



JavaOne™

java.sun.com/javaone

Semantic Web for the Working Ontologist

Dean Allemang, Chief Scientist, TopQuadrant

TS-5555



Data Challenges in a Distributed World

> In the good old days:

- We had a customer master file and a product master file
- They represented a "Single Version of the Truth"
- Even in large corporations, there was a relatively small number of IT systems and they were all centralized

> Today:

- Many systems: corporate, departmental, personal, internal, external
- Many databases, often hundreds and thousands:
 - reflecting some aspects of the total picture
 - referring to items that are conceptually the same, but not technically (in terms of their identity as understood by computers) the same
- No "Single Version of the Truth"

> How do we function in a multi-faceted, distributed world?

- In particular, how do we develop and deploy systems that can cope and thrive in this reality?

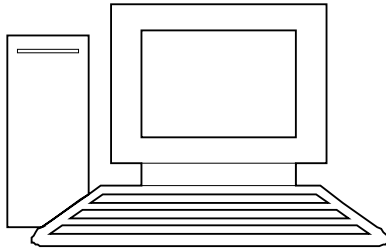


© Warren Photographic

What is RDF? How to distribute data . . .

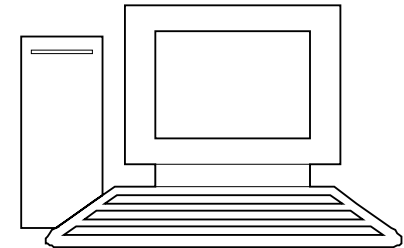
ID	Title	Author	Medium	Year
1	As You Like It	Shakespeare	Play	1599
2	Hamlet	Shakespeare	Play	1604
3	Othello	Shakespeare	Play	1603
4	Sonnet 78	Shakespeare	Poem	1609
5	Astrophil and Stella	Sir Phillip Sidney	Poem	1590
6	Edward II	Christopher Marlowe	Play	1592
7	Hero and Leander	Christopher Marlowe	Poem	1593
8	Greensleeves	Henry VIII Rex	Song	1525

By Rows?

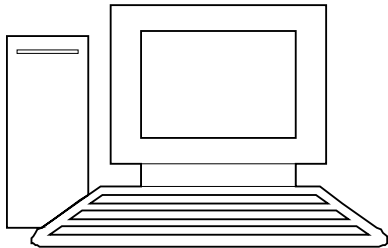


1	As You Like It	Shakespeare	Play	1599
---	----------------	-------------	------	------

Needs common schema - which column is which?

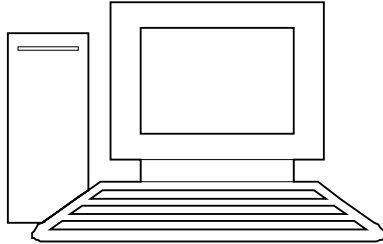


4	Sonnet 78	Shakespeare	Poem	1609
6	Edward II	Christopher Marlowe	Play	1592



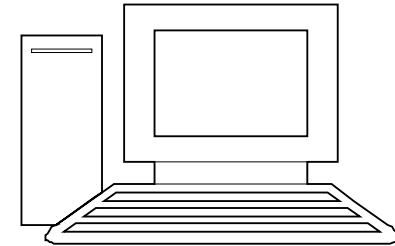
7	Hero and Leander	Christopher Marlowe	Poem	1593
3	Othello	Shakespeare	Play	1603

By columns?



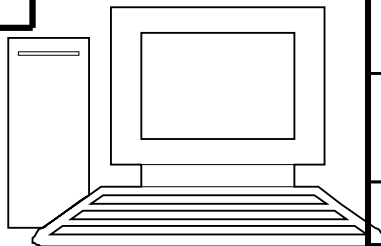
Year	Medium
1599	Play
1604	Play
1603	Play
1609	Poem
1590	Poem
1592	Play
1593	Poem
1525	Song

*Needs to reference entities –
which thing are we talking about?*

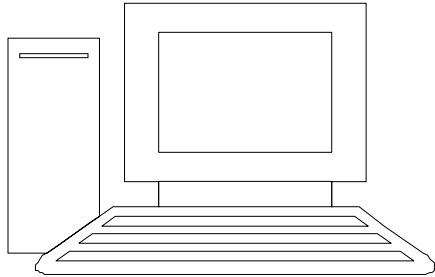


Title
As You Like It
Hamlet
Othello
Sonnet 78
Astrophil and Stella
Edward II
Hero and Leander
Greensleeves

Author
Shakespeare
Shakespeare
Shakespeare
Shakespeare
Sir Phillip Sidney
Christopher Marlowe
Christopher Marlowe
Henry VIII Rex

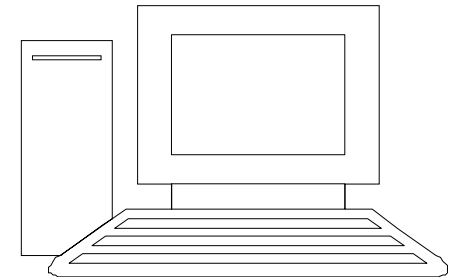


By cells?

