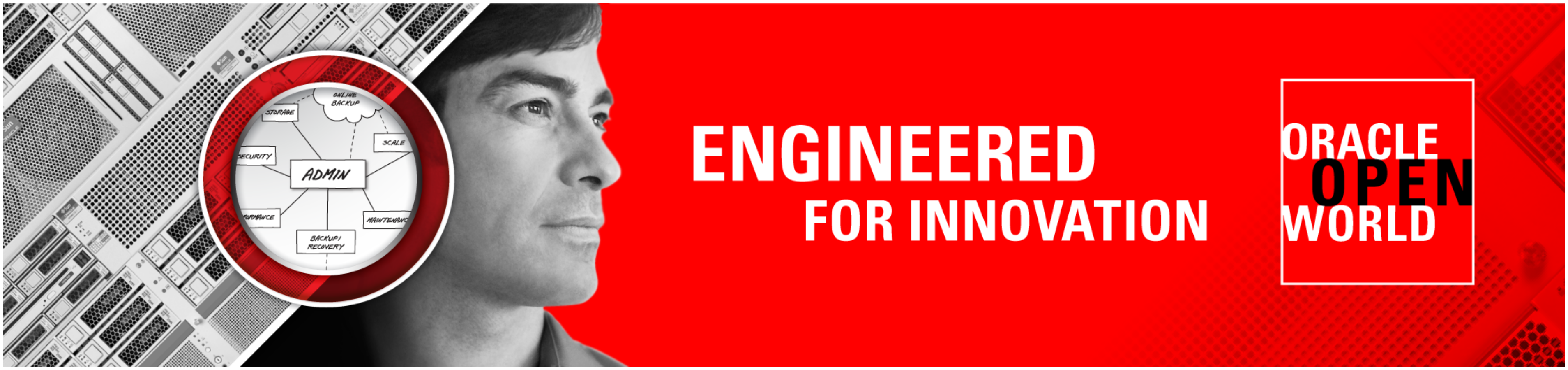
ORACLE®



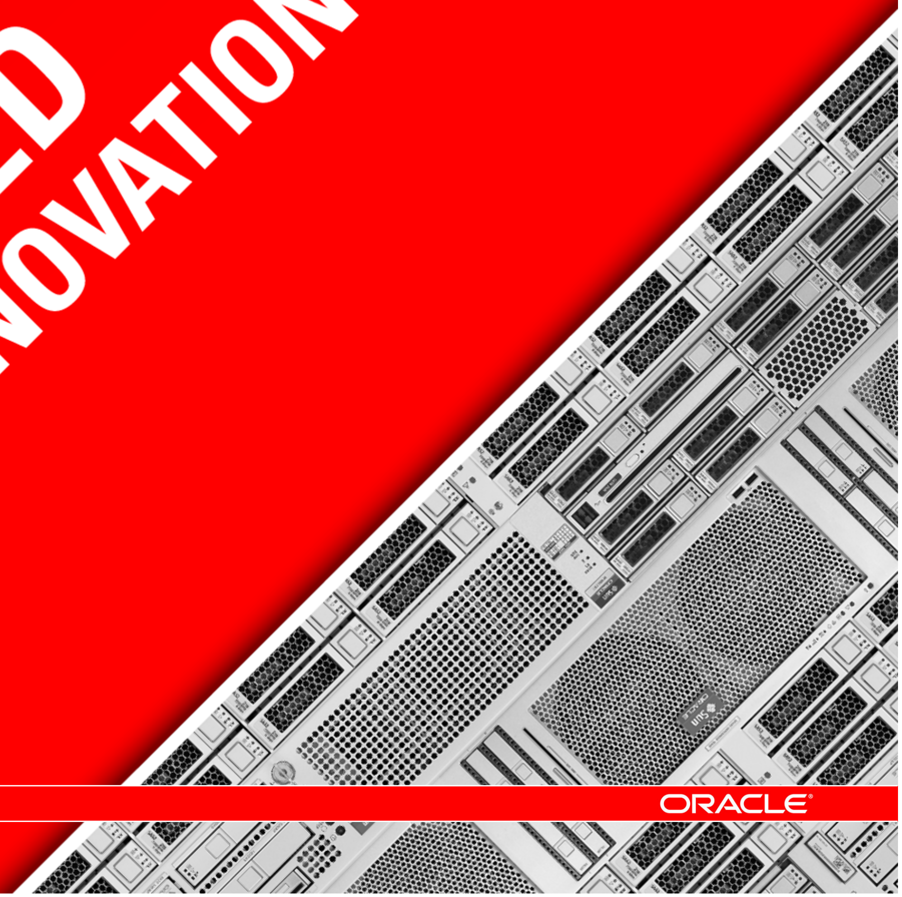
What can “semantic technologies" do for database applications?

Jim Steiner
Vice President, Server Technologies

Zhe Wu
Consultant Member of Technical Staff

ORACLE OPEN WORLD

ENGINEERED FOR INNOVATION



ORACLE

ORACLE OPEN WORLD

Latin America 2011

December 6–8, 2011

Tokyo 2012

April 4–6, 2012

Oracle OpenWorld Bookstore

- Visit the Oracle OpenWorld Bookstore for a fabulous selection of books on many of the conference topics and more!
- Bookstore located at Moscone West, Level 2
- All Books at 20% Discount

DigitalGuru
Technical Bookshop

ORACLE

 |

Oracle Products Available Online

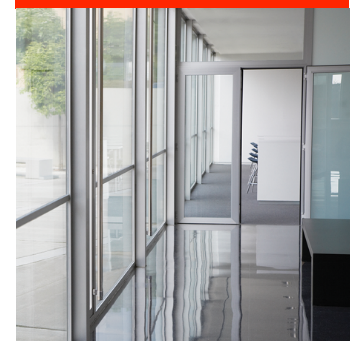
SHOP NOW

Oracle Store

Buy Oracle license and support online today at
oracle.com/store

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Program Agenda



- Oracle’s Semantic Strategy
- Introduction to Semantic Technologies
- How semantics complements existing approaches
- Semantics in Oracle Database
- Under the hood - Semantic Technologies in Oracle Database
- Customers and partners

