How semantic technology can help you do more with production data

Doing more with production data

EPIM and Digital Energy Journal

2013-04-18

David Price, TopQuadrant
London, UK
dprice at topquadrant dot com

Agenda

- Quick introduction to Semantic Technology
- Production data needs and How Semantic Technology helps

The Web: The World's Largest Information System

What does the Web do?

Doing more with production data

Thursday, April 18, 2013
Stavanger
Norwegian Petroleum Museum

Organised by Digital En

Expedia.co.uk
People shaped travel

Welcome - Already a member? [S

Home | Hotels | Flights | Flight + Hotel | Car Hire | Beach Holidays | Deals | Last Mi

TRAVEL INFORMATION Winter weather travel advice

CREATE YOUR TRIP

Flight
Hotel
Car
Activities

Flight + Hotel
Flight + Car
Flight + Hotel + Car
Eurostar + Hotel

BOOK A FLIGHT + HOTEL TOGETHER
SAVE 16%*
*Average savings 2011

Flight + Hotel

Leaving from: London, England, United Kingdom (LON-All Ai
Departing: 17/04/2013
Time: Anytime

Going to: Stavanger (and vicinity), Norway
Returning: 18/04/2013
Time: Anytime

1 night trip

Direct flights only

I only need a hotel for part of my trip

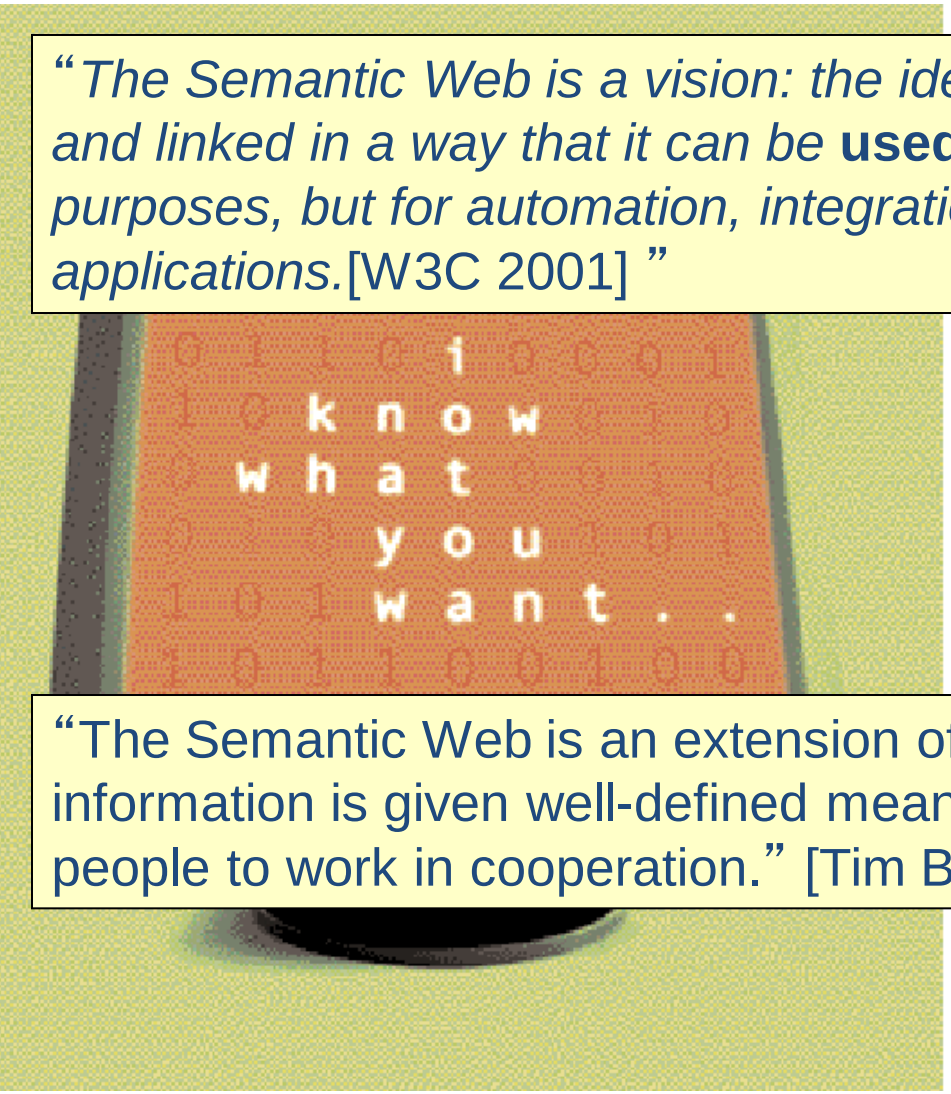
Rooms:	Adults (18-64)	Seniors (65+)	Children (0-17)
1	Room 1: 2	0	0

SEARCH FOR FLIGHT + HOTEL

Web page interaction today – people are the medium!

Semantic Web: Make Web content machine-readable!

"*The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be* **used by machines** *not just for display purposes, but for automation, integration and reuse of data across various applications.*[W3C 2001] "



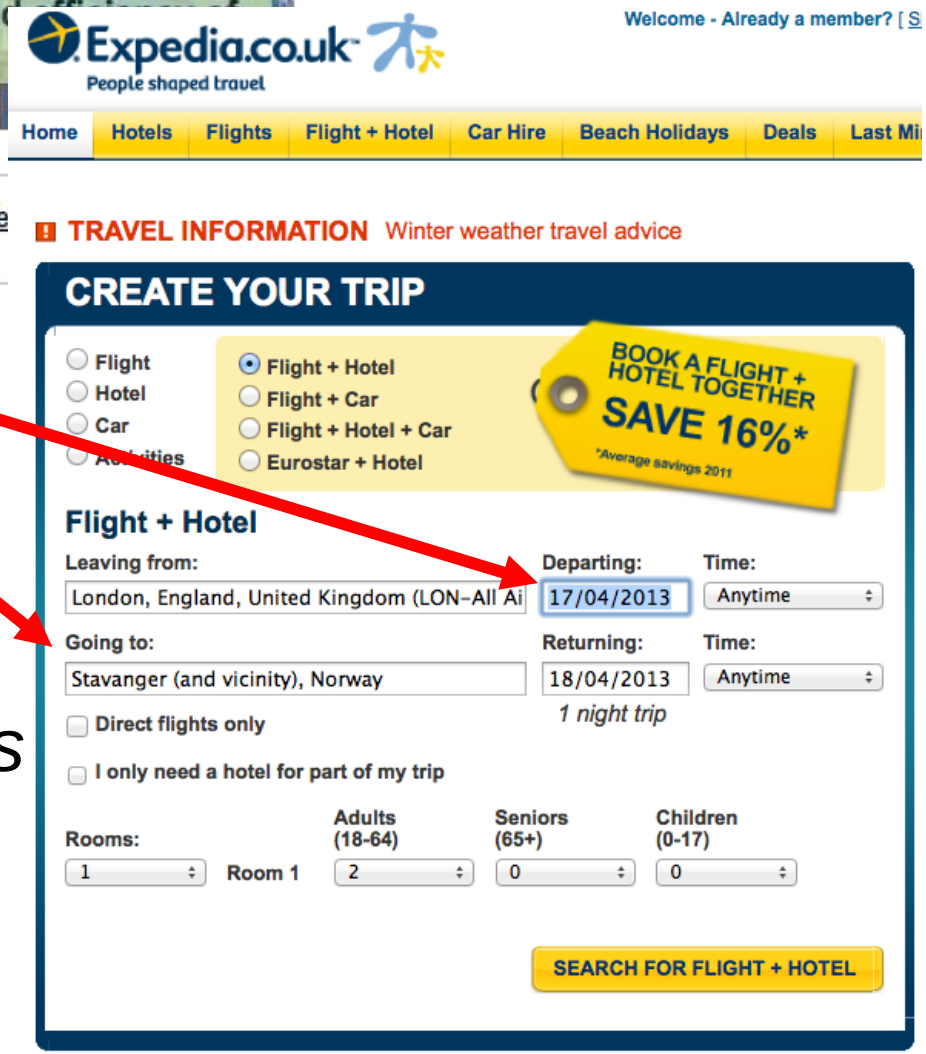
"The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." [Tim Berners-Lee et al 2001]

What could a Semantic Web do?