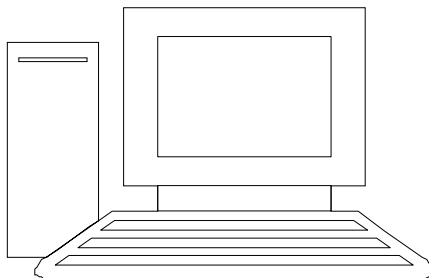


Medium		Title
7	Poem	
2	Hamlet	

Needs to reference both schema and entities



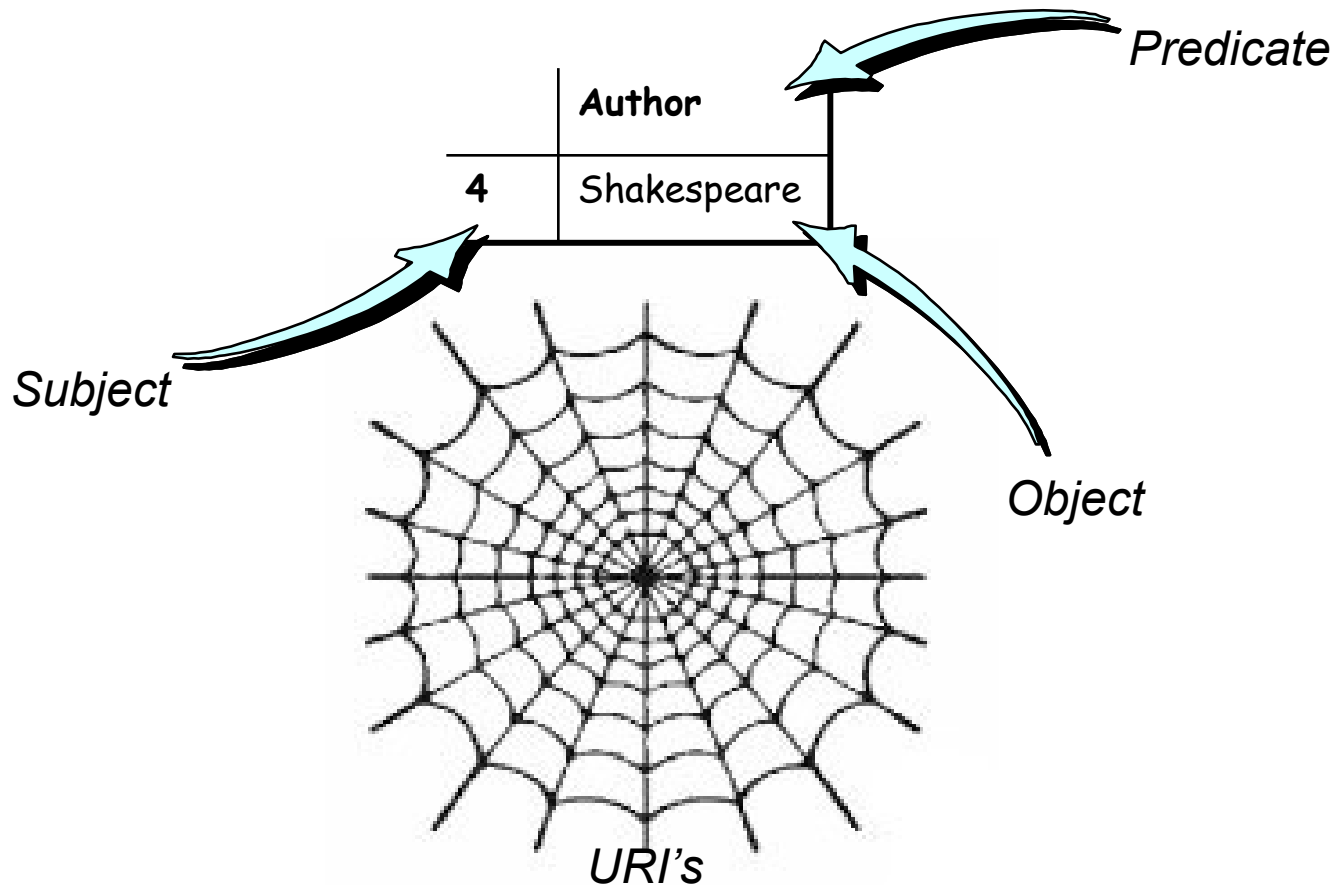
Author	
4	Shakespeare



Year	
2	1604

Medium	
6	Play

A web of Data!

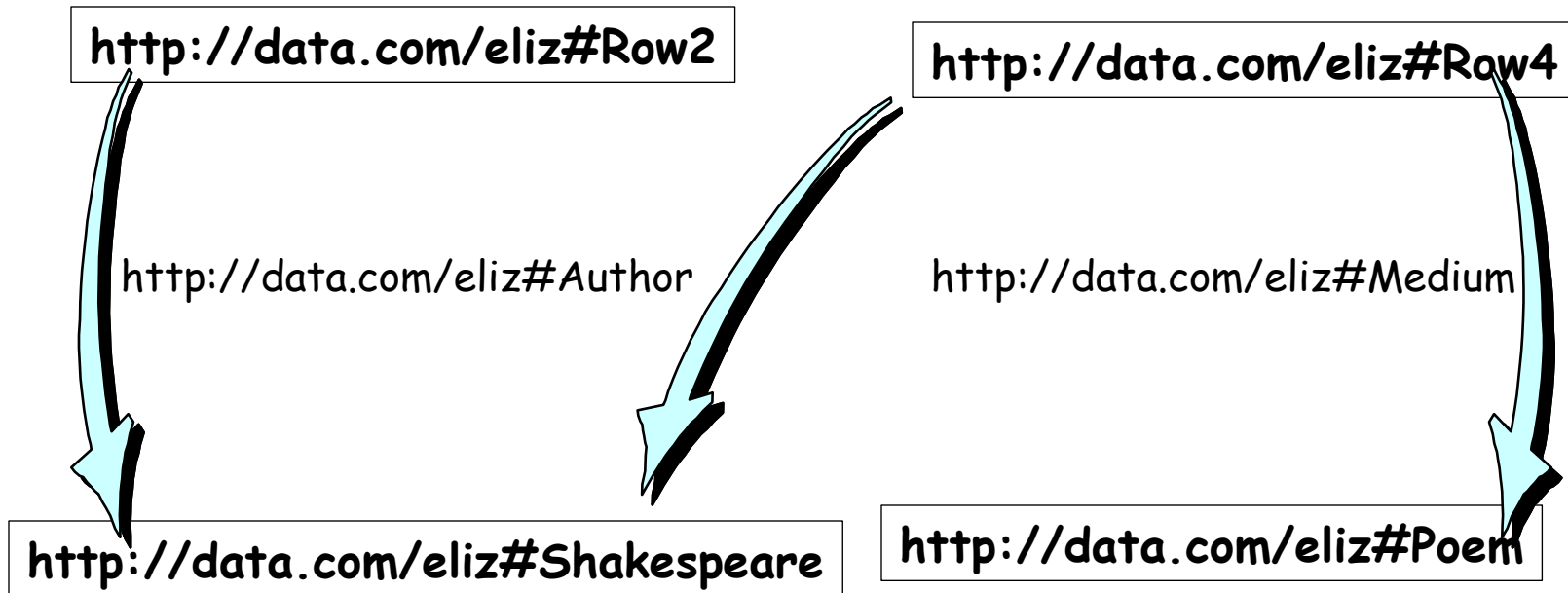


A web of Data!