Oracle’s Semantic Strategy

Eight Years Ago

ORACLE DATABASE 10g

W3C WORLD WIDE WEB consortium

RDF

ORACLE SPATIAL

Vision of Linked Open Data & Business Applications

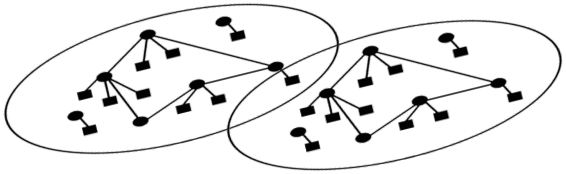
Integrate full breath of enterprise content (structured, spatial, email, documents, web services)

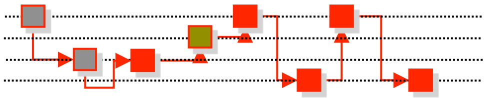
Reconcile differences in data semantics so that they can all “talk”and interoperate;

Resolving semantic discrepancies across databases, applications

Create consolidated “single” views across business applications

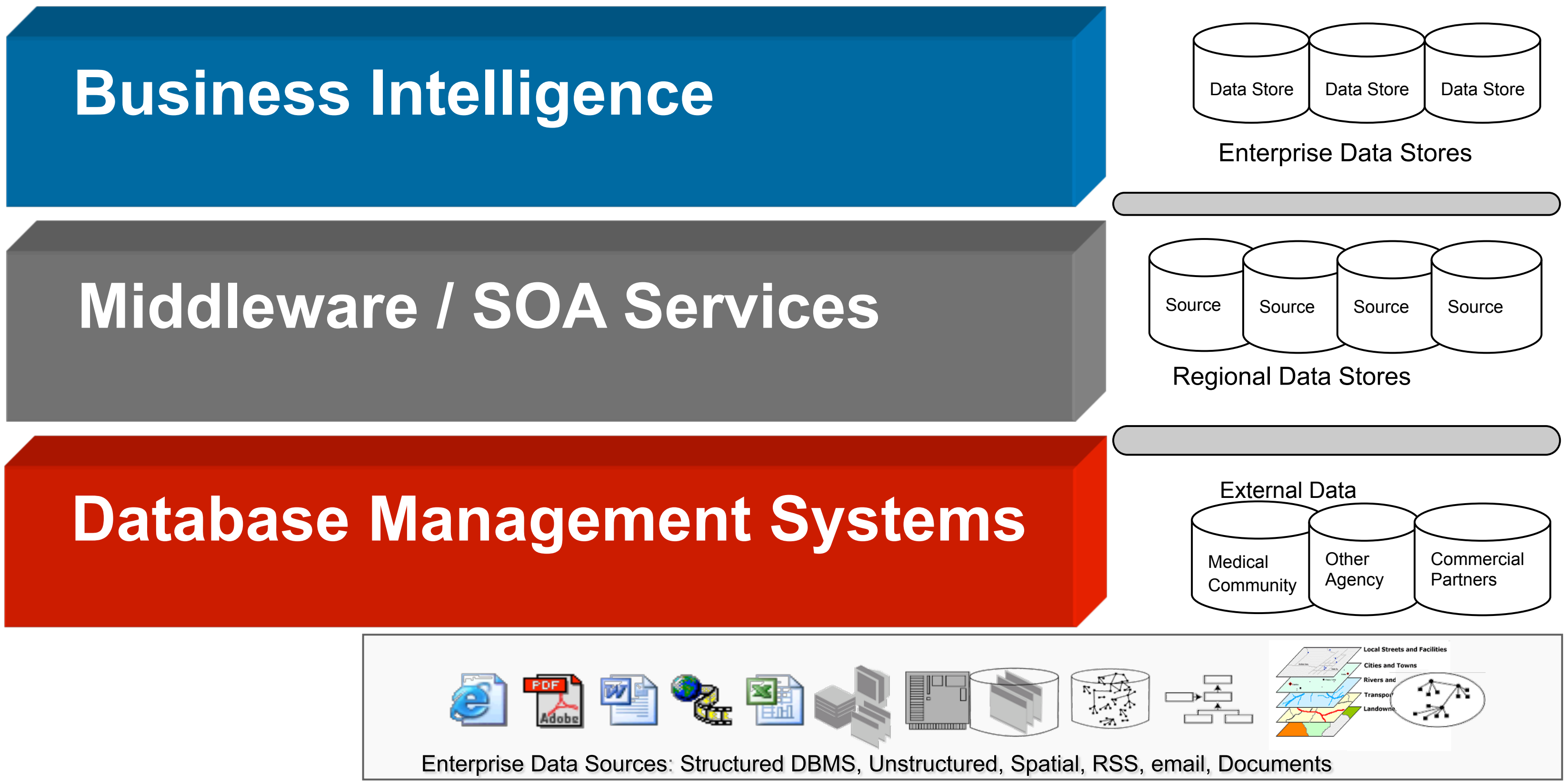
Model and implement common Business Processes





 |

Mainstream IT Platform



Introduction to Semantic Technologies

"Semantic technologies include software standards and methodologies that are aimed at providing more explicit meaning from the information that's at our disposal."

The CIO's Guide to Semantics

Dave McComb, Semantic Arts, Inc.

semantic arts Semantic and Service Oriented Architects

ORACLE

Fundamental Concepts and “building blocks”

1) Anything can be described by its unique relationship to something else

- John Smith Is At OpenWorld
- OpenWorld Is In San Francisco
- Seema Is Presenter of OOW Semantic Session

- This is called a **“triple”**
- Uniqueness in the triple is enforced by the inclusion of a **URI**

Subject	Relationship	Item
John Smith	Is At	Openworld
Openworld	Is In	San Francisco
Oracle	Has A Conference Called	Openworld
Seema Rao	Works At	Oracle
Seema Rao	Is Presenter of	OOW Semantic Session
John Smith	Is Registered for	OOW Semantic Session
OOW Semantic Session	Is Held	10/6/11, 12:00 Noon

ORACLE

Fundamental Concepts and “building blocks”

