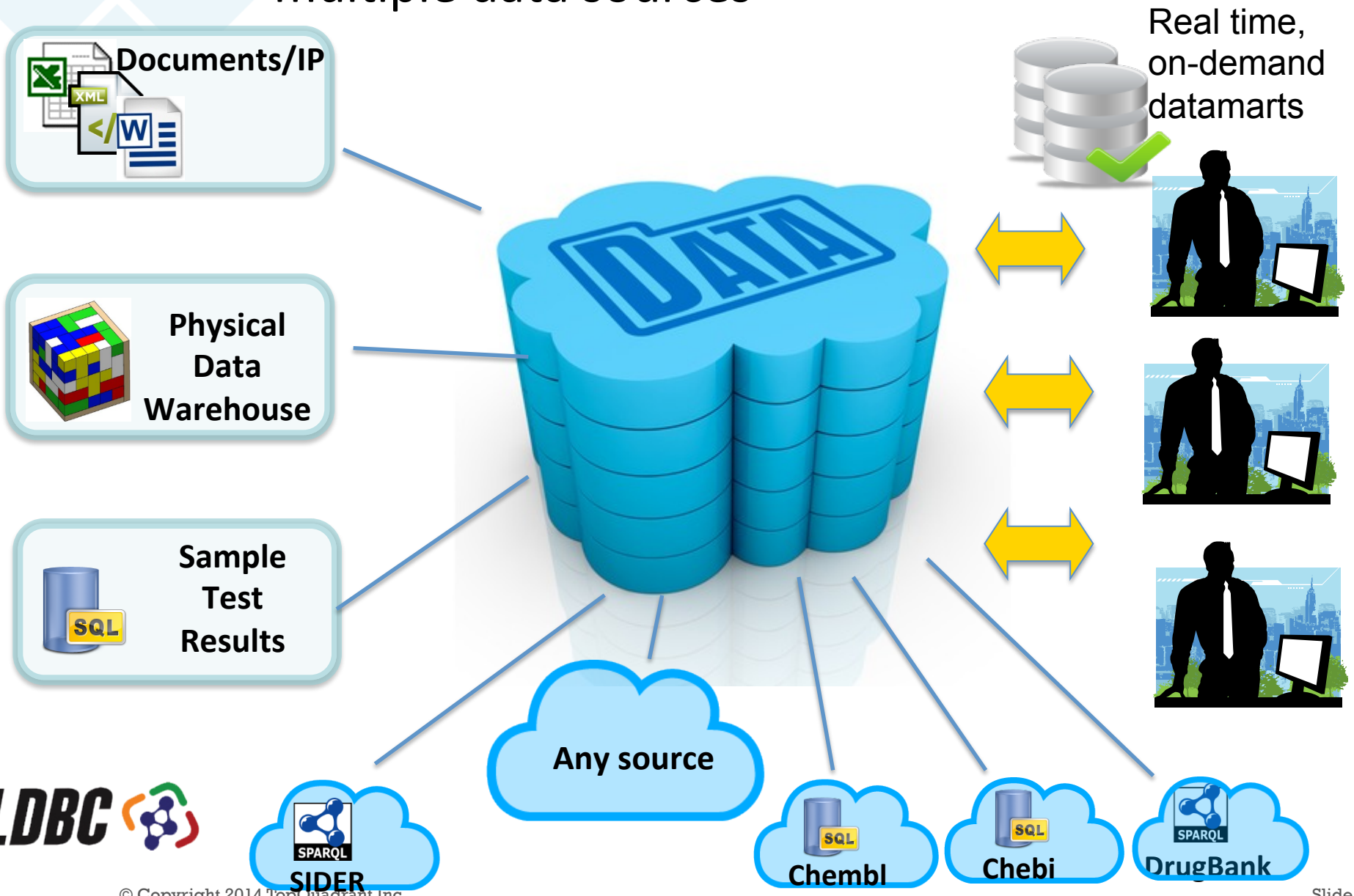
Case Study: Syngenta

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

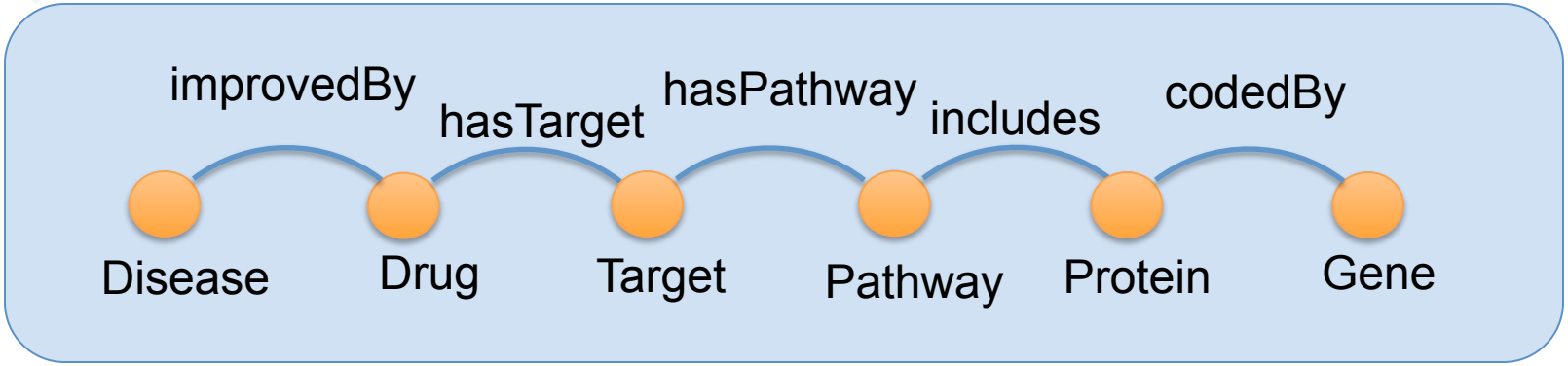


Topbraid Insight offers a layer of connection and meaning to the user and protects them from the distracting mechanics of data access.

