Use of Semantic Technologies at Eli Lilly and Company

J Phil Brooks
Information Consultant, SE Data Team
Discover IT
Eli Lilly and Company

Lilly

Answers That Matter.
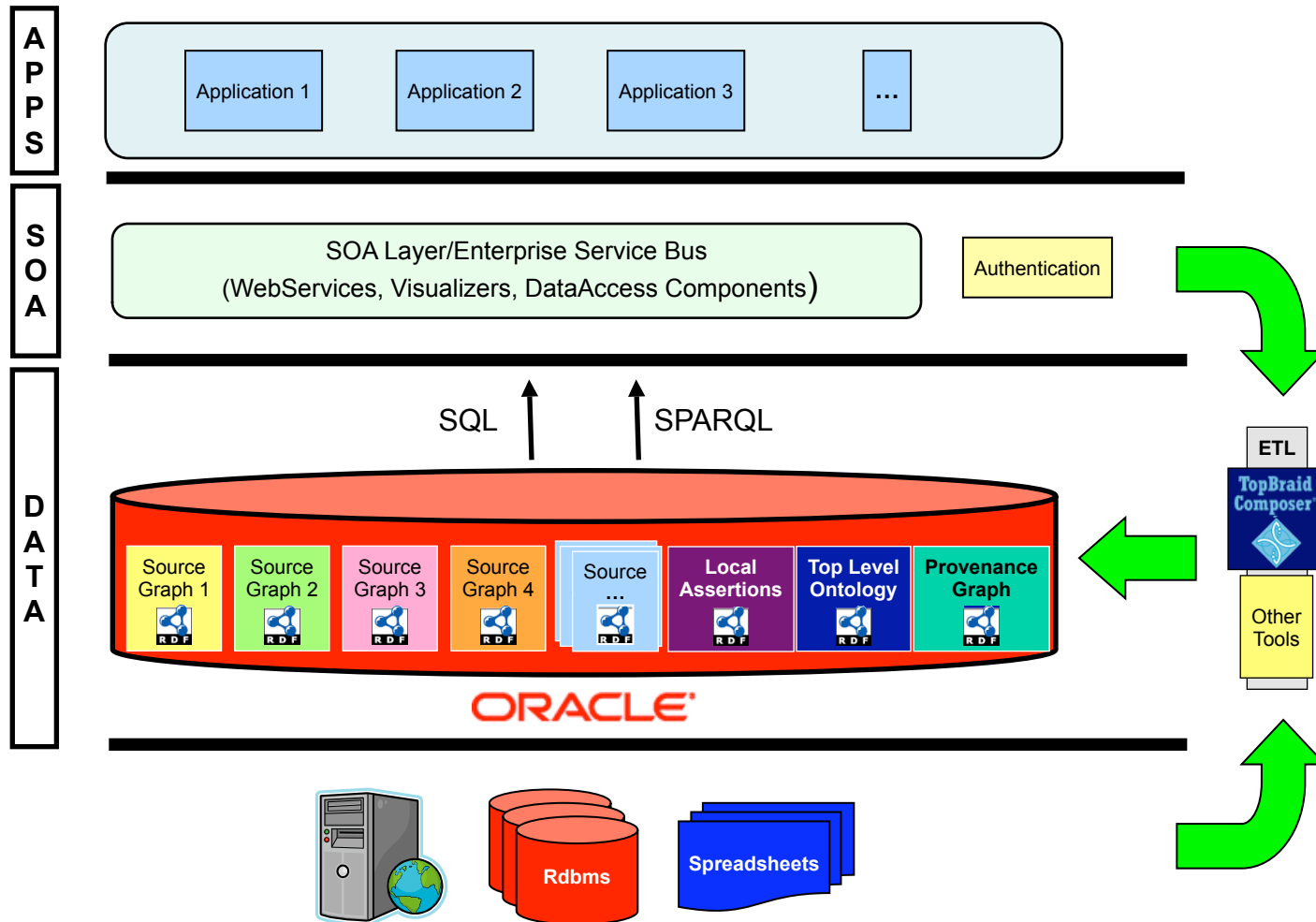
Notable Semantic Projects at Lilly

- Discovery Metadata
  - Integration engine for Discovery Master Data to enable translational research
- CATIE Integration
  - Integration of a large-scale antipsychotic clinical trial with public data sets to identify interactions between treatments, genes, receptors, and pathways
- Competitive Intelligence
  - Harmonization and integration of publicly available data sets for surveying the competitive landscape in the Endocrine disease area
- Experiment Metadata Repository
  - Aggregation and vocabulary synchronization of experiment-related data
- Hypothesis Generator
  - Framework combining Semantic Web technology, graph algorithms, and user profiling to discover and prioritize novel associations among biomedical entities across disciplines

Semantic Projects at Lilly: Discovery Metadata

Integrates Master Data throughout the pharmaceutical discovery process to enable information sharing/ integration for the scientific community

- Models key relationships between Master Data classes
- Provides ability to integrate disparate data sets quicker than the normal warehouse paradigm typically allows
- Utilizes inferencing to create new classifications of data
- Creates a re-usable and sustainable semantic framework
- Enables user-authored manual curation of relationships

Semantic Projects at Lilly: Discovery Metadata - Architecture

Semantic Projects at Lilly: Competitive Intelligence

Provides a mechanism for actively surveying publicly available information for competitive intelligence in the Endocrine area of Lilly's business

- Constructed an ontology of competitor companies, projects, molecules, and disease targets
- Utilized semantic normalization along with other Natural Language Processing (NLP) techniques to reconcile data content
- Utilized inferencing (in conjunction with the company ontology) to normalize companies, their subsidiaries, and other collaborators

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

Why Oracle?

- Longstanding relationship with Oracle
- Significant Oracle expertise in-house
- Mature, centralized support organization
  - Backup/recovery
  - Software/patch installation
  - On-call support
- Ability to experiment with "minimal" investment
  - Oracle databases readily available
  - Dependent RDBMS options already licensed
- Support from Oracle's Semantic Technologies Group has been great!