2) Implied relationships can be found in the data using rules

This is called **“inferencing”**

RULE:
1. OOW is the same as Openworld

“John and Seema were in San Francisco on 10/6/11”

Subject	Relationship	Item
Openworld	Is In	San Francisco
Openworld	Has A Session Called	OOW Semantic Session
Seema Rao	Is Presenter of	OOW Semantic Session
John Smith	Is Registered for	OOW Semantic Session
OOW Semantic Session	Is Held	10/6/11, 12:00 Noon

ORACLE

Fundamental Concepts and “building blocks”

3) Standard sets of related concepts can be stored to describe relationships and referenced to enhance query and discovery

This is called an **“ontology”**

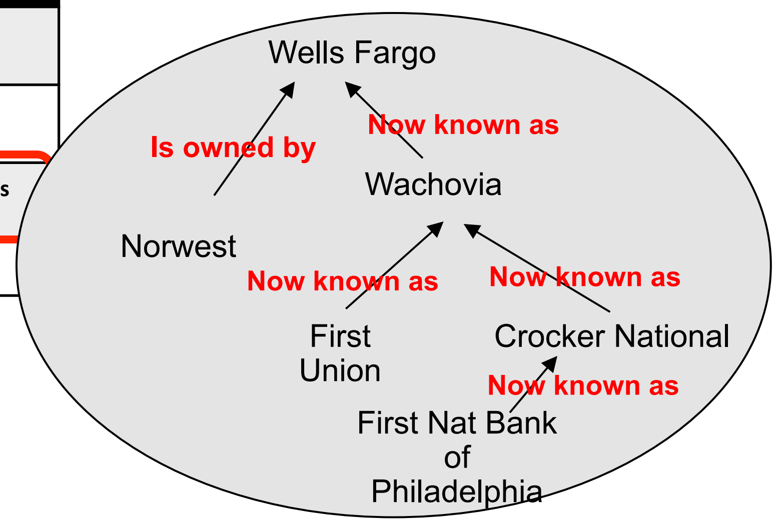
Type of Relationship	What you evaluate	What you compare	Opposite/Inverse Relationship
Lends to	Businesses and related parties	Businesses	Borrows from
Owns	Institutions and related parties	Institutions	Is owned by
Now known as	Corporate names and symbols	Corporate names	Previously known as
Operates in	Geographic hierarchy	Geographic name	No presence in

-- Holding companies own banks ➔ banks lend to other institutions ➔ …

Fundamental Concepts and “building blocks”

4) Conceptually, Semantic applications look at things as being represented as graphs, rather than tables

Type of Relationship	What you evaluate	What you compare	Opposite/Inverse Relationship
Lends to	Businesses and related parties	Businesses	Borrows from
Owns	Institutions and related parties	Institutions	Is owned by
Now known as	Corporate names and symbols	Corporate names	Previously known as
Operates in	Geographic hierarchy	Geographic name	No presence in



In Oracle Database, we use Triples and Key relationships to represent nodes and links in the Graph.

 |

Fundamental Concepts and "building blocks"

2) Implied relationships can be found in the data using rules

This is called **"inferencing"**

RULE:
1. OOW is the same as Openworld

"John and Seema were in San Francisco on 10/6/11"

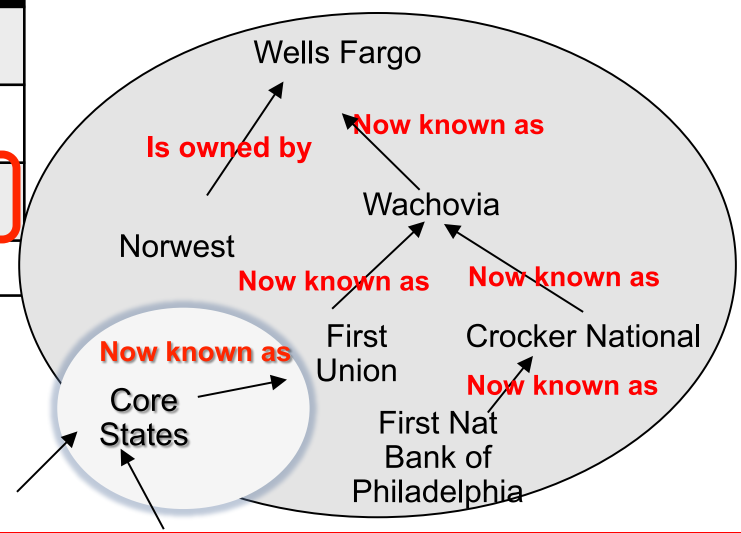
Subject	Relationship	Item
Openworld	Is In	San Francisco
Openworld	Has A Session Called	OOW Semantic Session
Seema Rao	Is Presenter of	OOW Semantic Session
John Smith	Is Registered for	OOW Semantic Session
OOW Semantic Session	Is Held	10/6/11, 12:00 Noon

 |

Fundamental Concepts and “building blocks”

4) Conceptually, Semantic applications look at things as being represented as graphs, rather than tables

Type of Relationship	What you evaluate	What you compare	Opposite/Inverse Relationship
Lends to	Businesses and related parties	Businesses	Borrows from
Owns	Institutions and related parties	Institutions	Is owned by
Now known as	Corporate names and symbols	Corporate names	Previously known as
Operates in	Geographic hierarchy	Geographic name	No presence in

