Program Agenda

1. Oracle’s Semantic Strategy
   – Xavier Lopez, Oracle

2. Using Semantics for Complex Financial Workflows
   – Seamus Hayes, Polar Lake
The preceding 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.
Please don’t miss:

“What Can Semantics Do For Database Applications?” Thursday, October 6, 12:00 Marriott Marquis, Golden Gate 6
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.
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
“Linked Open Data” driven Applications

Data.gov
OpenSource Datasets

Geographic Names

Spatial Data

Census Statistics

Data.gov
OpenSource Datasets

• DoD
• USGS
• Census
...

Data Models and Business Workflows

• Data Integration
• BPMN
• Finance
• Procurement
• Domain vocabularies

Applications

Emergency Response

DoD

State & Local Government

ORACLE®

9 | Copyright © 2011, Oracle and/or its affiliates. All rights reserved.
What Applications Benefit From Semantics?

- **Master Metadata Management**
  - Management of multiple, and inconsistent metadata registries
  - Graphs represent existing metadata models and enable merging of models

- **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

- **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
Semantics with Data Integration: MDM, EII EDI

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
- Standard enterprise metadata model definition

Complements Traditional Data Integration Approaches
- Can support existing MDM, EII metadata repositories
- Supported by key Oracle technologies: Golden Gate (ETL, XML, BPM)
- Supports standard Business Intelligence tools (OBIEE)
Semantic Projects at Lilly: Metadata Repository

Aggregates metadata for experiments from diverse relational databases into a semantic repository for scientific investigation

- Provides a unified vocabulary for scientific investigation
- Implemented faster and provided benefits sooner than the traditional warehousing paradigm
- Allows semantic and relational databases to work together

Sample Queries:

- Identify all interactions for methylases involved in Colon cancer
- Find cell lines in which RNAi data has been generated using Dharmacon reagents
- Retrieve the antibodies that have been used to assess the AKT1 pathway activity in MCF7
- Find all the experiments that were done using my sample
- Find all samples which are grade III colorectal cancer and retrieve the expression, mutation and aCGH data
Lilly: Drug Discovery / Target Assessment
Cisco Knowledge Network

RDF Store

Social Graph

People

Info

Communities

User Entered Tags Import

Entity and Property Extraction

Structured Data

Content Repositories

Enterprise Vocabulary Import

Transactional Applications

e.g. "product catalog"
"Directory"

->

e.g. "web content",
"wikitopics",
"expertise"

e.g. "job roles",
"customer accounts"

Copyright © 2011, Oracle and/or its affiliates. All rights reserved.
Dreamworks Entertainment Repository

- Requires database to scale for millions of movie shot files, thousands of artists
- RDF graph describing a movie shot allows sharing and reuse
- UI uses SPARQL graph pattern query to find movie shots
Customers use tags to find publications of interest

Application tags publications using vocabularies

Vocabularies managed in RDF and terms related using RDFS/OWL
What Does Oracle Bring to this Area?
Semantically Enhanced IT Framework

ANALYTICAL LAYER
Oracle Business Intelligence

INFORMATION EXTRACTION
Golden Gate

Enterprise Data Sources: Structured DBMS, Unstructured, Spatial, RSS, email, Documents
Extraction, Modeling, Reasoning & Discovery

Transform & Edit Tools
- ETL
- NLP Entity Extraction
- Ontology Engineering
- Categorization
- Custom Scripting

Load, Query & Inference
- RDF/OWL Data Management
- SQL & SPARQL Query
- Inferencing
- Semantic Rules
- Scalability & Security
- Semantic Indexing

Applications & Analysis Tools
- OBIEE, Analytics
- Graph Visualization
- Social Network Analysis
- Metadata Registry
- BPM

Oracle & Partner Tools
Oracle 11g
Oracle & Partner Tools
Oracle 11g as RDF Database

- Secure platform for wide-range of semantic applications
- Readily scales to ultra-large repositories
- Choice of SQL or SPARQL query
- Exploits proven technologies – partitioning, clustering, parallelism, security

Key Capabilities:

**Load / Storage**
- Native RDF graph data store
- Manages billions of triples
- Bulk and incremental load

**Query**
- SQL
- SPARQL
- Ontology assisted query of RDBMS data

**Reasoning**
- Forward chaining model
- RDFS++ OWL, OWL 2
- User defined rule base
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
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 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
Customers Deploying Semantic Technologies

Life Sciences
- Lilly
- Pfizer

Defense/Intelligence
- Hutchinson 3G

Education
- The University of Michigan

Clinical Medicine & Research
- Cleveland Clinic

Telecomm & Networking
- Swiss Institute of Bioinformatics
- Cisco

Publishing
- Hutchinson 3G Austria
- Thomson Reuters
## Semantic Technologies Partners
### Integrated Tools and Solution Providers:

<table>
<thead>
<tr>
<th>Ontology Engineering</th>
<th>Reasoners</th>
<th>NLP Entity Extractors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TopQuadrant</strong></td>
<td><strong>clarkparsia, llc</strong></td>
<td><strong>LYMBA</strong></td>
</tr>
<tr>
<td><strong>protégé</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ontoprise</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>know how to use Know-how</strong></td>
<td></td>
<td><strong>the power to answer</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Open Source Frameworks</th>
<th>Standards</th>
<th>SI / Consulting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>openRDF.org</strong></td>
<td><strong>OGC</strong></td>
<td><strong>infoMENTUM</strong></td>
</tr>
<tr>
<td><strong>Jena</strong></td>
<td><strong>W3C</strong></td>
<td><strong>Raytheon</strong></td>
</tr>
<tr>
<td><strong>Joseki</strong></td>
<td>Semantic Web</td>
<td><strong>NORTHROP GRUMMAN</strong></td>
</tr>
<tr>
<td><strong>Sesame</strong></td>
<td><strong>CALAIS</strong></td>
<td><strong>McDonald Bradley</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tools &amp; Applications</th>
<th><strong>SI / Consulting</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PolarLake</strong></td>
<td><strong>Cognia</strong></td>
</tr>
<tr>
<td><strong>TERANODE</strong></td>
<td><strong>accenture</strong></td>
</tr>
<tr>
<td><strong>MONDECA</strong></td>
<td><strong>BOEING</strong></td>
</tr>
<tr>
<td><strong>Tom Sawyer</strong></td>
<td><strong>orbis</strong></td>
</tr>
<tr>
<td><strong>SOFTWARE</strong></td>
<td><strong>TenForce</strong></td>
</tr>
<tr>
<td><strong>Tom Sawyer</strong></td>
<td><strong>The Prognostic Company</strong></td>
</tr>
</tbody>
</table>

**ORACLE®**
For More Information

Oracle RDF

Xavier.Lopez@oracle.com

oracle.com