Model-driven Data Virtualization: TopBraid Insight provides a unified query interface over multiple data sources



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
E&P Information Management Association

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

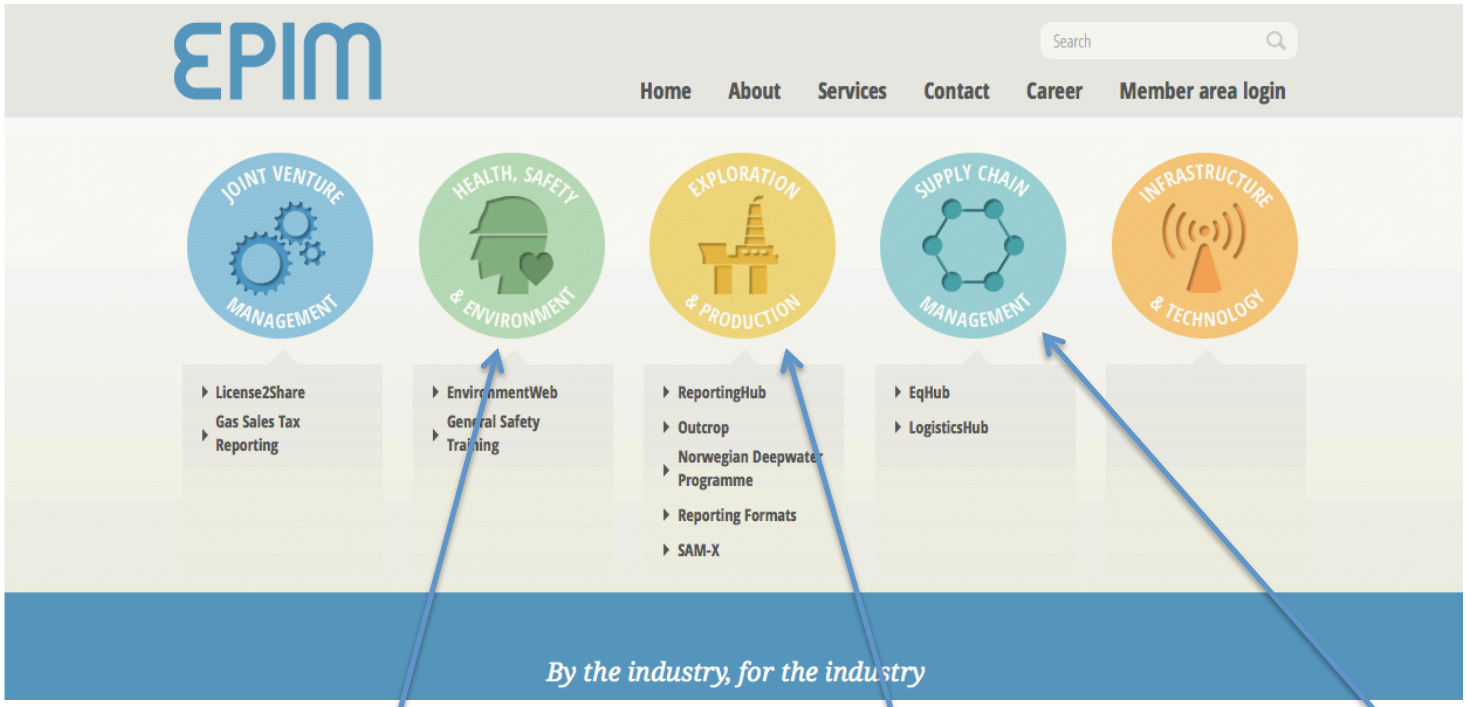
- EPIM ReportingHub (ERH)
- License2Share (L2S)
- EqHub
- EnvironmentWeb