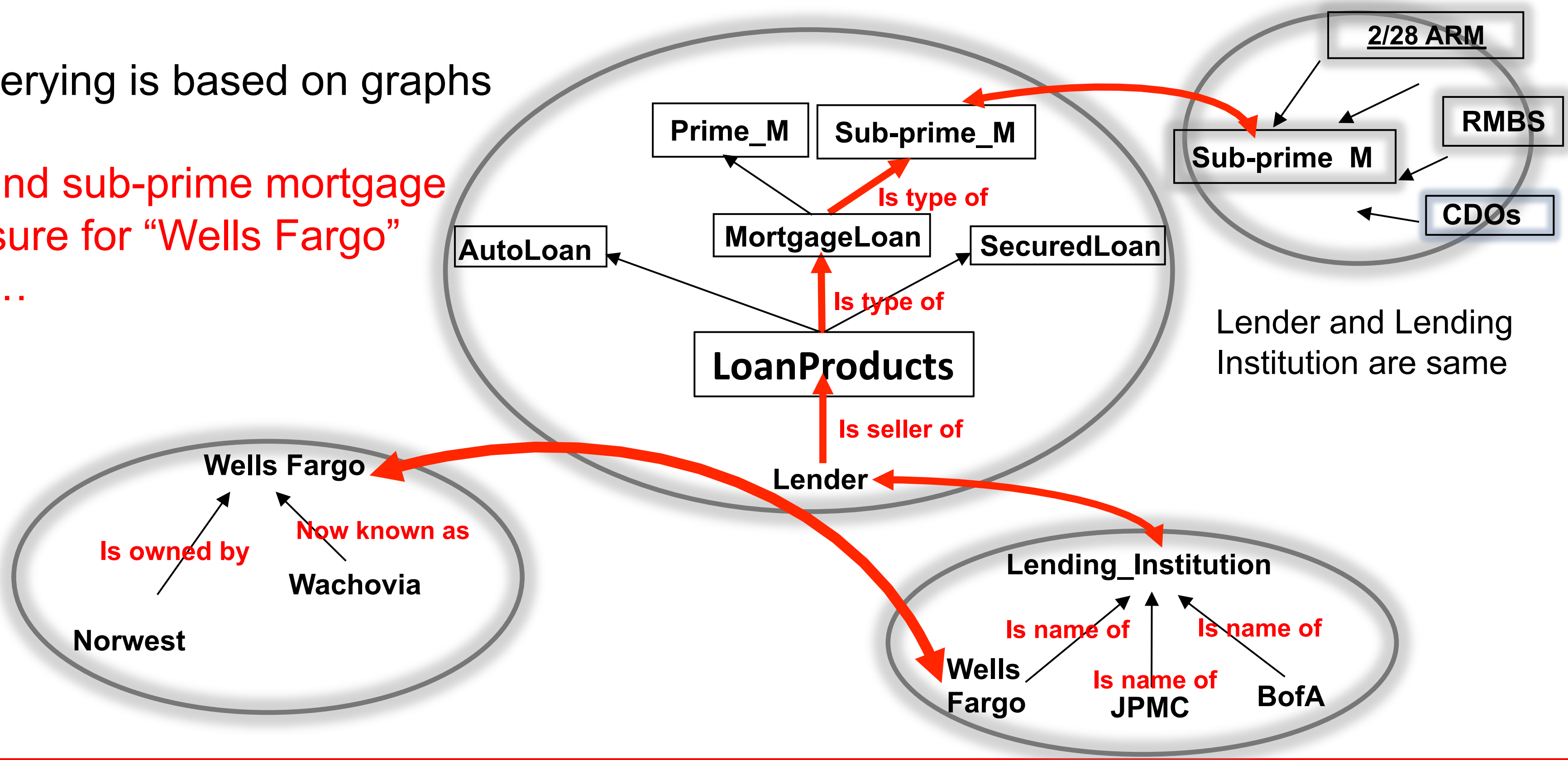


In Oracle Database, we use Triples and Key relationships to represent nodes and links in the Graph.

Fundamental Concepts and “building blocks”

5) Querying is based on graphs

Ex: Find sub-prime mortgage exposure for “Wells Fargo” bank…



Recap: Key ideas

- Based on fundamentally different Open World Assumption
  – What is unknown is undefined (not false) - that supports discovery
- Schema are flexible, evolving, can’t be known in advance
  – Rich, real world relationships are modeled in the data
- Every data element is uniquely identified - supports integration
  – Data & relationships are machine-readable
- Pattern query language supports discovery workflows

How semantics complements existing approaches

What Use Cases Benefit From Semantic Analysis?

- Data Warehousing / BI
  - Supports rapidly changing schema, inexact queries, and addition of new facts to support analysis
  - Query across internal OLAP, OLTP, and unstructured content repositories
- Metadata Management
  - Management of multiple,and inconsistent metadata registries
  - Graphs represent existing metadata models and enable merging of models
- Data Mining
  - Extend search w/ the meaning of terms and concepts
  - Discover new relationships w/ built-in rules, like transitivity
- Social Applications
  - Perform social graph analytics (e.g. clustering, centrality, degrees of separation)
  - Easily merge data from different sources using “same-as” relationships

ORACLE

 |

Can't These Use Cases Be Addressed By Existing Approaches?

To a degree…

- With Data Warehousing, but …
- With Master Data Management, but…
- With Data Mining, but…
- With Social Applications, but…

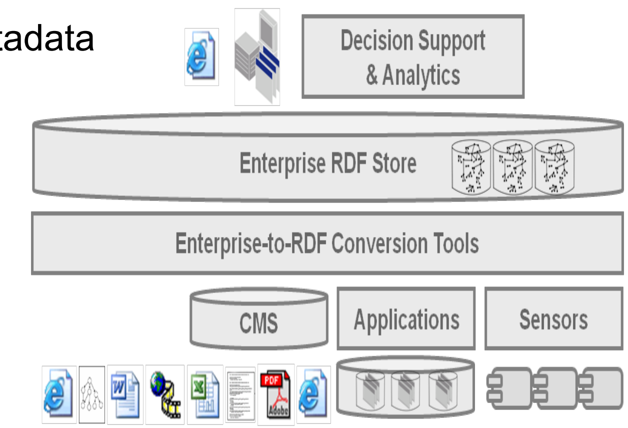
Semantics with Data Warehousing

Traditional Data Warehousing is challenged when:

- Data sources are not well known
- Data is non-relational: Text, XML, Spatial
- Goal is discovery, rather than analysis of known dimensions

How Semantics Can Help

- Metadata model designed to “evolve” without impacting existing metadata models and application schemas
- Automatically classify all data based on business concepts
- Perform inferencing and pattern queries to discover new facts and relate hidden relationships



Complements Traditional Data Warehousing and BI

Ontology-assisted SQL queries on existing DW tables

- Graph queries on semantic indexes derived from the DW tables
- Oracle Business Intelligence EE Support

 |

Semantics with Data Integration:

Master Data Management, Enterprise Information, Integration, Electronic Data Interchange

Traditional Metadata Registries are Challenged By:

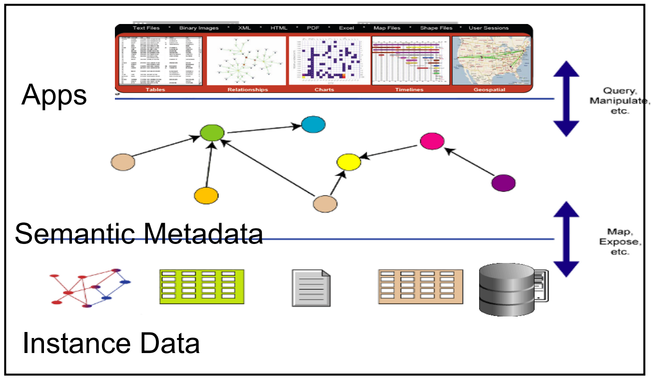
- Complex management of many inconsistent metadata registries
- Integrating registries for Text, XML, and relational data
- Difficult to analyze patterns and relationship with relational metadata

How Semantics Can Help

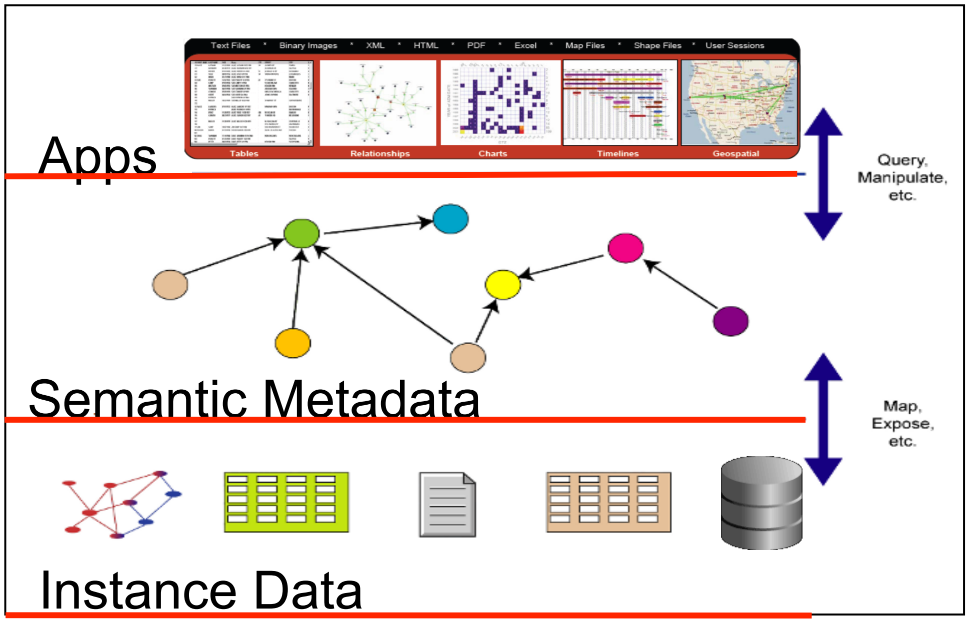
- Flexible schema that can support all data types
- Facilitates definition of unified metadata content vocabulary

Complements Traditional Data Integration Approaches

- Can support existing MDM, EII metadata repositories
- Supported by key Oracle technologies: Golden Gate (ETL, XML, BPM)



How Does Semantics Address Data Integration?



Semantics introduces an enterprise metadata framework. The metadata graph associates underlying instance data to other data resources based on their semantics. This linking of resources enables interoperability between apps that exchange data.

Semantics Complements Data (Text) Mining

Traditional Data Mining is Challenged When:

- Nature of end user problems cannot be defined in advance
- Querying across databases, syndicated content, web sites, real-time feeds, etc. is complex when underlying schemas are inconsistent
- Discovering patterns on wide ranging data types and sources


COMPUTERWORLD

Print Article Close Window

N.Y. bomb plot highlights limitations of data mining

Like weather forecasting, data mining can predict major storms but not where each raindrop will fall

Jaikumar Vijayan

May 5, 2010 (Computerworld)

Saturday's botched bombing attempt in New York City provides an example of why the use of data mining approaches to uncover potential terrorism plots is a little like weather forecasting.

How Semantics Can Help:

- Apply rules on underlying data and concepts to discover hidden patterns
- Easily merge data from different sources using “same-as” relationships
- Enable new discovery techniques: Question and Answer, Search, Navigation

Complements Traditional Data Mining Approaches:

- Supports rich variety of 3rd party Natural Language Processing tools
- Can leverage rich clustering algorithms from Oracle Data Miner and R

ORACLE

 |

Semantics Complements Social and Collaboration Apps

Traditional Social Applications have unique requirements:

- Integrate metadata model for blogs, wikis, calendars, IM, WebEX, voice, and video news feeds, public cloud, etc.
- Reconcile messy, user-generated content tagging to common model – automatically.

How Semantics Can Help

- Validate semantic/structural consistency
- Unify content metadata model
- Support high volume and highly transactional workloads

Complements Existing Social Apps:

- Supports standards relevant to social applications: XML, RDF, OWL, SPARQL
- Chosen by Cisco for its Social App integration platform

Semantics in Database

How Does Semantics Enable Better SQL Results?