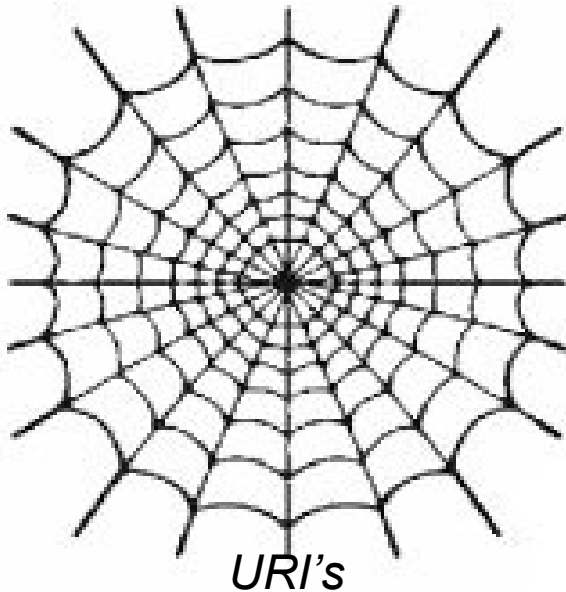
	Author
4	Shakespeare

	Author
2	Shakespeare

	Medium
4	Poem

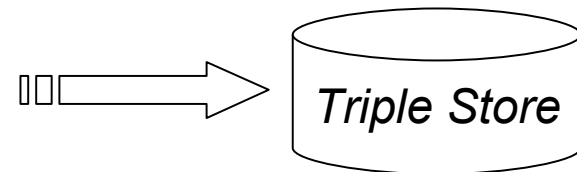


RDF Triple Store – a database for the Semantic Web



Row2	Author	Shakespeare
Row4	Author	Shakespeare
Row4	Medium	Poem

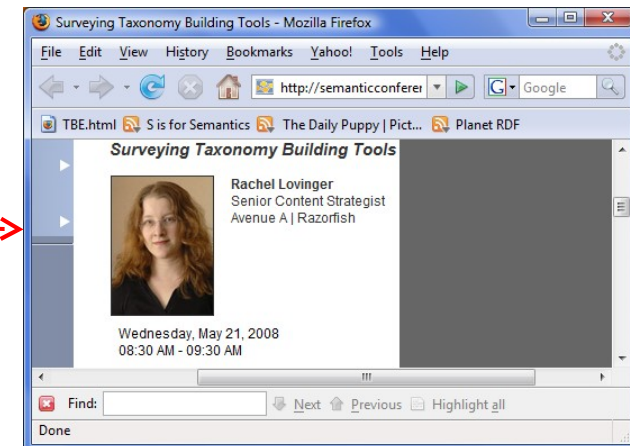
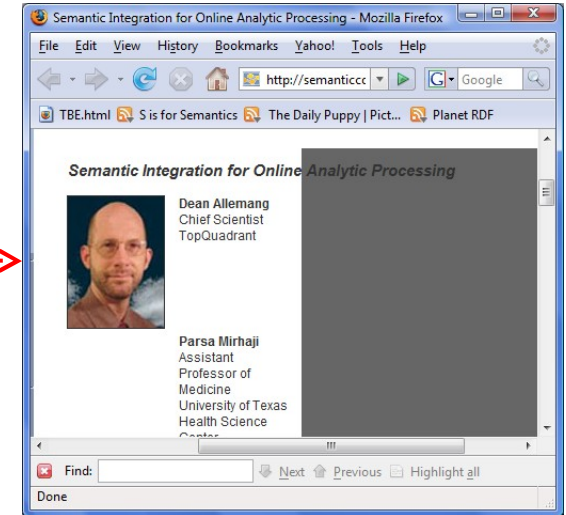
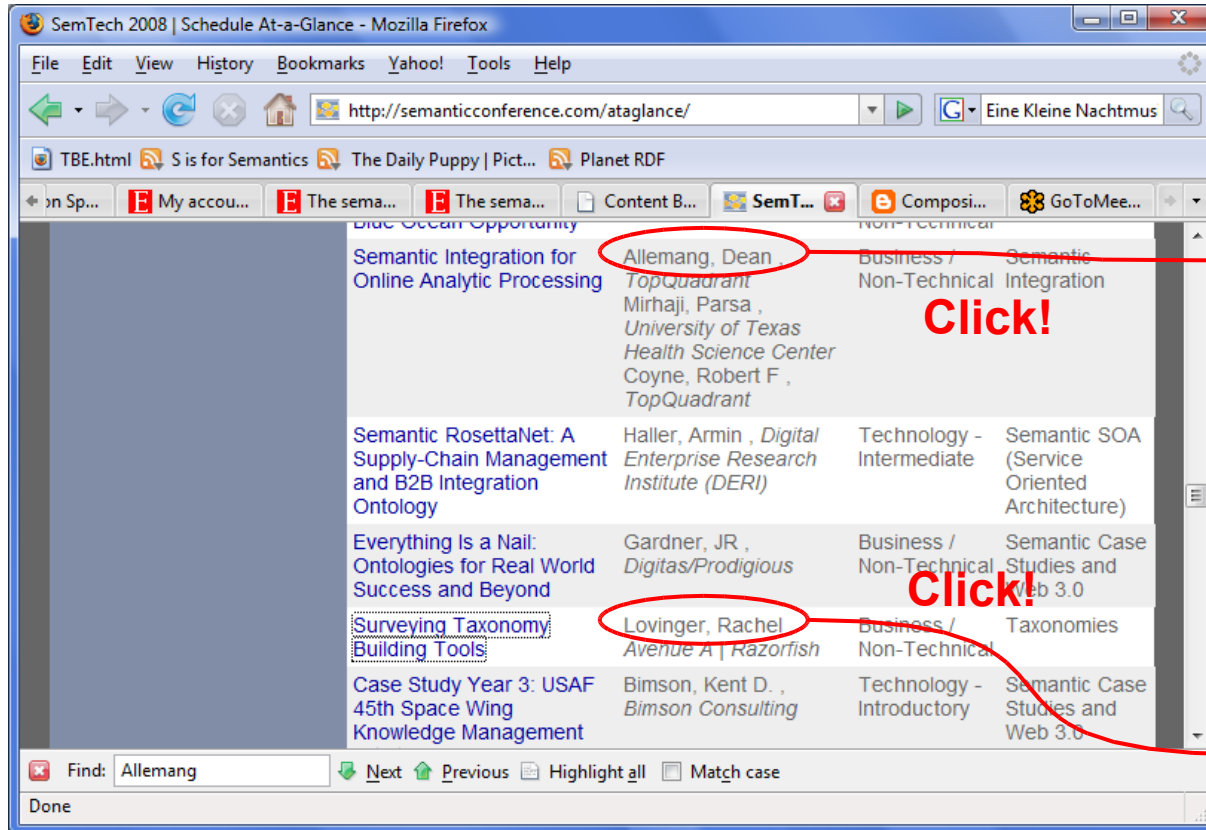
- *Store,*
- *Index, and*
- *Federate*
these triples



Semantic, Schemantic!