EPIM E&P Information Management Association

LDBC

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

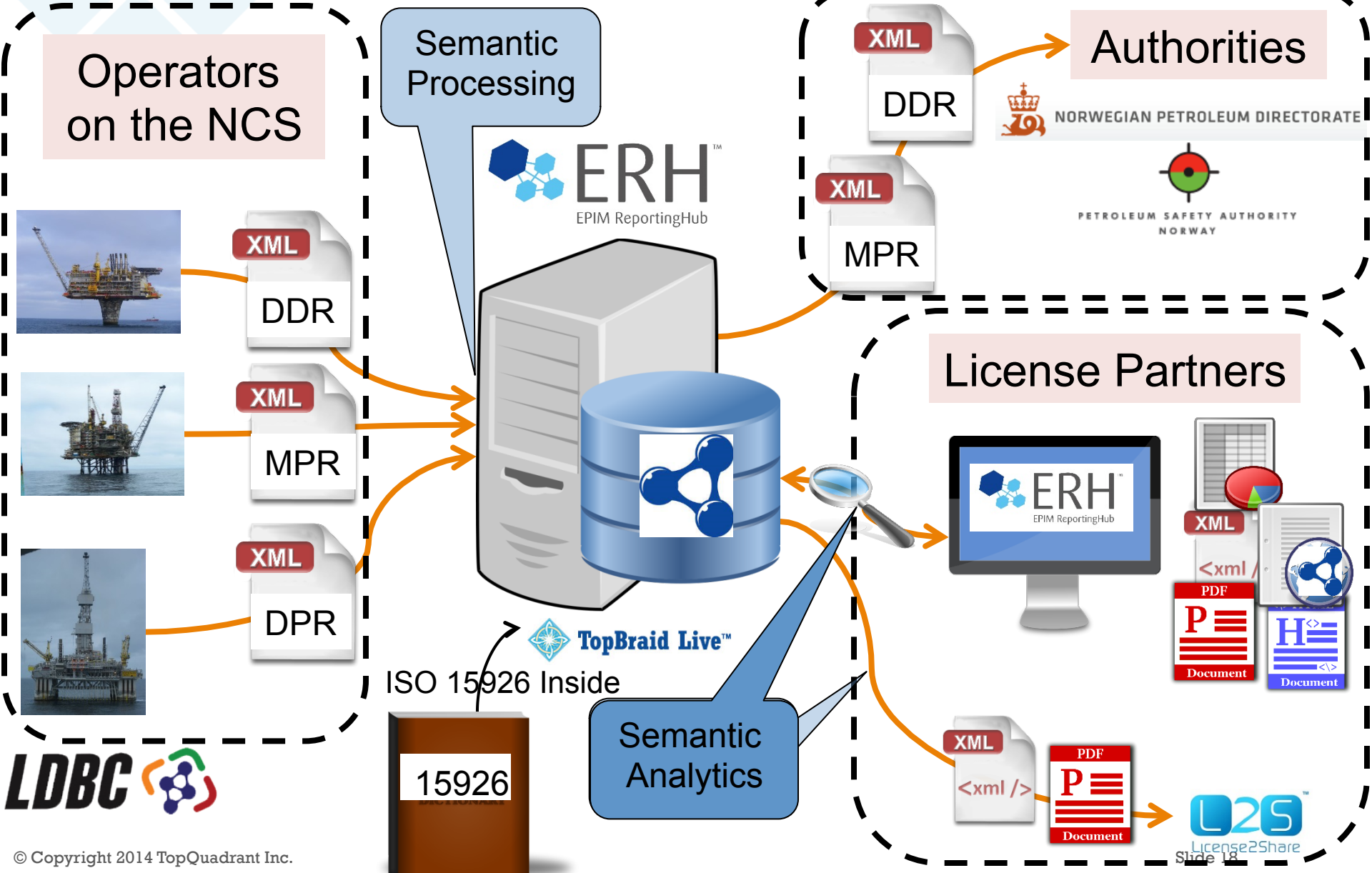


EnvironmentHub

ReportingHub

LogisticsHub

EPIM ReportingHub (ERH) Architecture – in production for nearly two years now



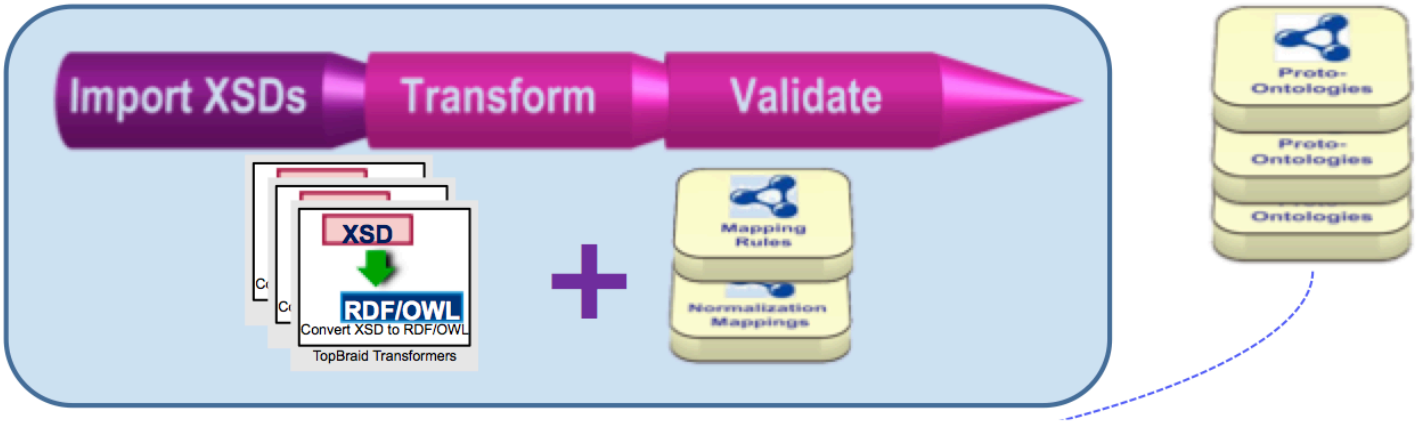
Scale of ERH Solution

- 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 fallover

Co-existing in the XML World

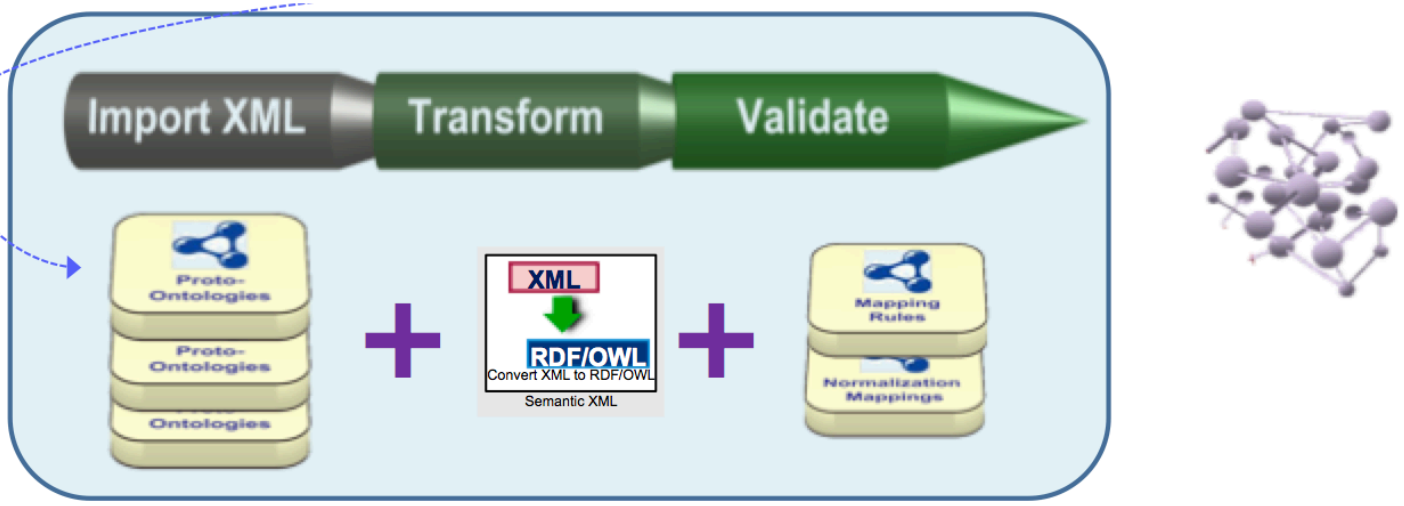
1

Make OWL Schemas from XSD Schemas

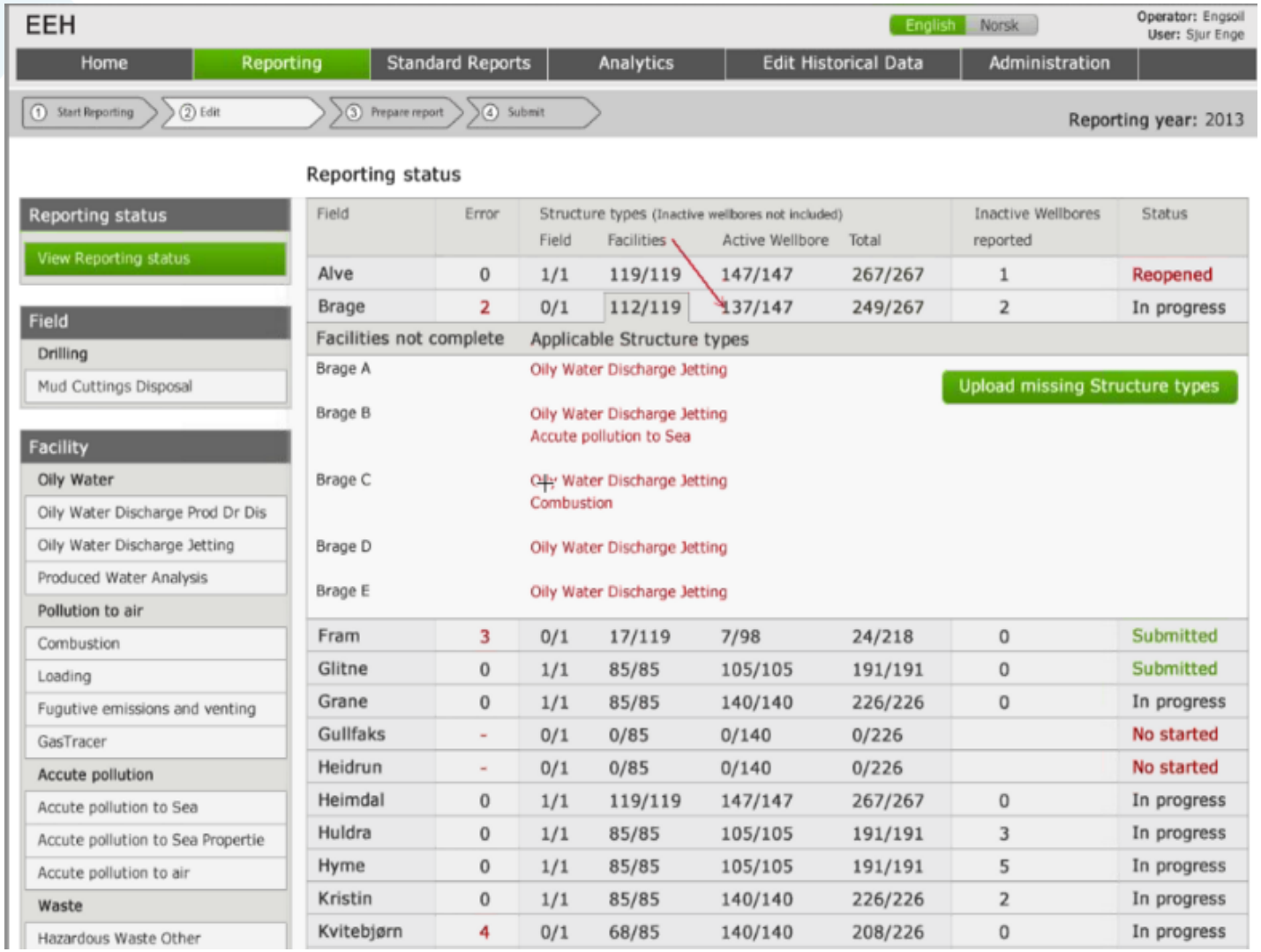


2

Use the OWL Schemas to make RDF from XML Messages



EPIM EnvironmentHub (EEH) Example User Interface



EEH Example Query – Generating a Report Table (KLIF 3.1) – 1 of 3

Årsrapport

Til
Klima- og forurensningsdirektoratet

2012

Varg