Query: “Find all entries in patient diagnosis column related to ‘Upper_Extremity_Fracture’ "

Patients diagnosis table

P_ID	DIAGNOSIS
1	Hand_Fracture
2	Rheumatoid_Arthritis
3	Finger_Fracture

No results w/ the usual SELECT

```
SELECT p_id, diagnosis
 FROM Patients
 WHERE diagnosis =
'Upper_Extremity_Fracture';
```

ORACLE

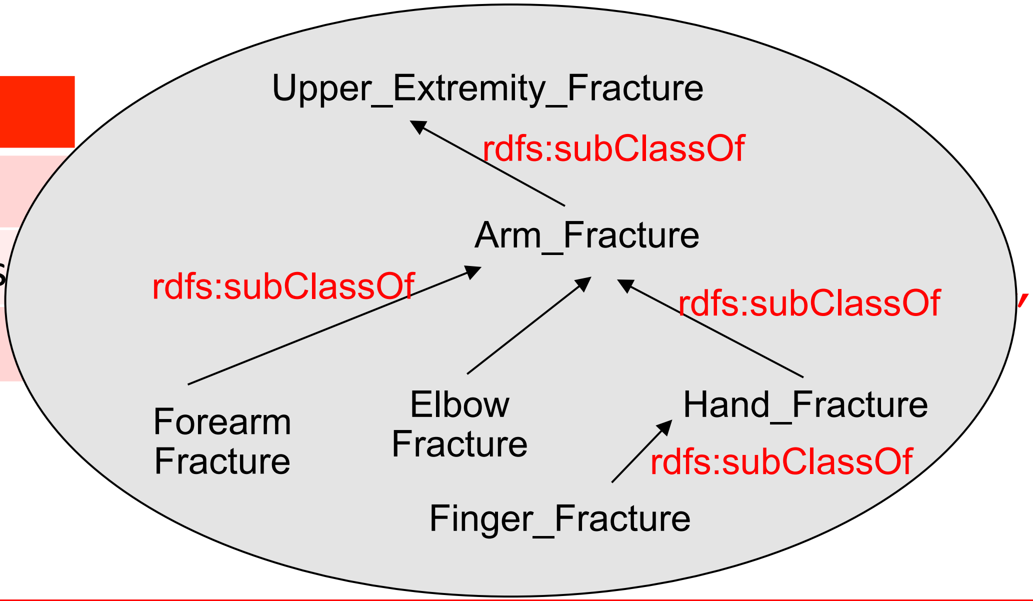
How Does Semantics Enable Better SQL Results?

Ontology-Assisted SQL Query

National Cancer Institute (NCI) medical ontology

Patients diagnosis table

P_ID	DIAGNOSIS
1	Hand_Fracture
2	Rheumatoid_Arthritis
3	Finger_Fracture



’ ;

 |

How Does Semantics Address Better SQL Results

SEM_RELATED SQL operator expands the SQL WHERE clause with related terms from the Ontology

Patients diagnosis table

P_ID	DIAGNOSIS
1	Hand_Fracture
2	Rheumatoid_Arthritis
3	Finger_Fracture

NCI medical ontology:

Finger_Fracture → Hand_Fracture → Arm_Fracture → Upper_Extremity_Fracture

Results: Hand_Fracture, Finger Fracture

```
SELECT p_id, diagnosis
 FROM Patients
WHERE SEM_RELATED (
    diagnosis,
    'rdfs:subClassOf',
    ':Upper_Extremity_Fracture',
    'Medical_ontology' = 1)
```

 |

ORACLE

How Does Semantics Address Discovery: Inferencing

- Business Constraint: An applicant can have only 1 line of credit
- Finds implicit (unstated) relationships in the data
- Allows machine-driven discovery based on W3C standard

Data:
Bank Data: John T Smith
hasSocialSecNum ’123’
Credit bureau report: JT Smith
hasSocialSecNum ’123’
Credit bureau report: JT Smith
hasHomeEquityLoan ’789’

User-defined rule:
if PersonA hasSocialSecNum SS1
and PersonB hasSocialSecNum SS1
then PersonA **sameAs** PersonB

Using OWL Construct:
hasSocialSecNum rdf:type
owl:InverseFunctionalProperty

Machine adds the inference:
John T Smith **sameAs** JT Smith

Therefore: John T Smith hasHomeEquityLoan ’789’ and isn’t qualified for another line of credit

How Does Semantics Address Finding Patterns in Data?

- Has its own graph query language - W3C SPARQL
- It's a simpler way to write query patterns that need to be joined together

e.g., Find pairs of siblings (same parents)

SPARQL		SQL
SELECT ?x ?y FROM <rdf_graph> WHERE { ?x hasFather ?f . ?x hasMother ?m . ?y hasFather ?f . ?y hasMother ?m . FILTER(?x != ?y) }	➔	SELECT g1.subject x, g3.subject y FROM rdf_graph g1, rdf_graph g2, rdf_graph g3, rdf_graph g4 WHERE g1.predicate = 'hasFather' AND g2.predicate = 'hasMother' AND g3.predicate = 'hasFather' AND g4.predicate = 'hasMother' AND g1.subject = g2.subject AND g3.subject = g4.subject AND g1.object = g3.object AND g2.object = g4.object AND g1.subject != g3.subject

Does this mean I can't use SQL?

- Not at all, Oracle SQL extended to include graph queries
- SEM_MATCH table function is based on SPARQL
- Uses patented SQL table function rewrite
  - Converts graph query to SQL, the whole SQL query is optimized
  - Returns a whole set of results rather than one result per table function call

SEM_MATCH: Adding SPARQL to SQL

Query example: “List everyone’s name and anyone else they know”

(Analogous to an outer join)

```
SELECT n1, n2
FROM
 TABLE(
  SEM_MATCH(
   'PREFIX foaf: <http://...>
    SELECT ?n1 ?n2
    FROM <http://g1>
    WHERE {?p foaf:name ?n1
           OPTIONAL {?p foaf:knows ?f .
                     ?f foaf:name ?n2 }
          }
    ORDER BY ?n1 ?n2',
  SEM_MODELS('M1'),…));
```

n1	n2
Alex	Jerry
Alex	Tom
Alice	Jerry
Alice	Tom
Alice	John
Jerry	Tom
Tom	Jerry
Zack	-

 |

What about Java Programmers?

- Jena/Sesame Adapter provides the following features:
  - A set of easy-to-use and performant Java APIs to access Oracle database
  - A standard-compliant SPARQL web service endpoint
    - SPARQL Protocol, Federated SPARQL, SPARQL update
  - Various data loading (RDF/XML, N-TRIPLES, Turtle, …) with long literal support
  - Oracle-specific extensions for query execution control and management
    - Timeout, abort, S2S, hints in SPARQL syntax, property path, result cache, mid-tier cache, user-defined functions…
  - Integration with various tools including (TopBraid Composer, Pellet)

ORACLE

Semantics in Oracle Database: Under the Covers

Is It Complicated To Represent Semantic Data?