- In what sense is this “Semantic”?
 - It doesn’t understand Natural Language
 - It can’t figure out meanings of terms unless I tell it!
 - It doesn’t automatically reconcile misunderstandings!
-
- Semantic as reference – “What does this term refer to?”

Web 3.0?



We already had data on the web. Now we have a Web of Data

```
SELECT ?photo ?name
WHERE {?s a :Session .
       ?s :speaker ?p .
       ?p :label ?name .
       ?p :image ?photo .}
```



We already had data on the web. Now we have a Web of Data

Distributed data requirement #1: universal names

- “My notion of the word ‘shoe’ is different from yours”



blacksmith:shoe



macys:shoe

Distributed data requirement #2: Assertions about items



- “Nellie wear horseshoes”
- “Kelly wears Macy’s shoes”

wears



blacksmith:shoe



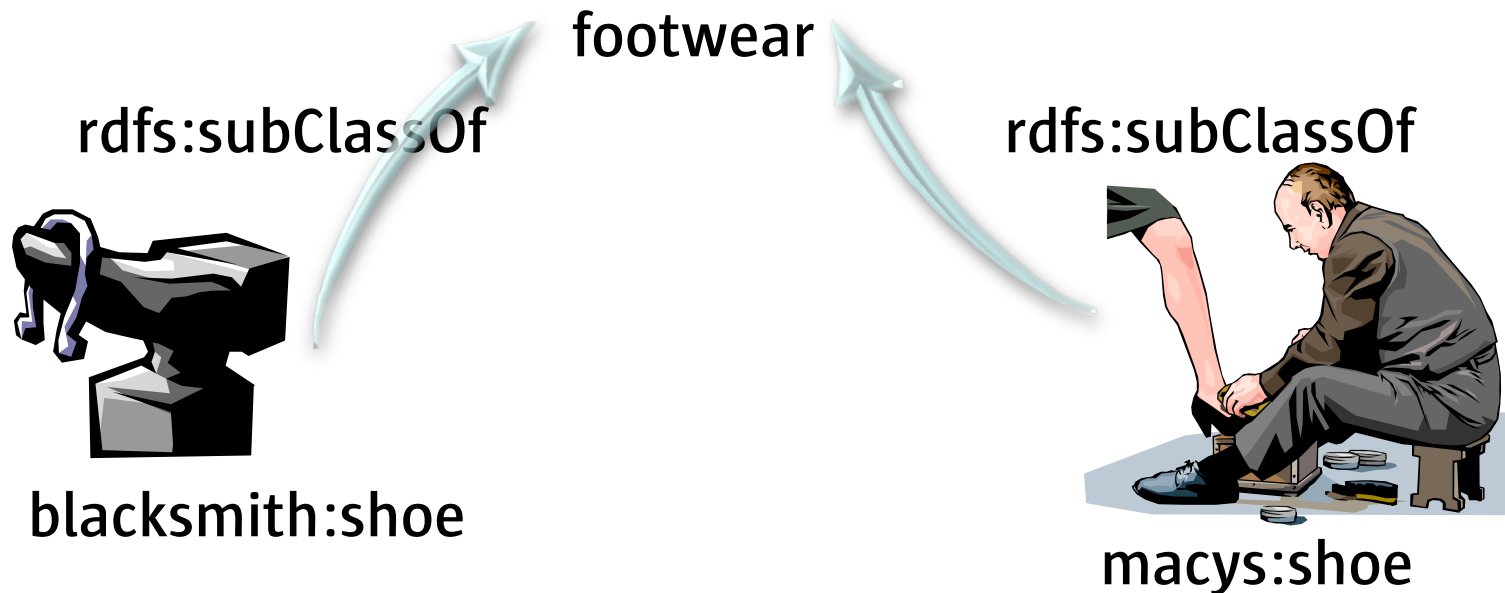
wears



macys:shoe

Distributed data requirement #3: Mapping items

Blacksmith's shoes do have something in common with Macy's shoes; they are all footwear!



A ***Semantic Model*** is a coherent set of statements like these

What is an Ontology?

^{wrong}
Lots of definitions
^

An Ontology is a reusable component of a distributed semantic model

Requirements of a Reusable data model?

> Extensible

Name	address	lat	long	Bike
Slausen	5585 Randolph St., Los Angeles 90032	-171.3	38.4	3
Union		-171.4	38.2	0
Vine		-170.9	37.9	12
McArthur		-170.4	38.1	3
Anaheim	1290 N. Long Beach Bl., Long Beach 90813	-171.3	38.2	12
Chinatown		-171.1	38.5	0
Beverly		-171.3	38.1	6

> Queryable Schema

```
SELECT ?property ?class
WHERE ?property rdfs:domain ?class .
```

-> (Name, address, lat, long), Station

Requirements of a Reusable data model?

➤ Referenceable

(on the web!)

➤ Mappable

“My word “latitude” means the same as your word “lat”.

my:latitude owl:equivalentProperty your:lat .

How small can be “(re-)usable”?

➤ Dublin Core – 15 properties

dc:title	dc:contributor	dc:source
dc:creator	dc:date	dc:language
dc:subject	dc:type	dc:relation
dc:description	dc:format	dc:coverage
dc:publisher	dc:identifier	dc:rights

➤ Do I have to use these?

- No, I just have to decide which of my words correspond . . .

my:author owl:equivalentProperty dc:creator .

my:title owl:equivalentProperty dc:title .

Mine is really small, but it gets the job done .