Doing more with production data

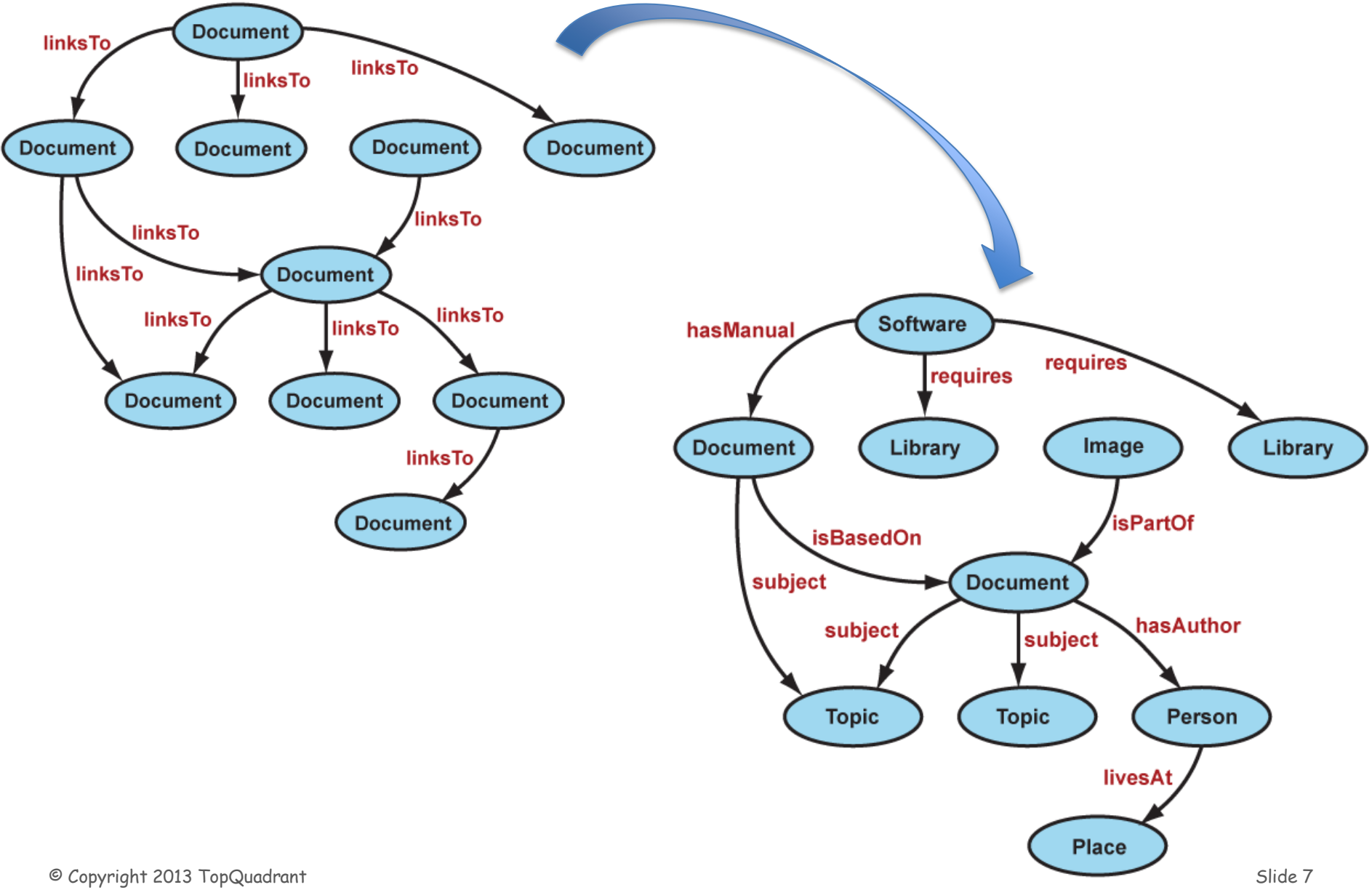
Thursday, April 18, 2013
Stavanger
Norwegian Petroleum Museum

Organised by Digital Ene



Add enabler for communication between apps that can be built-in by the Webmaster

From *The Web* to a *Semantic Web*



Features of The Web

- Anyone can say Anything about Any topic (AAA)
- Names are global so that anyone can refer to them
- Two people might have different names for the same thing . . . (non-unique naming)
  - . . . Or the same name for different things!
- You never know everything on the Web ("Open World")

This isn't what we **want** the Web to be, it is how the Web **is** (and how it supports the network effect that makes the Web so valuable)

The Web rides the Internet

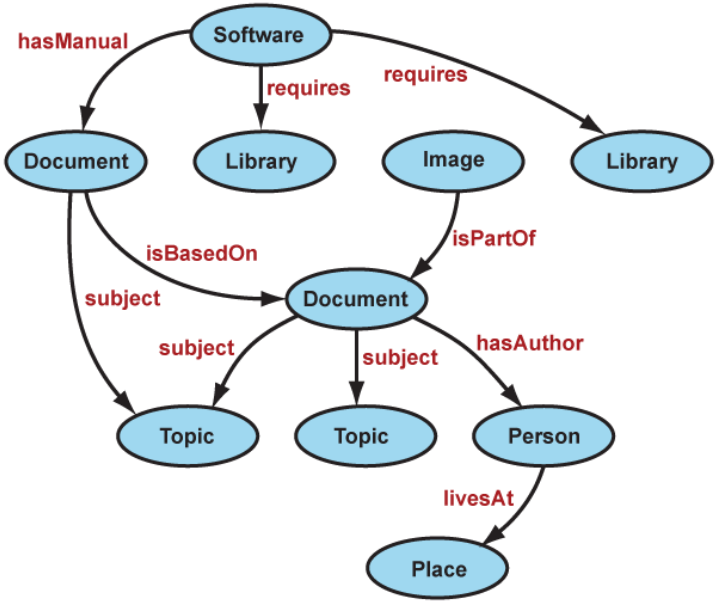
- Internet technology includes
  - Standards for identifying things globally : Web addresses (aka Uniform Resource Identifiers or URIs)
  - Protocols for accessing the identified things: Hypertext Transfer Protocol (HTTP)
- Businesses and users have a lot of experience using, managing, securing and scaling Web sites and applications that use Internet technologies
- And all this works just as well on internal, secure networks as on the Web/Internet

A plan comes together

- Build on existing Web infrastructure – after all it is the Semantic **Web**
- Be flexible and extensible
  - Enable easily reuse of and addition to what's known about a topic without forcing translation or duplication
  - Enable naming issues to be addressed
- Do not stop simple things being done simply, yet enable automation and complexity where useful
  - Draw on *computer science meets philosophy* research – the 'O word' comes into use (i.e. ontology)
- Develop standards to support the Semantic Web vision

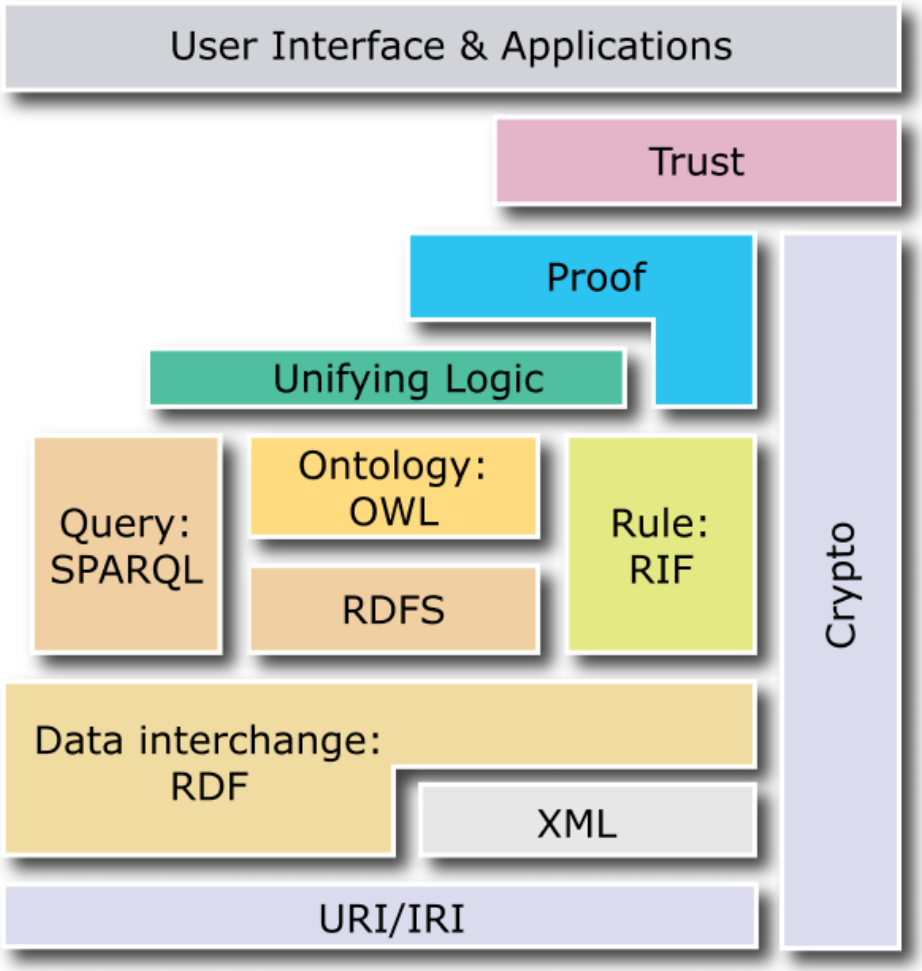
Technical Solution

- Web data is not tables or a hierarchy … it is a network
  - Same is true of data about any even mildly complex topic, such as Oil and Gas Production
- Obvious solution : manage the data as the graphs that they naturally are
  - Remembering that names are global and ride on Internet technology
- And add semantics over that