- No, semantic data is stored in relational tables
- Each row has 3 columns and describes 1 data attribute

Semantic triple	**Subject**	Property or Predicate	Object
Example data	**123-45-6789**	HasName	John Smith
Relational representation	**Primary key**	Column name	Column value

- An object can be a link (URI) to another subject
  E.g., John Smith sameAs J.T. Smith
- Linking 2 subjects by a property creates a relationship
- Relationships connect people, places, things, events, ideas, and data
- Data is commonly visualized as a network (a graph) of Resource Description Framework (RDF) data

ORACLE

How is Semantic Data Stored in Oracle Database?

Semantic (RDF) data is mainly stored in two compressed tables in MDSYS schema

RDF_Value$ Table

- Stores the canonical & user value of each S, P and O
- Globally unique Hash id for each data element
- Reconciles non-differences like numbers 25 & 0025
- Canonical values enable more accurate queries & joins

RDF_Link$ Table

- Stores the triples (quads) using the hash Ids in Value$
- Partitioned w/ local indexing to accelerate loading, querying and inferencing

ORACLE

Performance, performance, performance!

Isn’t It Hard & Slow Joining All These Triple Patterns?

- No, the Oracle Database query optimizer determines an efficient approach to query the triple data

- Recent LUBM Benchmark results for 14 queries:
  - Sun M8000: returned over 149 million triples in 4.3 min.
  - Exadata X2-2 returned over 465.8 million triples in 4.3 min.

- The details
  - Uses nested loop lookups and hash joins as needed
  - Optimizer uses dynamic sampling to determine join order and type
  - Optimizer plans can be influenced w/ hints
  - Resulting hashed triples are expanded into subject-predicate-object values by a join to the values table

- **Storing data in a common triple format simplifies integration**

ORACLE

Performance, performance, performance!
Does This Get Memory Bound?

- No, operations are performed in Oracle Database
- SPARQL queries are converted to SQL
- Query processing is performed by the SQL processing engine
- Inferencing is performed by the inferencing engine in the database and persisted before query time

Performance, performance, performance!

Tune for Performance!

- A balanced hardware setup is **required** for performance
  - I/O throughput is critical to performance (therefore, avoid RAID5, avoid single-harddrive setup, use SSDs, use ASM w/ multi-disks, etc.)
- OS has to be tuned
  - shmmax, shmall, aio-max-nr, sem, …
  - Network MTU, Oracle SQL*Net parameters including SDU, TDU, SEND_BUF_SIZE, RECV_BUF_SIZE
- Database init parameters need to be tuned
  - SGA, PGA, filesystemio_options, db_cache_size, …
- Statistics need to be maintained. Parallel executions can help
- Follow best practices described in
  - Oracle Database Semantic Technologies Developer's Guide
  - SemTech (2010, 2011) Workshop slides

ORACLE

 |

Performance, performance, performance!

Oracle Performs Best on a Balanced Hardware Setup!

Degrees of Parallelism	Data set	Load (B triples/min.)	OWL Inference (M triples/min.)	Query (M answers/min.)
128	LUBM 8K	1.1B / 28’ 11” *	869M / 62’	149+ m in 4.3’
64	LUBM 8K	1.1B / 53’ 49”	869M / 114’	149+ m in 4.3’

Setup:

Hardware: Sun M8000 -- 16 SPARC64 VII+ 3.0 GHz CPUs [64 total Cores 128 Parallel Threads] -- 512 GB Ram -- Dual F5100 Flash Arrays (160 total drives)

Storage required: 330GB + 110GB of temporary table space

Software: Oracle Database 11.2.0.2.0 + Patch 9825019: SEMANTIC TECHNOLOGIES 11G R2 FIX BUNDLE 3

SGA_TARGET=256G and PGA_AGGREGATE_TARGET=206G

* 1.1B / 28’ 11” means 1.1 billion triples in 28 minutes and 11 seconds

ORACLE

Performance, performance, performance!

Oracle Performs Best on a Balanced Hardware Setup!

Degrees of Parallelism	Data set	Load (B triples/min.)	OWL Inference (M triples/min.)	Query (M answers/min.)
128	LUBM 25K	3.4B /105’	2.7B / 160’	0.47B in 8.7’
64	LUBM 25K	3.4B / 186’	2.7B / 210’	0.47B in 12.3’

Setup:

Hardware: Sun M8000 -- 16 SPARC64 VII+ 3.0 GHz CPUs [64 total Cores 128 Parallel Threads] -- 512 GB Ram -- Dual F5100 Flash Arrays (160 total drives)

Storage required: 900GB + 300GB of temporary table space

Software: Oracle Database 11.2.0.2.0 + Patch 9825019: SEMANTIC TECHNOLOGIES 11G R2 FIX BUNDLE 3

SGA_TARGET=256G and PGA_AGGREGATE_TARGET=206G

* 1.1B / 28’ 11” means 1.1 billion triples in 28 minutes and 11 seconds

Semantics in Oracle Database: The Benefits

Why Bother?

Semantics enriches relational and XML technologies through richer modeling, integration, reasoning, and discovery.

Machine-driven discovery

- Relationships connect people, places, things, events, ideas, and data
- Patterns of relationships can be queried for insights
- Implicit relationships in the data are discoverable

Flexibility

- Schema evolves easily by adding a triple
- Queries don't have to be planned for in advance

Why Use Oracle Database As A Semantic Store?