• •

- TopBraid Geography Ontology has one class and three properties:

map:DisplayLocation (class)

geo:lat (domain DisplayLocation)

geo:long (domain DisplayLocation)

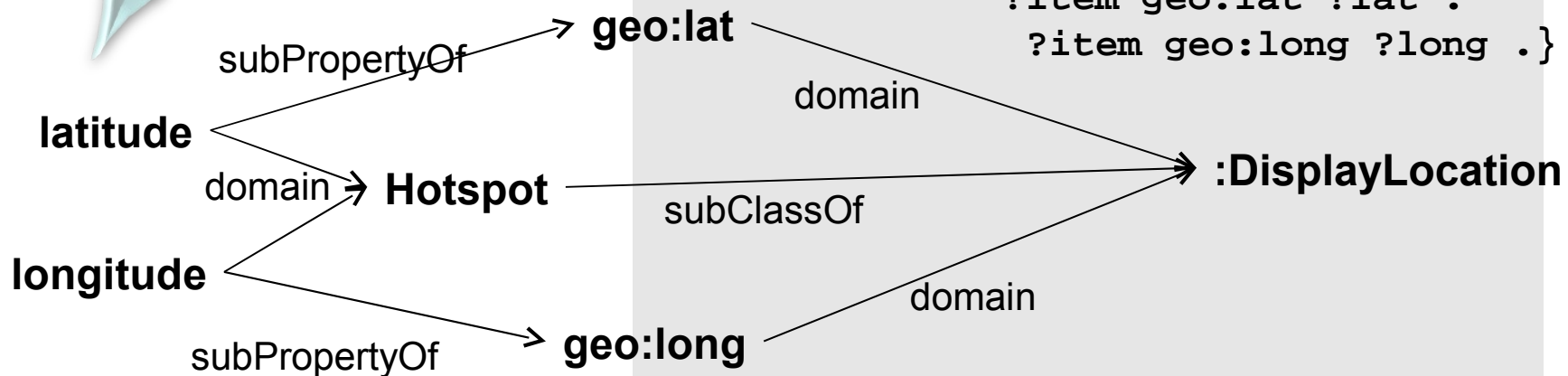
map:hasIcon (domain DisplayLocation)

Role of Ontologies in Model-based deployment

Name	latitude	longitude
Slausen	-171.3	38.4
Union	-171.4	38.2
Vine	-170.9	37.9
McArthur	-170.4	38.1
Anaheim	-171.3	38.2
Chinatown	-171.1	38.5
Beverly	-171.3	38.1

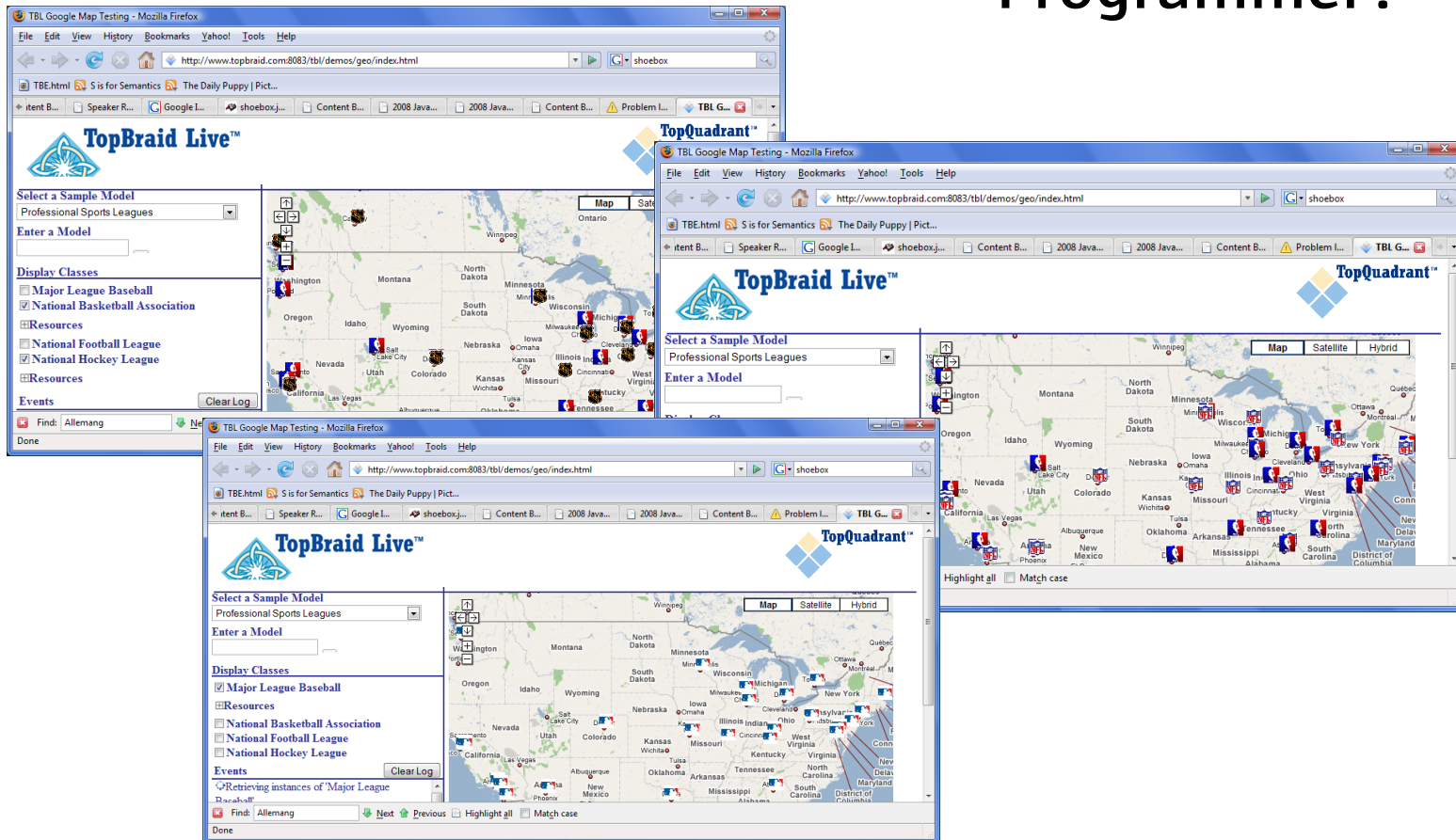


```
SELECT ?lat ?long
WHERE {?item a :DisplayLocation .
       ?item geo:lat ?lat .
       ?item geo:long ?long .}
```



Who is responsible for mash-ups?