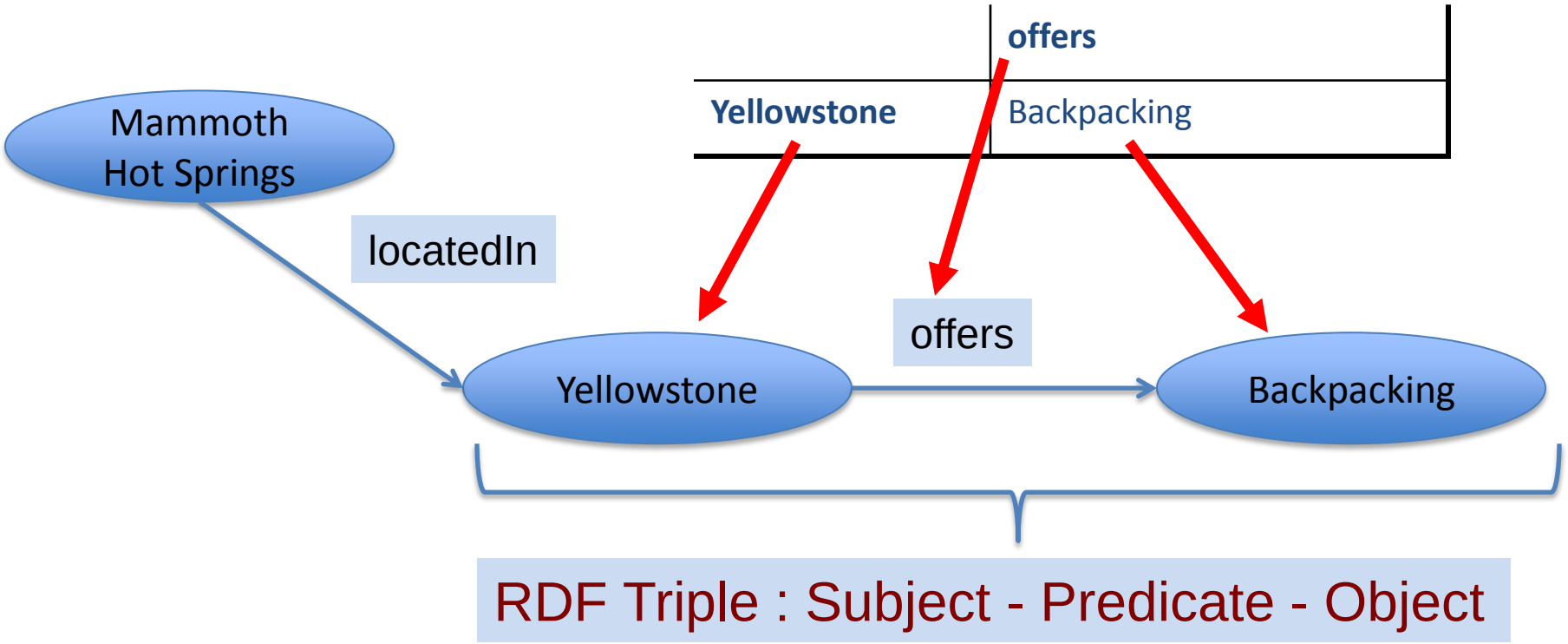
W3C Standards Stack

- *RDF* lets data be brought together (as graphs)
- *RDF Schema* enables simple data modeling
- *OWL* enables complex data modeling and logical inferences
- *SPARQL* queries over any RDF



Resource Description Framework

- RDF : basic infrastructure, a directed graph language
  - Resource means *thing identified by a Web address*
  - E.g http://www.example.org/places/offers
- The node-edge-node pattern is called an *RDF Triple*, which equates to a cell in a spreadsheet :



RDFS and OWL

RDF Schema is the schema language for RDF defines:

- things can be members of classes (individual/instance)
- class hierarchies (e.g. **Company** subClassOf **Organisation**)
- simple property hierarchies (e.g **nickname** subPropertyOf **name**)

Web Ontology Language (OWL) adds

- logic-based classes (e.g. **A** is unionOf **B,C**)
- restrictions on class-property relationships (all **Company** instances shall be **incorporatedBy** *Companies- House*)

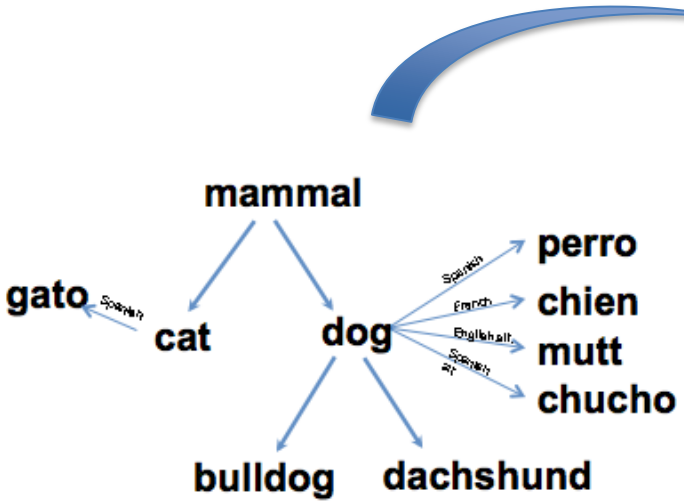
... for Owl, wise though he was in many ways, able to read and write and spell his own name ...

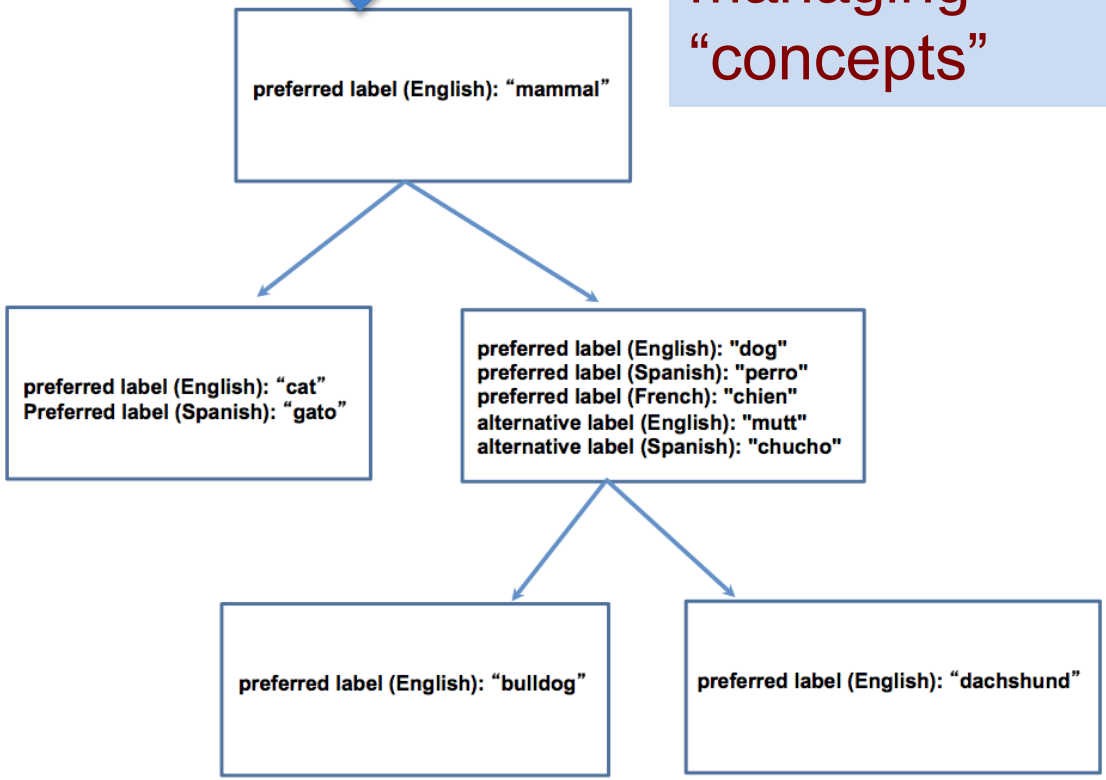
No new syntax, RDFS is specified using RDF so the schema is just more data (same for OWL)