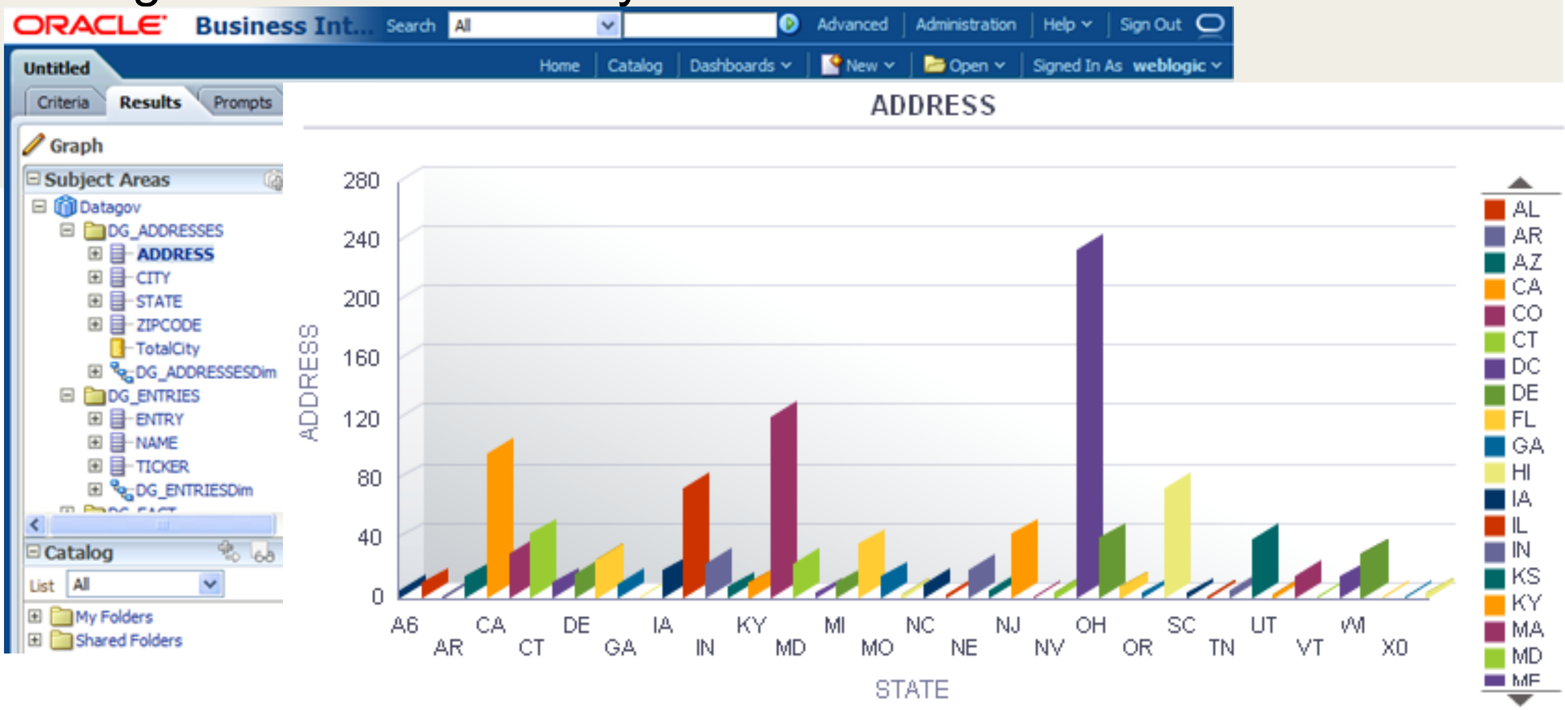
- ✓ Scalable for tens of billions of triples and more
- ✓ High availability w/ DataGuard
- ✓ DoD-strength security

- Support for Oracle Exadata, RAC, SQL*Loader direct path load, Parallelism, Oracle Label Security
- Table compression reduces storage upwards of 60%
- Partitioning and local indexing by application table
- B-tree indexing of triples
- More accurate results: canonical data stored once w/ hash ID
- Hash IDs minimize the size of the triple table, speeds up joins

 |

Why Use Oracle Database As A Semantic Store?

- ✓ Easy integration of RDF data with Business Intelligence (OBIEE) through SPARQL Gateway



 |

Why Use Oracle Database for Semantic Queries & inferencing?

✓ Support for SPARQL and SPARQL patterns in SQL

✓ Native inferencing engine supports W3C standards

✓ Plug-in architecture supports more inferencing engines

- Support for SPARQL 1.1 w/ open source Jena and Sesame
- Mixed SPARQL and SQL queries
- Patented Table Function Rewrite converts SPARQL to SQL
- Support for incremental and parallel inferencing
- Support for W3C standards RDFS, OWL2 RL, EL+, SKOS
- 3rd party in-memory inference supported w/ Oracle's inferencing
- Support for user-defined rules

ORACLE

 |

Summary: Oracle Database Semantic Technologies

- Scalable to billions of triples
- RAC & Exadata scalability
- Compression & partitioning
- SQL*Loader direct path load
- Parallel load, inference, query
- Oracle DataGuard availability
- Triple-level DoD-strength security
- Choice of SPARQL or SQL
- Native inference engine
- W3C standards compliance
- Semantic Indexing of text & docs
- Growing ecosystem of 3rd party tools partners

W3C WORLD WIDE WEB consortium

RDF

Key Capabilities:

Load / Storage
- Native RDF graph data store
- Manages billions of triples
- Optimized storage architecture

Query
- SPARQL-Jena/Joseki, Sesame
- SQL/graph query, b-tree indexing
- Ontology assisted SQL query

Reasoning
- RDFS, OWL2 RL, EL+, SKOS
- User-defined SWRL-like rules
- Incremental, parallel reasoning
- Plug-in architecture

ORACLE

 |

Customers Deploying Semantic Technologies

Life Sciences

Lilly

Pfizer

SIB
Swiss Institute of Bioinformatics

Defense/ Intelligence

National Geospatial-Intelligence Agency

Education

The University of Michigan 1817

Clinical Medicine & Research

Cleveland Clinic

Telecomm & Networking

Cisco

Hutchinson 3G Austria

Publishing

Westlaw
Thomson Reuters

ORACLE

 |

Semantic Technologies Partners
Integrated Tools and Solution Providers:

Ontology Engineering

Reasoners

NLP Entity Extractors

Open Source Frameworks

Joseki

Sesame

Standards

Applications

SI / Consulting

For More Information

Oracle RDF

James.Steiner@oracle.com

Alan.Wu@oracle.com

oracle.com

ORACLE

 |

Q&A