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 beregeninger av korrelasjonsfaktor for drenasjevann.

Tabell 3-1 Utslipp av olje og oljeholdig vann

Vanntype	Totalt vannvolum (m3)	Midlere oljeinnhold (mg/l)	Midlere oljevedheng på sand (g/kg)	Olje til sjø (tonn)	Injisert vann (m3)	Vann til sjø (m3)	Eksportert prod. vann (m3)	Importert prod. vann (m3)
Produsert	996 928	6.44		6.42	0	997 582	0	0
Fortregning		0.00						
Drenasje	7 348	6.81		0.06	0	8 708	0	0
Annet		0.00						
	1 004 275			**6.48**	**0**	**1 006 289**	**0**	**0**

EEH Example Query – Generating a Report Table (KLIF 3.1) – 2 of 3

```
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 {
   {
      BIND ("Produsert,Fortregning,Drenasje,Jetting,Annet" AS ?Vanntype_LIST) .
      ?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)) 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 {
            {
               BIND ("Produsert,Fortregning,Drenasje,Jetting,Annet" AS ?Vanntype_LIST) .
               ?Vanntype spif:split ( ?Vanntype_LIST "," ) .
               {
                  {
                     FILTER (?Vanntype IN ("Produsert", "Fortregning", "Drenasje", "Annet")) .
                     ....
```

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
 {
            FILTER fn:contains("Jetting", ?Vanntype) .
            ?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 {
               ?oilyWaterDischarge eeh-ui:oilyWaterDischargeJetting-OilToSeaISOmethod-tonnes ?Olje_til_sj_UNITtonn_TMP_MAYBE .
            } .
            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