- the difference is in the inferences

Simple Knowledge Organization System

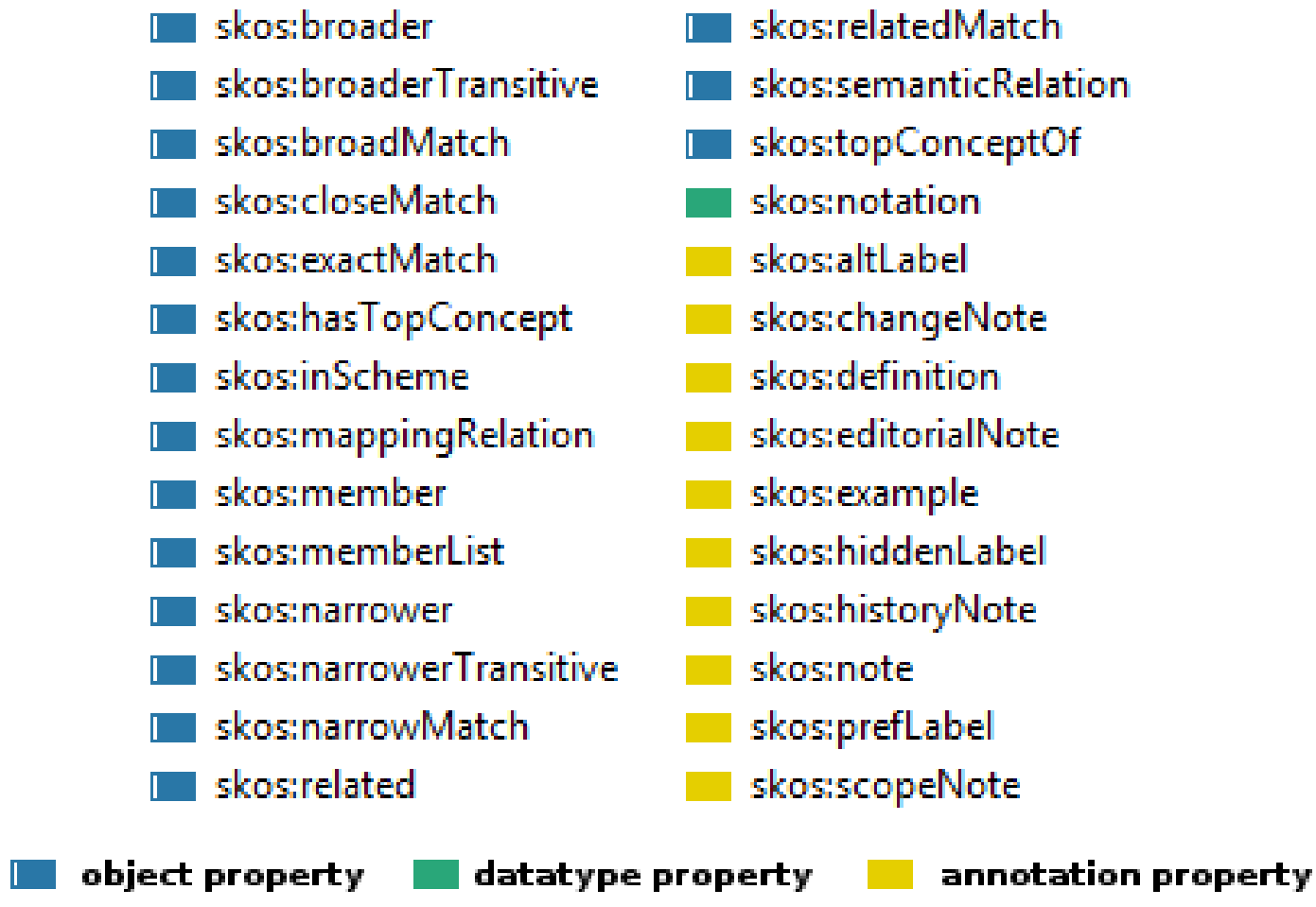
- SKOS: the W3C RDF/OWL standard for thesauruses, taxonomies, and controlled vocabularies





Not managing "terms", instead managing "concepts"

SKOS Properties



Key Characteristics of Results

- Flexibility
  - Designed assuming a bottom-up approach so mixing schemas and data from many different sources is simple
  - Adding new concepts is simple and low cost
- Distribution
  - Anything can be anywhere reachable by Internet protocols
    - Public servers, government servers, secure in-house servers, files on the Web, files on my laptop
  - URI basis in RDF means you can point to anything with global name scope
- Standardization at W3C
  - World Wide Web Consortium
    - This is where HTML and XML were standardized
  - No vendor lock-in
    - unlike comparable approaches like relational databases

Agenda

- Quick introduction to Semantic Technology
- Production data needs and How Semantic Technology helps

Example Production Data Needs

- Find existing data
  - Where is the analysis of Wellbore 7/4-3 performed last week?
- Relate existing data
  - Data relating Morvin field and Åsgard B platform that exists in different IT systems (and Åsgard B is called “ASB” in one and “Åsg-B” in the other)
- Exchange data
  - I need to extract the 2009 Kristin volumes, pressures and temperatures and convert to spreadsheet to load into my reporting application
- Integrate data
  - I’d like to know last months production volume total for all fields in which GDF Suez E&P Norge AS is a licensee.
- Analyze data
  - Over the past 12 weeks, what’s the trend in barrels of oil per day for Kristin field?