➤ Programmer?



Information Exploration requires queryable schema!

TopBraid Ensemble User: test Logout
Project: kennedys.owl **TopQuadrant**

Concept Navigator

- College (19)
- Gender (2)
- Person (73)
 - Matriarch (1)
 - Patriarch (1)
 - Profession (34)
 - world:Address
 - world:Area
 - world:Country (5)
 - world:State

Results

Name	Label	Comment	First name	Year of death	Year of birth
Alfred Tucker	Alfred Tucker		Alfred		1967
Alina Mojica	Alina Mojica		Alina		1965
Amanda Smith	Amanda Smith		Amanda		1967
Andrew Cuomo	Andrew Cuomo		Andrew		1957

Graph

Graph visualization showing relationships between individuals:

- John Kennedy (Image) has parent Rose Fitzgerald (Image)
- John Kennedy has parent Eunice Kennedy (Image)
- Eunice Kennedy has child Rose Fitzgerald
- Eunice Kennedy has child Maria Shriver (Oval)
- Arnold Schwarzenegger (Image) has spouse Maria Shriver

Search **Basket** **SPARQL**

Search fields:

- Label:
- Comment:
- First name:
- Year of death: Min Max
- Year of birth: Min Max

Major brick-and-mortar retailer

- Why are people shopping online?
- What do we have that online shops don't?
 - Famous service
 - Parts
 - Follow-through
- What keeps people from taking advantage of follow-thru?
- They can't find their
 - Receipts
 - Service contracts
 - Model number
 - . . .

Solution

➤ Requirements:

- A web portal for consumers to maintain information about their homes and belongings
- Many different product lines and products need to be supported, all have different type of information
- Flexibility for future extensions and ease of maintenance
- Quick deployment – 10 weeks project to launch

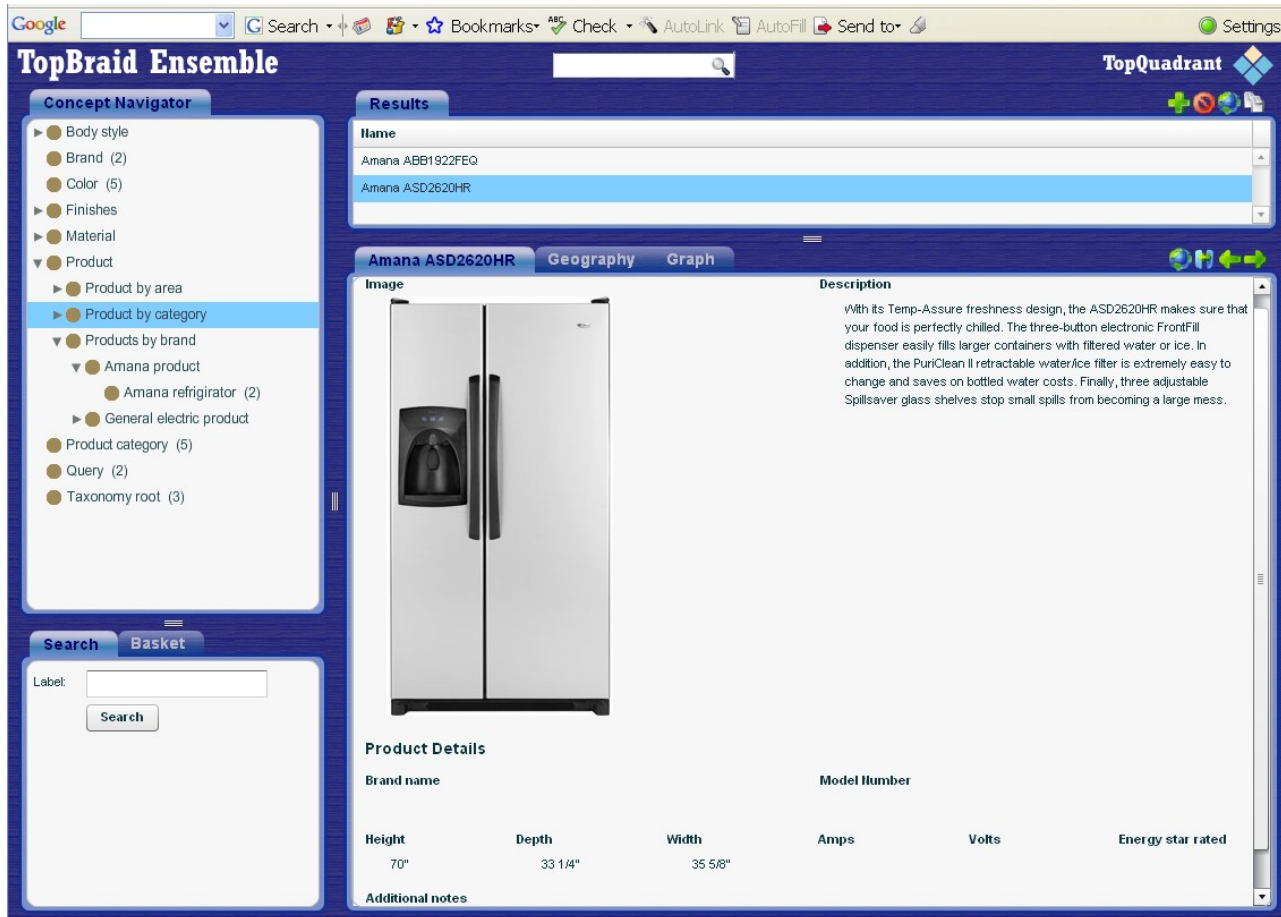
Need user forms to conform to data schema – FOREVER!!