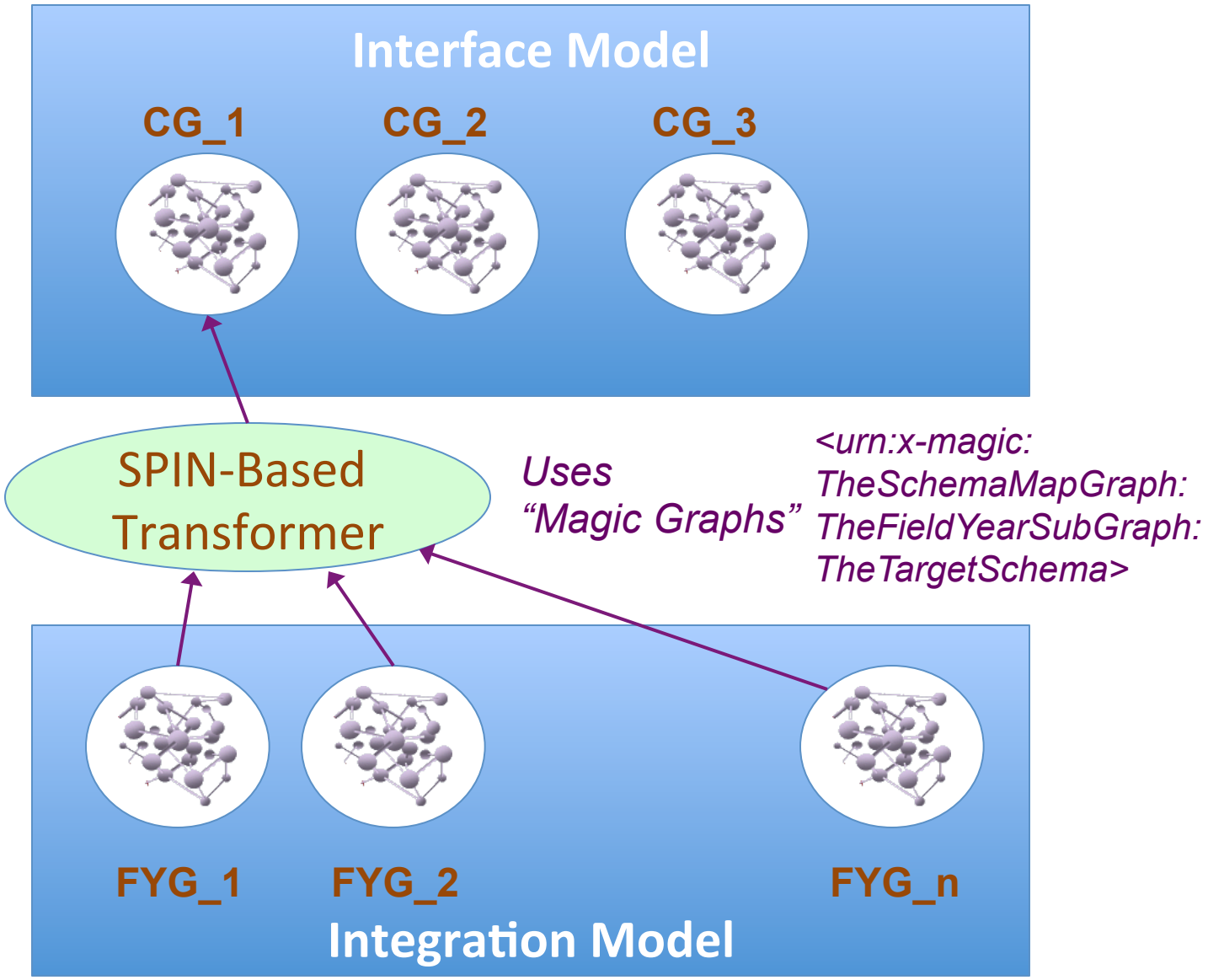
```
SELECT ?Avfallstype ?Beskrivelse ?Avfallsnummer ?Nuklider ?Sendt_til_land_UNITtonn_SUM ?Total_aktivitet_UNITGBq_SUM
WHERE {
   {
      SELECT ?Avfallstype ?Beskrivelse ?Avfallsnummer
          ((eeh-lib:sortStringListRemovingDuplicates(GROUP_CONCAT(?Nuklides; SEPARATOR=','))) AS ?Nuklider)
          ((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 .
         ?hazardousWasteOther eeh-ui:hazardousWasteOther-Category ?AvfallstypeU .
         ?AvfallstypeU rdfs:label ?Avfallstype .
         OPTIONAL {
            ?hazardousWasteOther eeh-ui:hazardousWasteOther-Description ?Beskrivelse_MAYBE .
         } .
         BIND (COALESCE(?Beskrivelse_MAYBE, "") AS ?Beskrivelse) .
         ?hazardousWasteOther eeh-ui:hazardousWasteOther-WasteCode ?Avfallsnummer .
         FILTER fn:starts-with(?Avfallsnummer, "3") .
         OPTIONAL {
            ?hazardousWasteOther eeh-ui:hazardousWasteOther-NuclideName ?Nuklides_MAYBE .
         } .
         BIND (COALESCE(?Nuklides_MAYBE, "NULL") AS ?Nuklides) .
         BIND (eeh-lib:sumStructureTypeValuesYearDataWithOneLiteralConstraint(eeh-ui:hazardousWasteOther-WasteCode,
                  ?Avfallsnummer, eeh-ui:hazardousWasteOther-SentOnShore-tonnes, eeh-ui:HazardousWasteOther, ?yearData) AS
                 ?Sendt_til_land_UNITtonn_SUM_TMP1) .
         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) .
   BIND (IF((datatype(?Sendt_til_land_UNITtonn_SUM_TMP2) != xsd:double), "", ?Sendt_til_land_UNITtonn_SUM_TMP2) AS
         ?Sendt_til_land_UNITtonn_SUM) .
```