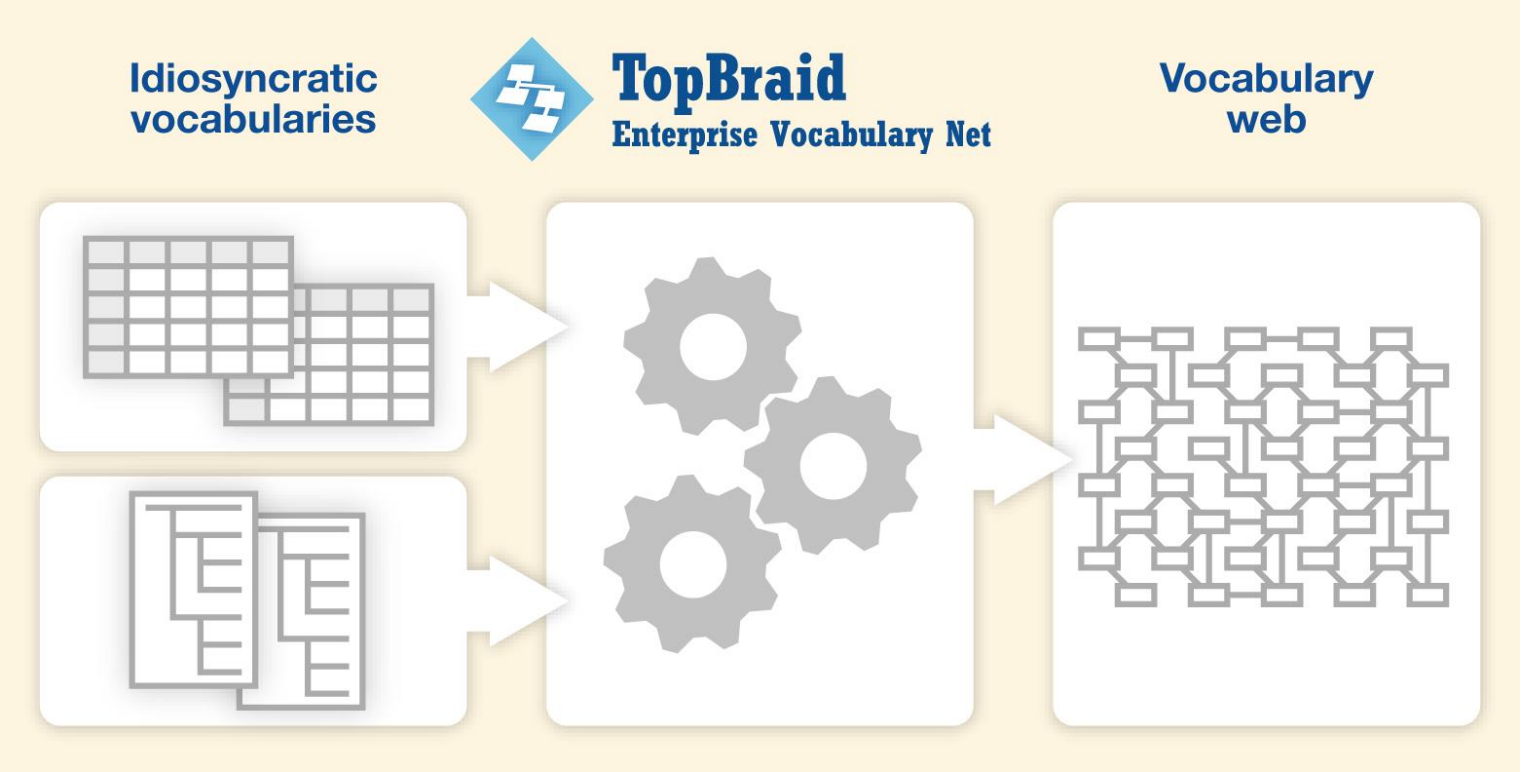
Matching Needs to Technology

Need	Technology Example
Find existing data	• Vocabulary-enhanced search (RDF, OWL, SKOS, SPARQL) • Logical data warehouse (R2RML RDB to RDF)
Relate existing data	• Linksets in Logical data warehouse (RDF,OWL)
Exchange data	• Triple-ize any data format (e.g XSD to OWL) • Query or graph for subset (RDF Graph, SPARQL) • SPARQL/SPIN transform (SPARQL Construct) • Export in multiple formats (XML, text, JSON, etc)
Integrate data	• Semantic repository (OWL, RDF database)
Analyze data	• Query over temporal data (OWL, RDF database, ISO 15926)

Vocabulary-enhanced search

1. Create an industry, corporate or project vocabulary to to enhance search;
2. Tag content with those terms, including auto-tagging using text extraction tools; AND/OR
3. Integrate vocabulary with search tool or content management system

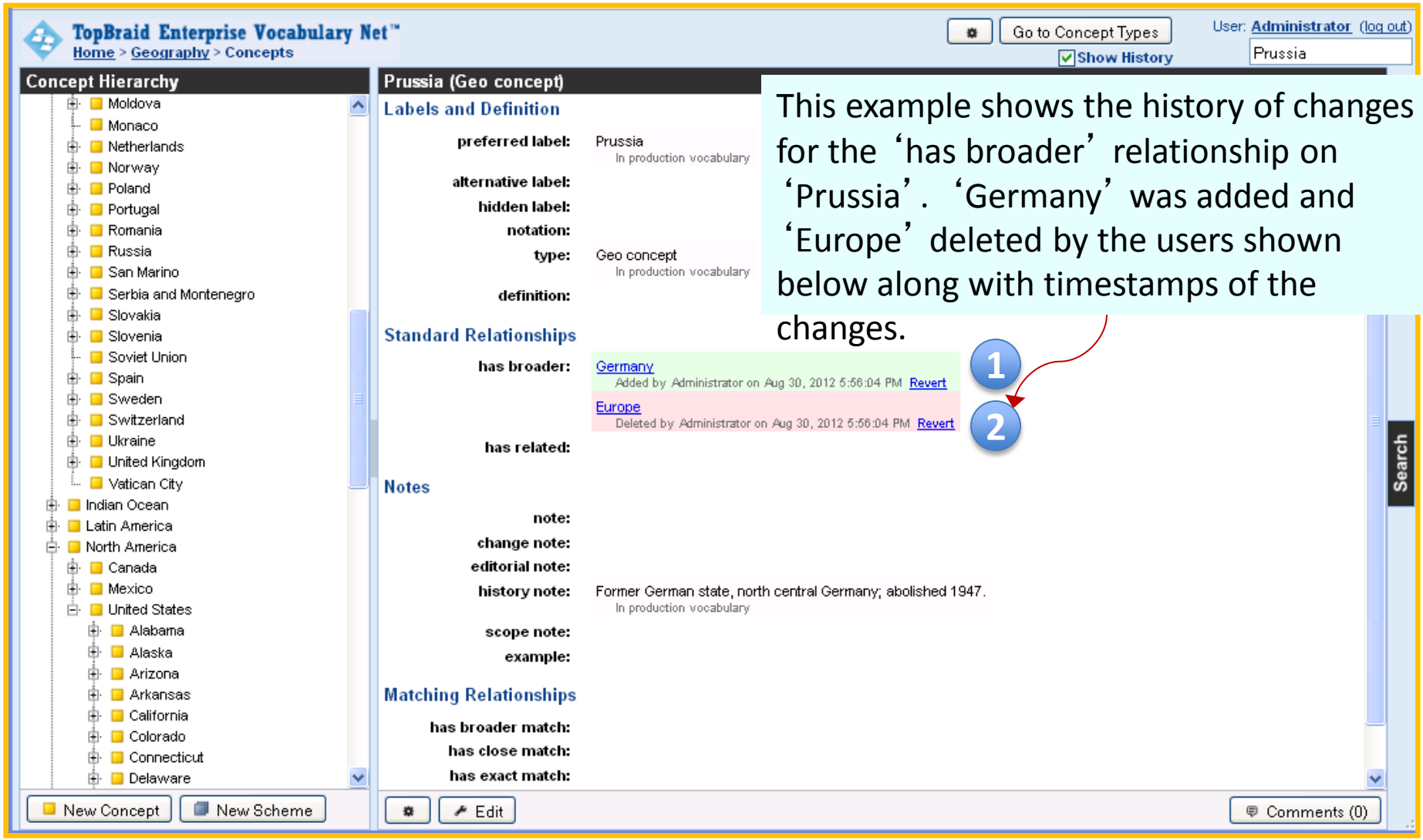
Step 1. TopBraid Enterprise Vocabulary Net



How it Works

- Constructs a Dynamic Web of Terminology
- Creates links between terminology elements that were unconnectable (using SKOS)

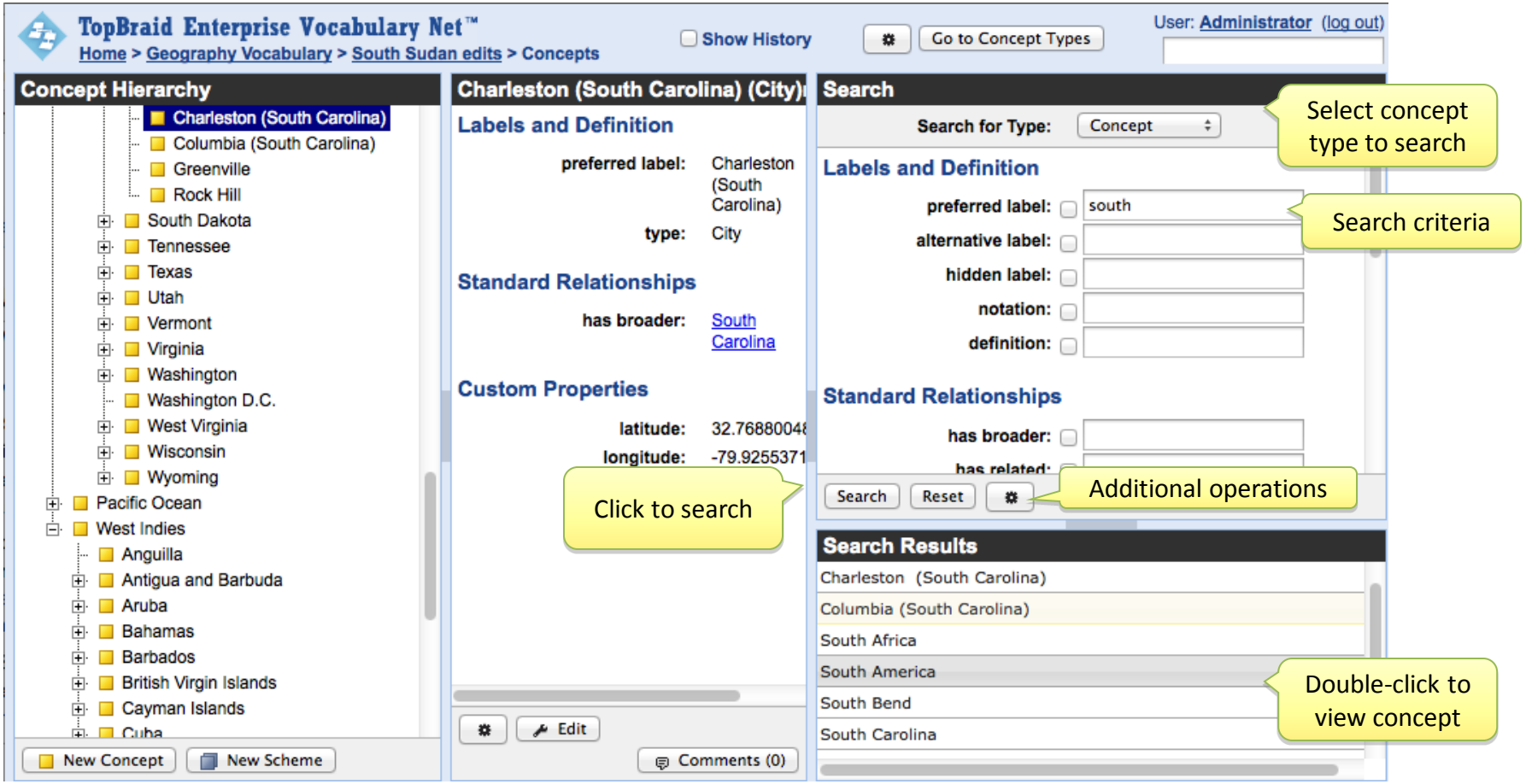
Being based in RDF, EVN provides granular history and audit trail of every change



This example shows the history of changes for the 'has broader' relationship on 'Prussia'. 'Germany' was added and 'Europe' deleted by the users shown below along with timestamps of the changes.