The screenshot displays a Java-based IDE with the following components:

- Navigator (Classes):**
 - owl:Thing (0 + 123)
 - brand:Brand (2)
 - catalog:ProductCategory (5)
 - catalog:TaxonomyRoot (3)
 - owl:Nothing
 - products:BodyStyle (0 + 3)
 - products:Color (5)
 - products:Finishes (0 + 20)
 - products:Material (0 + 80)
 - products:Product (0 + 3)
 - catalog:ProductInArea (0 + 3)
 - catalog:ProductsWithBrand (0 + 3)
 - catalog:ProductWithCategory (0 + 3)
 - sparql:Query (2)
- Form Layout for Instances of catalog:ProductsWithBrand**
 - Element Nesting Hierarchy:**
 - Form "Resource Form"
 - Panel (2 columns)
 - Widget for products:image
 - Widget for products:description
 - Panel
 - Label "Product Details"
 - Panel (2 columns)
 - Widget for brand:brandName
 - Widget for products:modelNumber
 - Panel (3 columns)
 - Widget for products:height
 - Widget for products:depth
 - Widget for products:width
 - Layout Preview:**
 - products:image | products:description
 - Product Details
 - brand:brandName | products:modelNumber
 - products:height | products:depth | products:width
 - products:amps | products:volts | og:energyStarR
 - products:notes
 - Details of Selected Element:**
 - Display Label: Resource Form
 - Layout Columns: 1
 - ☒ Show properties with matching domain
 - ☐ Use rdf:label where possible
- Properties:**
 - brand:brandName
 - catalog:hasNavigationCatego
 - catalog:subCategoryOf
 - products:availableColor
 - products:hasBodyStyle
 - products:hasInsulationMater
 - products:hasWiringMaterial
 - products:image
 - products:madeOfMaterial
 - catalog:energyStarRated
 - products:amps
 - products:depth
 - products:description
 - products:height
 - products:modelNumber
 - products:notes
 - products:volts
 - products:width
 - owl:allValuesFrom
 - owl:cardinality
 - owl:maxCardinality
 - owl:minCardinality
 - owl:someValuesFrom
 - rdf:first
- Instances Table:**

[Resource]	rdf:type	rdfs:label	rdfs:comment
brand:GE_GSH25JF...	brand:GeneralElectric...	GE GSH25JFTWW	
catalog:Amana_AB...	brand:AmanaRefrigira...	Amana ABB1922FEQ	
products:Amana_A...	brand:AmanaRefrigira...	Amana ASD2620HR	

Render forms on web



The screenshot displays the TopBraid Ensemble web application interface. The page is divided into several sections:

- Top Navigation:** Includes a Google search bar, a search button, and a settings icon.
- Concept Navigator:** A sidebar on the left with a tree view of product categories. The 'Product by category' section is expanded, showing 'Amana product' and 'Amana refrigerator (2)'.
- Results:** A list of search results showing 'Amana ABB1922FEQ' and 'Amana ASD2620HR'.
- Product Details:** The main content area displays details for the 'Amana ASD2620HR' refrigerator. It includes an image of the refrigerator, a description, and a table of specifications.

Product Details Table:

Brand name	Model Number
Amana	ASD2620HR

Height	Depth	Width	Amps	Volts	Energy star rated
70"	33 1/4"	35 5/8"			

Description: With its Temp-Assure freshness design, the ASD2620HR makes sure that your food is perfectly chilled. The three-button electronic FrontFill dispenser easily fills larger containers with filtered water or ice. In addition, the PuriClean II retractable water/ice filter is extremely easy to change and saves on bottled water costs. Finally, three adjustable Spillsaver glass shelves stop small spills from becoming a large mess.

Solution features

- Hundreds of product lines supported
- Forms correspond to schema (queryable schema!)
- Schema can be updated (new products – new product features)
- Form continue to conform to schemas as system develops
- Deployed quickly (< 10 weeks) !

The Ontology Connection

➤ Queryable schema allows forms to be built/maintained automatically (e.g., from a query like this):

- ```
SELECT ?class ?prop ?type
WHERE {?prop rdfs:domain ?class .
 ?prop rdfs:range ?type .}
```

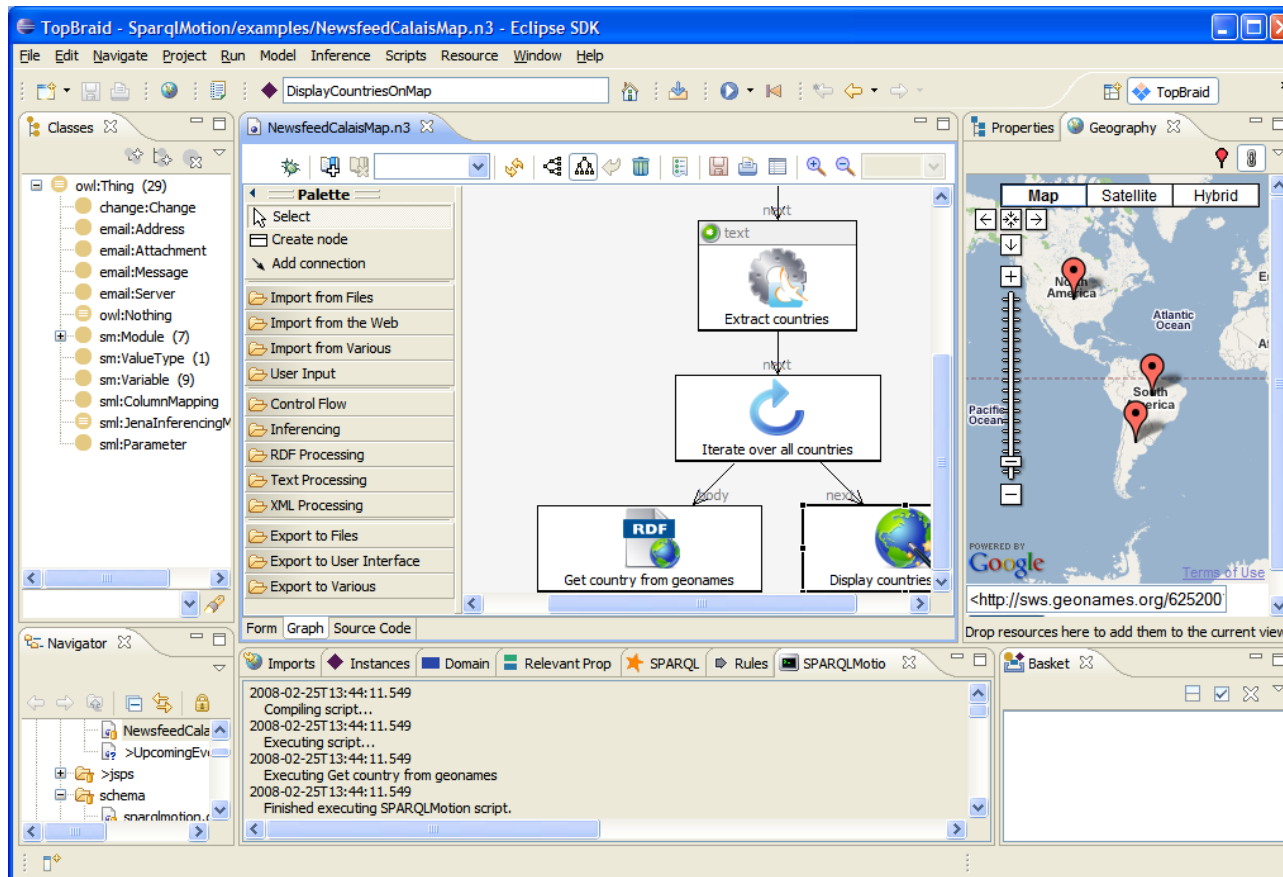
➤ Relations between product lines are represented and used