Ontology Architecture for Composite Graphs

Simple View Ontologies for UI, reporting and analytics

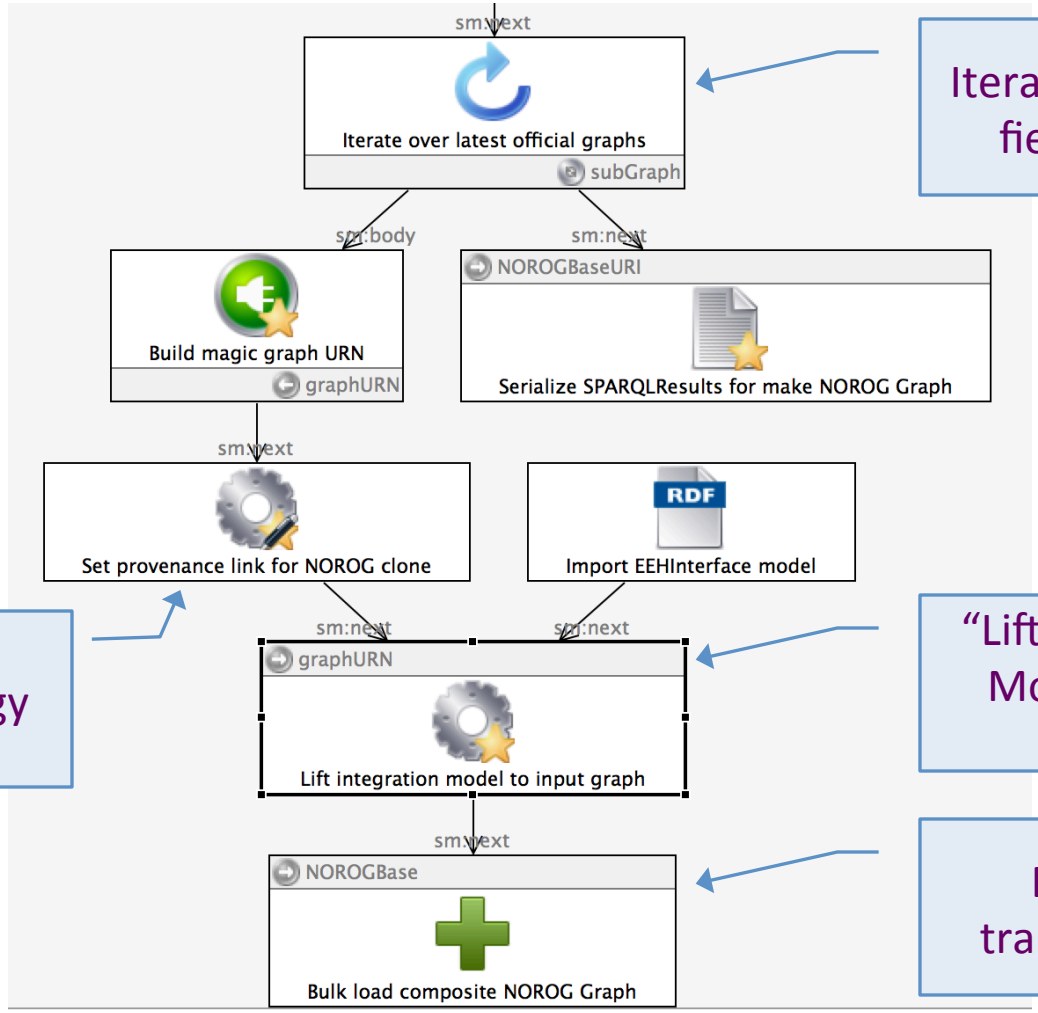
SPARQL Query-View-Transformation

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

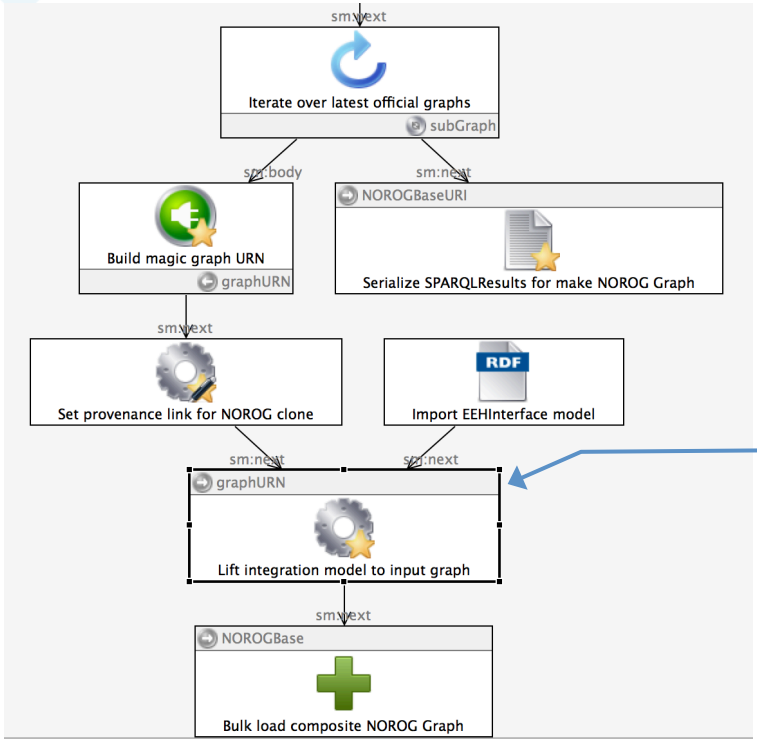
Use of W3C Provenance Ontology for traceability

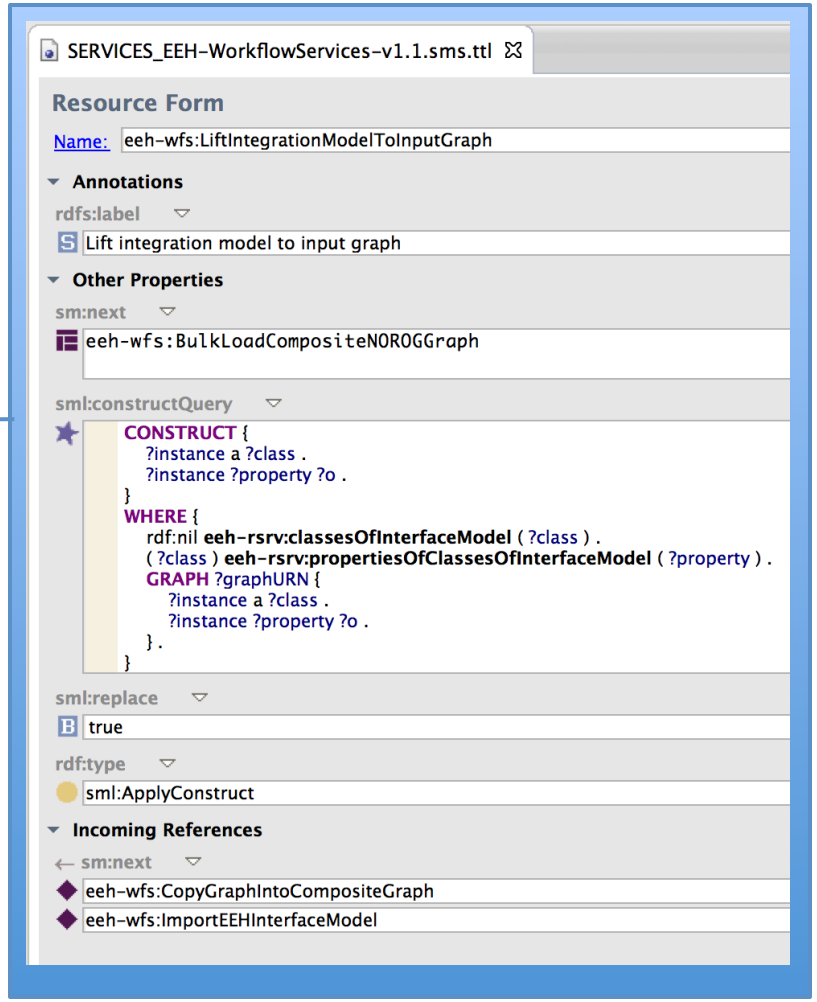
"Lift" the Integration Model using SPIN transforms