Forms on the Search panel enable concepts of interest to be found based on their property values

- Expand Search box



Step 2 : EVN Tagger Overview

- EVN Tagger: Manual tagging of content with SKOS vocabularies
  - EVN Tagger is an application that links “content” to SKOS vocabulary
  - Content is a set of resources in any RDF graph. Administrator identifies which graphs are “content graphs"
  - Content graph can be a virtual view into external sources, such as SharePoint files, Web sites, etc.
  - Change management is applied to tags

Step 3: EVN Search Enrichment Server

- Search Enrichment Server provides APIs for accessing vocabulary content by external systems
  - examples include:
    - AllBroaderConcepts: Gets all concepts that are broader than the provided ?narrowerConcept, including the broader values of broader values.
    - AltLabels: Gets the alternative labels of a SKOS concept. If a language tag is specified, only the labels of the language tag are returned. Otherwise, all labels are returned.
    - SynonymsOfConcept: For the purpose of this template, synonym is a resource which label matches any of the labels of the given ?concept.
  - Pre-build APIs are designed to support the requirements for search enhancement capabilities
  - Custom APIs can be added using tools in the TopBraid platform

Logical data warehouse

1. Wrap existing data sources in place, but triple-ize it
   - for relational databases the W3C RDB to RDF Mapping Language standard can be applied
2. Defined a “master model” or schema through which you’ll query the warehouse data
3. Define relationship between model of data sources and master model
4. Create “linksets” : links between unrelatable items in any data sources

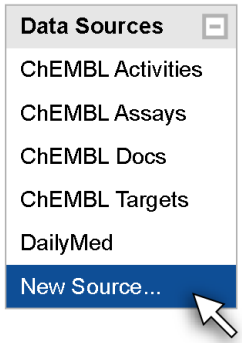
According to Gartner* *“...the Logical Data Warehouse (LDW) is a new data management architecture for analytics which combines the strengths of traditional repository warehouses with alternative data management and access strategy.”*

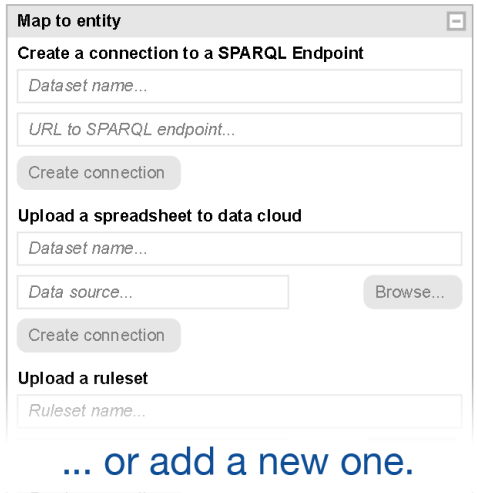
* *Gartner Hype Cycle for Information Infrastructure, 2012 - http://www.gartner.com/id=2101415*

Logical Data Warehouse Example : TopBraid Insight

TopBraid Life Sciences Insight™ Solution
A Logical Data Warehouse for the Life Sciences

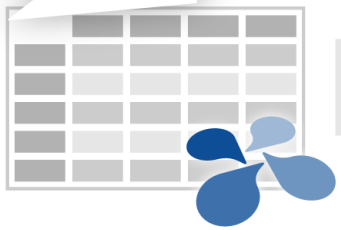
1. Choose an available data source...

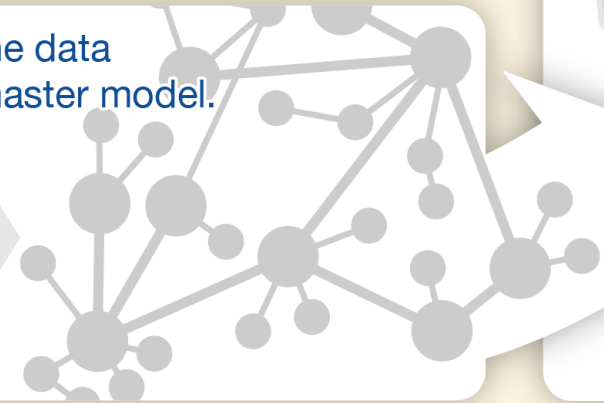




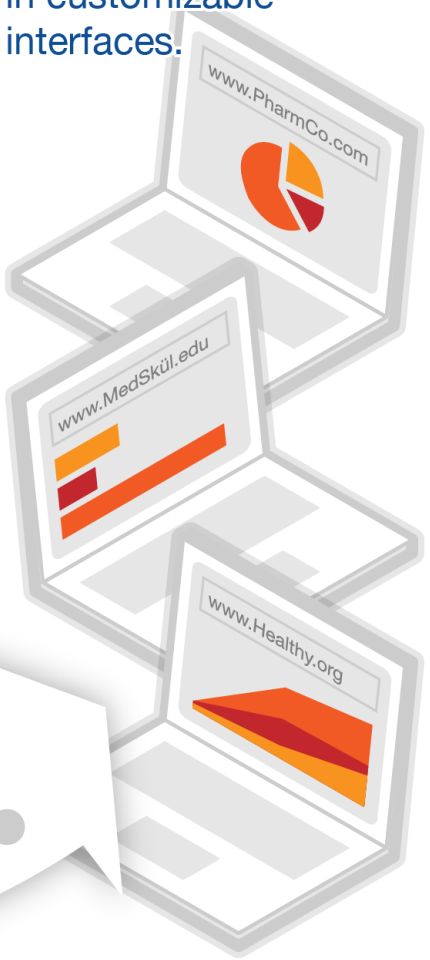
... or add a new one.

2. Map the data to your master model.





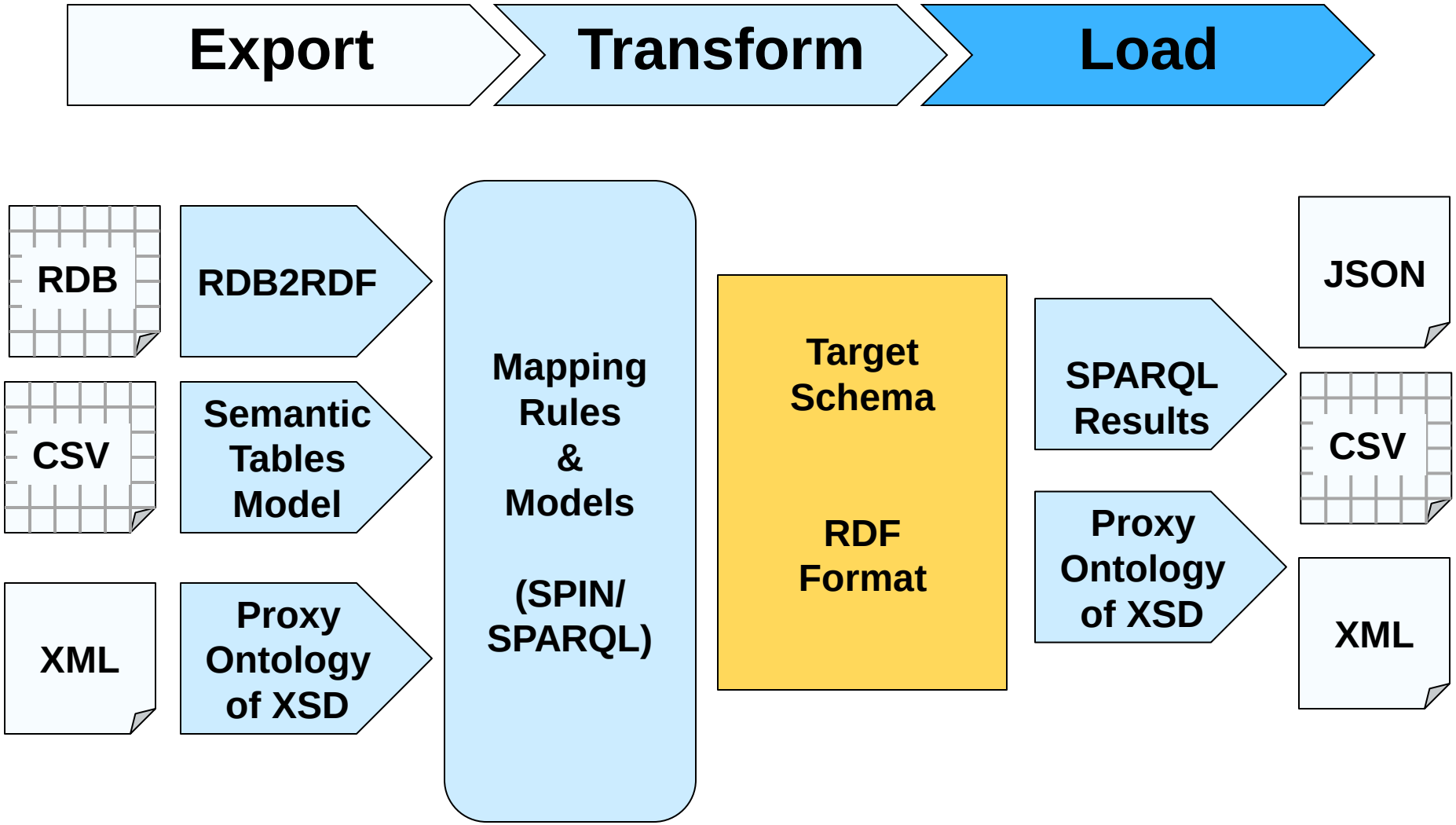
3. Interact with the federated data in customizable interfaces.