:Refrigerator rdfs:subClassOf :KitchenAppliance

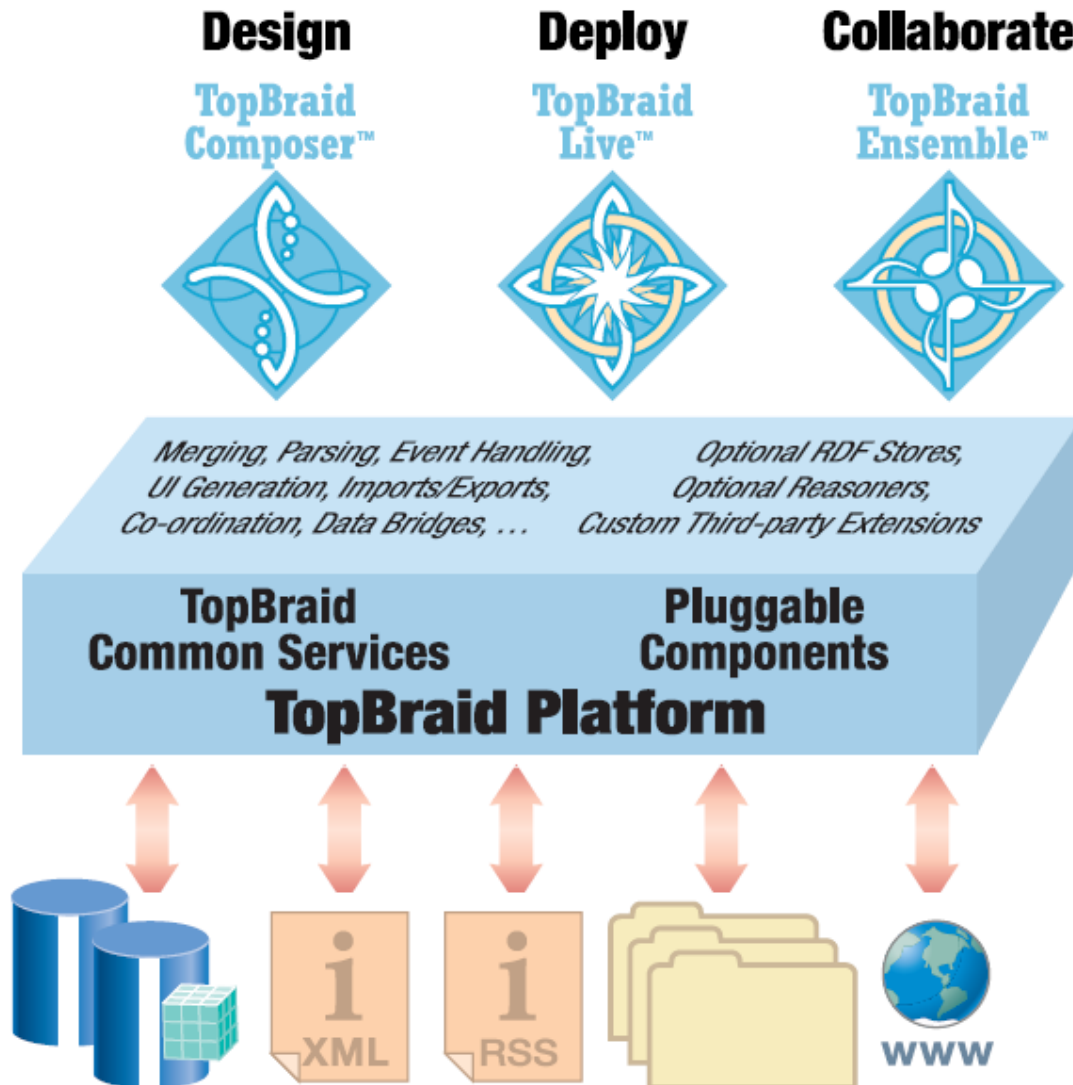
# The Java Connection

- Jena – open source system for parsing, storing, querying RDF
- Jena – de-facto API (regardless of store)
- Sesame – open source scalable triple store
- Eclipse – open platform for model development and deployment
- TopBraid Composer™ - IDE for semantic applications
- TopBraid Live™ - deployment platform for semantic applications

# Bringing it together – Java™ Technology, Eclipse and the Semantic Web



# Semantic Solution Architecture



# Evaluation

## ➤ Why Java Technology?

- Eclipse platform – plugins
- Platform independence
- Community support: Java Tools for RDF, OWL, Rules, etc.

## ➤ Why Semantic Web?

- Distribute data
- Survive in an uncertain world!

# Semantic Information Integration with TopBraid Suite™

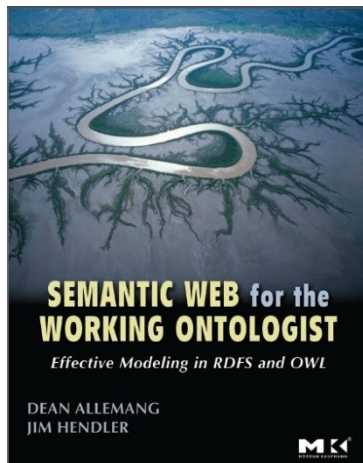
TopQuadrant, Inc.



DEMO

# For More Information

- <http://www.topquadrant.com/>
- Semantic Web for the Working Ontologist, Allemang & Hendler





# THANK YOU



Dean Allemang, Chief Scientist, TopQuadrant