Bulk Load the transformed triples

Transforming is controlled by a SPARQLMotion Web Service (2 of 2)

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



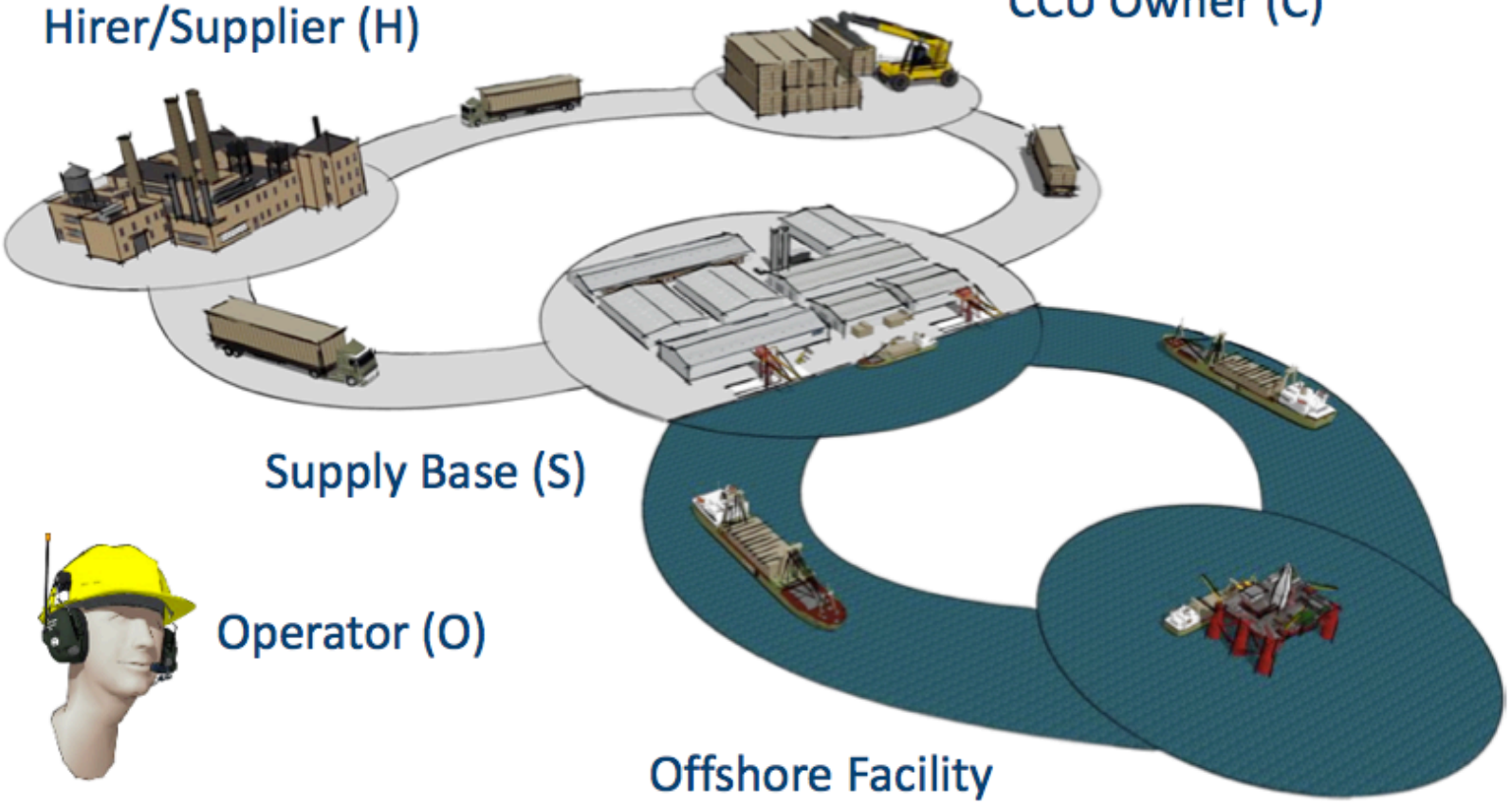


- Total triples in the Integration Model (ISO-15926 Ontologies): ~ 4,000,000
- Total triples in the Interface Model: ~400,000

EPIM LogisticsHub (ELH)

Offshore Supply Chain: CCU Tracking

computas



TopQuadrant™

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

LDBC

Thank you!

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