Exchange data

1. Provide access to source system data as triples
2. Provide target schema as RDF/OWL
3. Define transformation to neutral inter-change format OR to final target system data format
4. If using neutral inter-change format, define transformation to final target system data format

Semantic Data Exchange



Data transformed to triples using SPIN

- Implementing converters = writing SPARQL or using SPINMap, not Java development
- Same approach regardless of source being XML or CSV

spin:rule (incomplete)

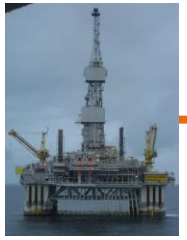
```
# STEP 101 Create DailyDrillingReport, DailyDrillingActivity, TemporalPartOfAWellBore and WellBore
# Minimum data - createDate, dTimStart, nameWellBore
CONSTRUCT {
   ?dailyDrillingReport a ep-report:DailyDrillingReport .
   ?dailyDrillingActivity a ep-activity:DailyDrillingActivity .
   ?dailyDrillingReport ep-activity:reportOn ?dailyDrillingActivity .
   ?dailyDrillingReport ep-activity:createdAt ?createDate .
   ?dailyDrillingActivity ep-activity:startedAt ?dTimStart .
   ?temporalPartOfAWellBore a ep-well:TemporalPartOfAWellBore .
   ?wellBore a ep-well:WellBore .
   ?temporalPartOfAWellBore ep-core:temporalPartOf ?wellBore .
   ?wellBore ep-core:name ?nameWellBore .
   ?dailyDrillingActivity ep-activity:onWellBore ?temporalPartOfAWellB
}
WHERE {
   ?this ddr:nameWellbore ?nameWellBore .
   ?this ddr:dTimStart ?dTimStart .
   ?this ddr:dTimStart ?dTimEnd .
   ?this ddr:createDate ?createDate .
   BIND (ep-spin-lib:normalizeString(?nameWellBore) AS ?normalized
   BIND (ep-spin-lib:buildDailyDrillingReportURI(?normalizedWellBorel
   BIND (ep-spin-lib:buildDailyDrillingActivityURI(?normalizedWellBore
   BIND (ep-spin-lib:buildTemporalPartOfAWellBoreURI(?normalizedW
 ?temporalPartOfAWellBore) .
   BIND (rhspin:wellBoreByName(?nameWellBore) AS ?wellBore) .
}
# STEP 101a Set name of DailyDrillingReport as label
```

Integrate data and Analyze Data

1. Create a conceptual schema covering all the data sources, this is the repository schema
2. Perform “Data exchange” where repository is the target system
3. Once integrated in this manner, interesting analysis options become available

Integration Example :EPIM ReportingHub

Operators on the NCS



XML DDR

XML MPR

XML DPR

Data Exchange

ERH EPIM ReportingHub

RDF Database

Authorities

XML DDR

XML MPR

NORWEGIAN PETROLEUM DIRECTORATE

PETROLEUM SAFETY AUTHORITY NORWAY

License Partners

ERH EPIM ReportingHub

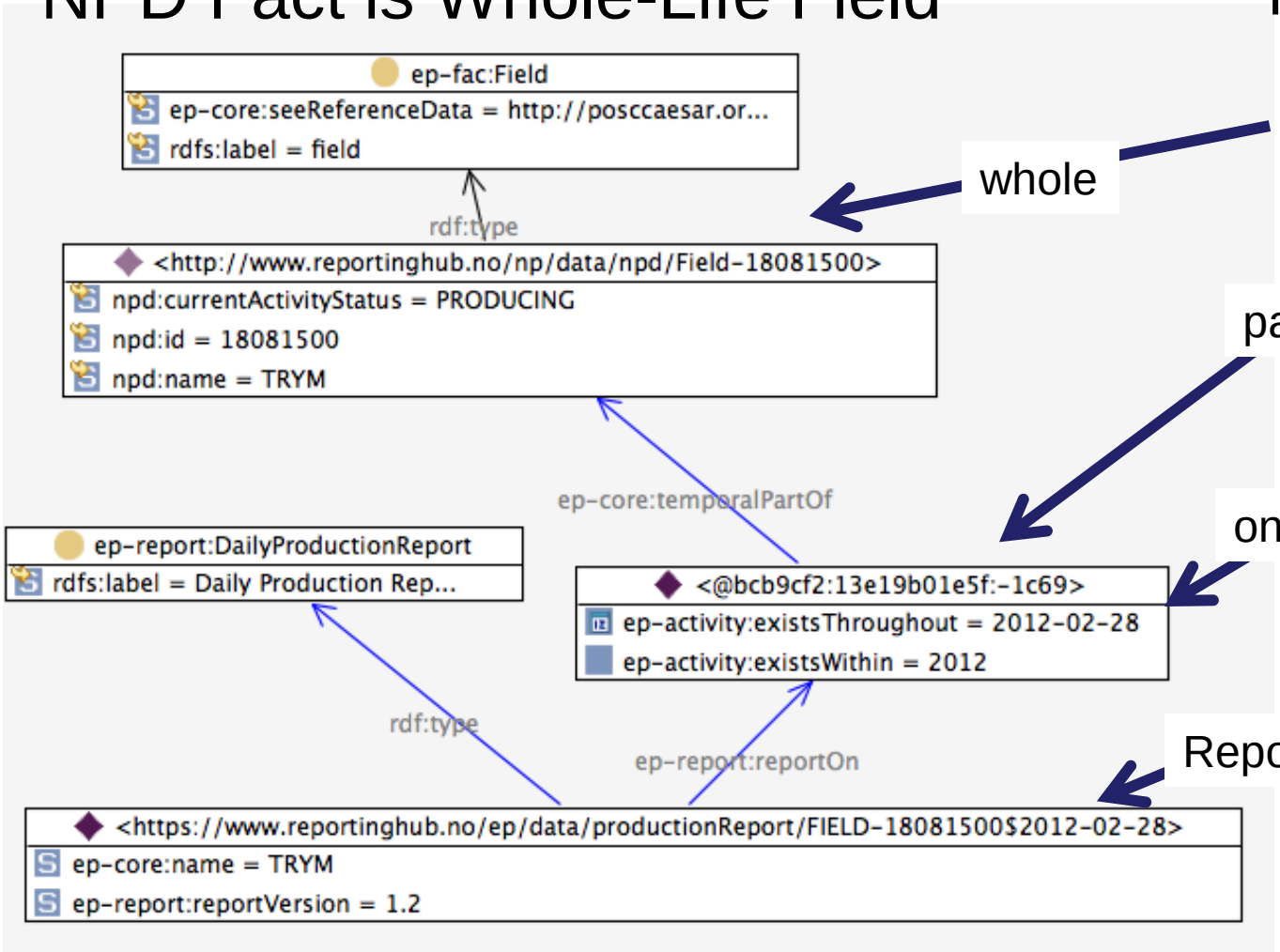
XML <xml /> PDF Document

Semantic Reporting

XML <xml /> PDF Document L2S License2Share

ERH manages temporal data (ISO 15926)

NPD Fact is Whole-Life Field



Field on a day is a *temporalPartOf* whole-life Field

whole

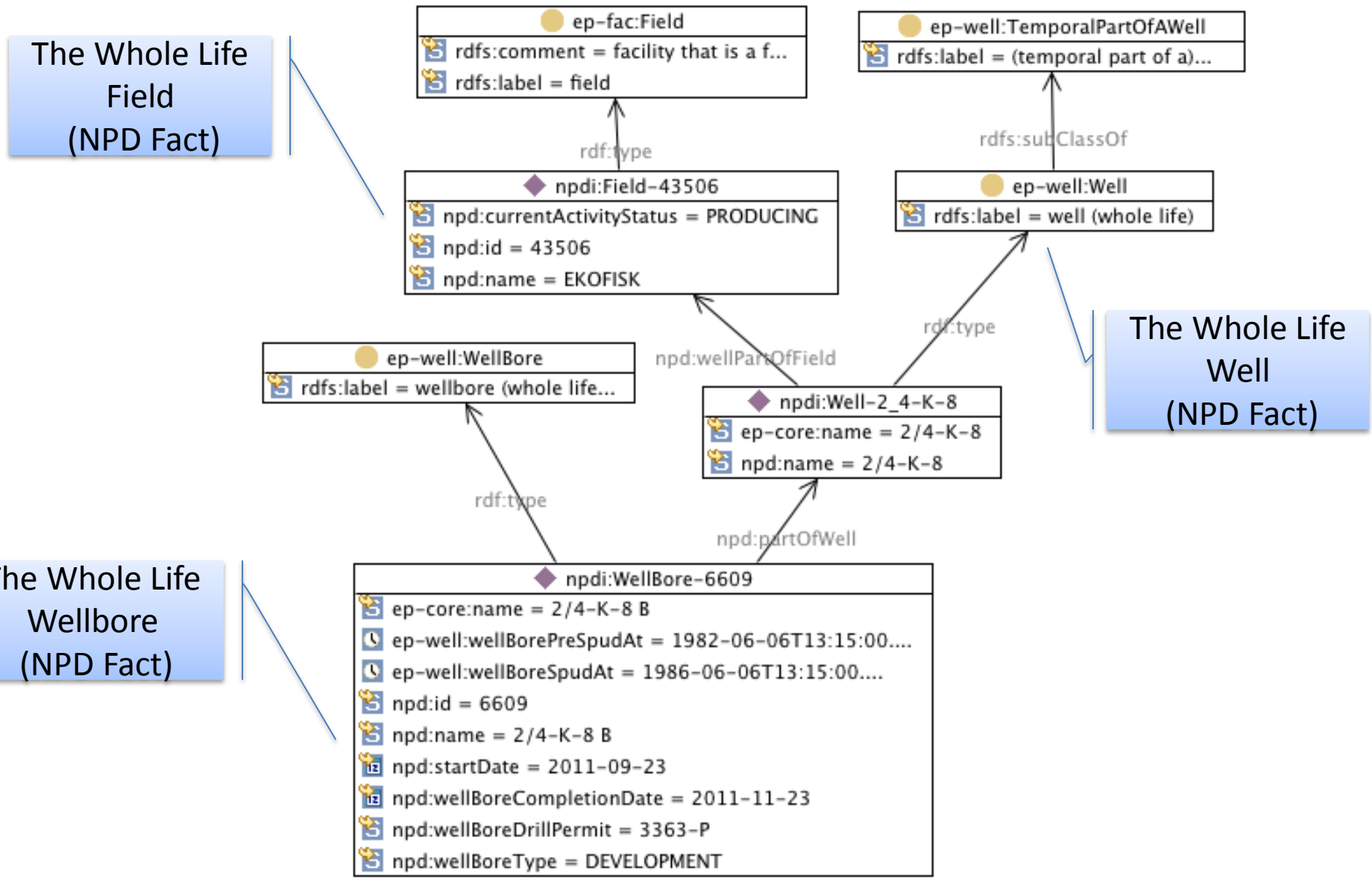
part

on

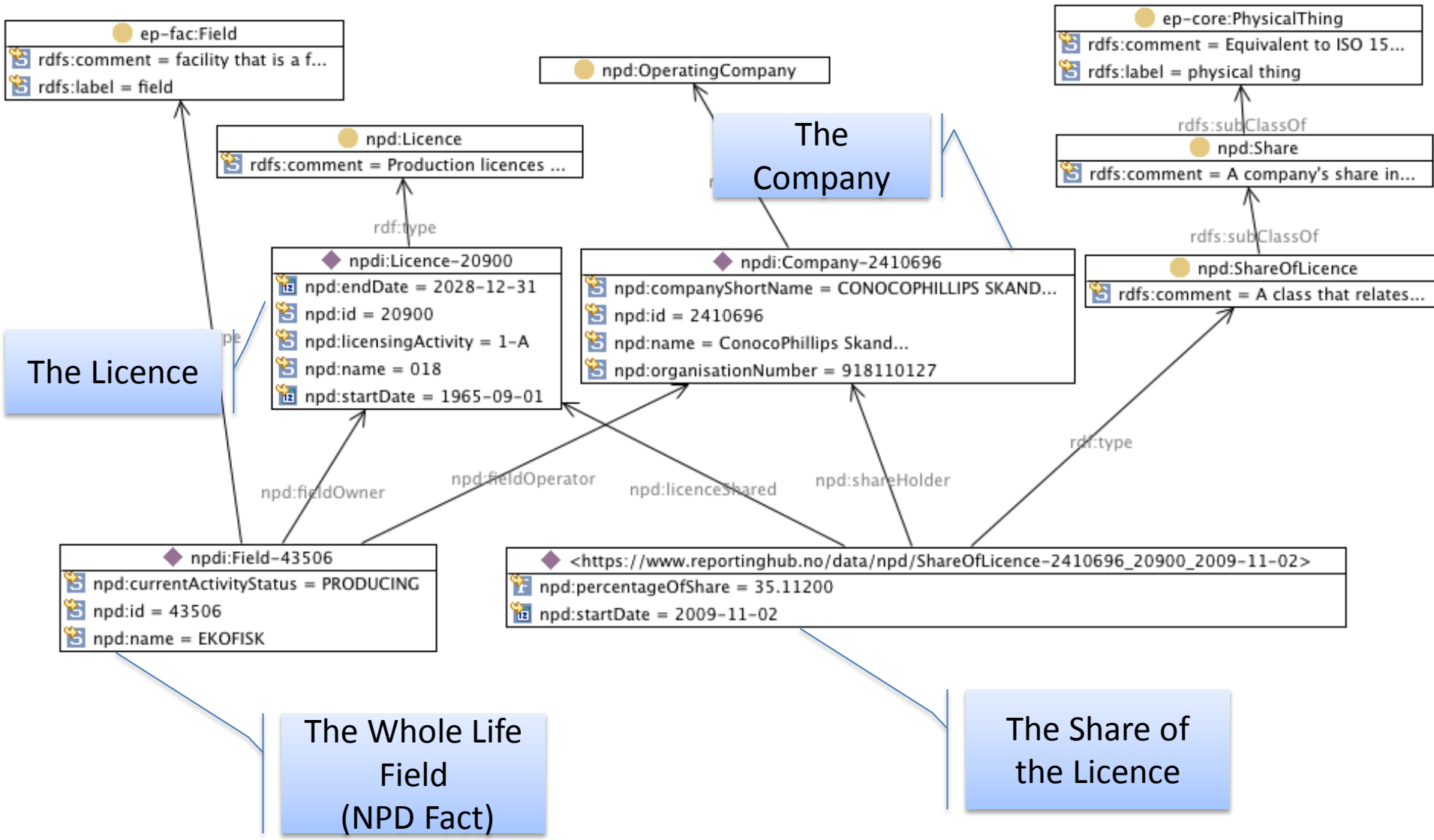
Report

Daily report is about what happens on *TemporalPartOf AField*

Wellbore – part of Well – part of Field



Field owner is License has Share Owners



Conclusions

- Semantic Web technology is a suite of standards and standards-based tools with a spectrum of capability
  - From natural language vocabularies to logic-based applications
- The core principles are:
  - Schema and data are one … schema is just more data, so changes over time are simplified
  - Everything has a globally unique name
  - Everything is accessible using Internet protocols
  - Distributed schemas and data are the norm, not the exception
- Production data is complex and inter-related … a perfect match for this technology