sandpiper software

# The Model-Driven Semantic Web

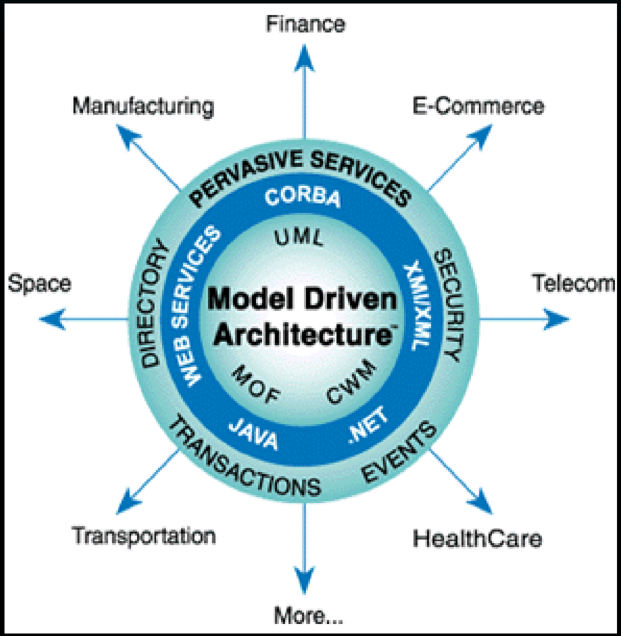
## Emerging Standards & Technologies

Elisa Kendall
Sandpiper Software

March 24, 2005

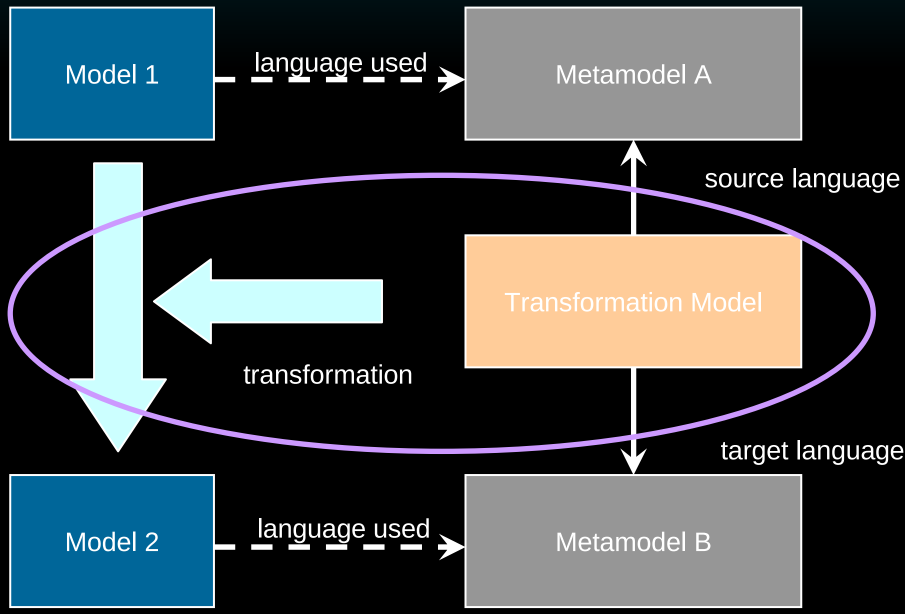
# Model Driven Architecture® (MDA®)

- Insulates business applications from technology evolution, for
  - Increased portability and platform independence
  - Cross-platform interoperability
  - Domain-relevant specificity
- Consists of standards and best practices across a range of software engineering disciplines
  - The Unified Modeling Language (UML®)
  - The Meta-Object Facility (MOF™)
  - The Common Warehouse Metamodel (CWM™)
- MOF defines the metadata architecture for MDA
  - Database schema, UML and ER models, business and manufacturing process models, business rules, API definitions, configuration and deployment descriptors, etc.
  - Supports automation of physical management and integration of enterprise metadata
  - MOF models of metadata are called *metamodels*



# MOF-Based Metadata Management

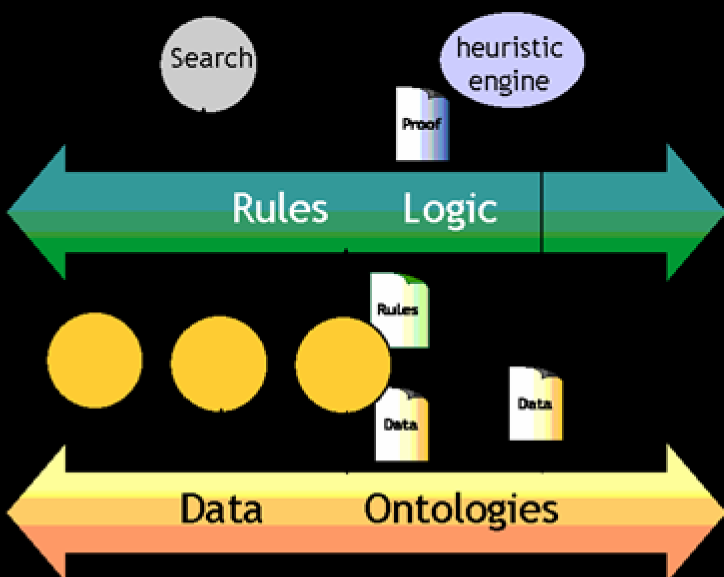
- MOF tools use metamodels to generate code that manages metadata, as XML documents, CORBA objects, Java objects
- Generated code includes access mechanisms, APIs to
  - Read and manipulate
  - Serialize/transform
  - Abstract the details based on access patterns
- Related standards:
  - XML Metadata Interchange (XMI®)
  - CORBA Metadata Interface (CMI)
  - Java Metadata Interface (JMI)
- Metamodels are defined for
  - Relational and hierarchical database modeling
  - Online analytical processing (OLAP)
  - Business process definition, business rules specification
  - XML, UML, and CORBA IDL



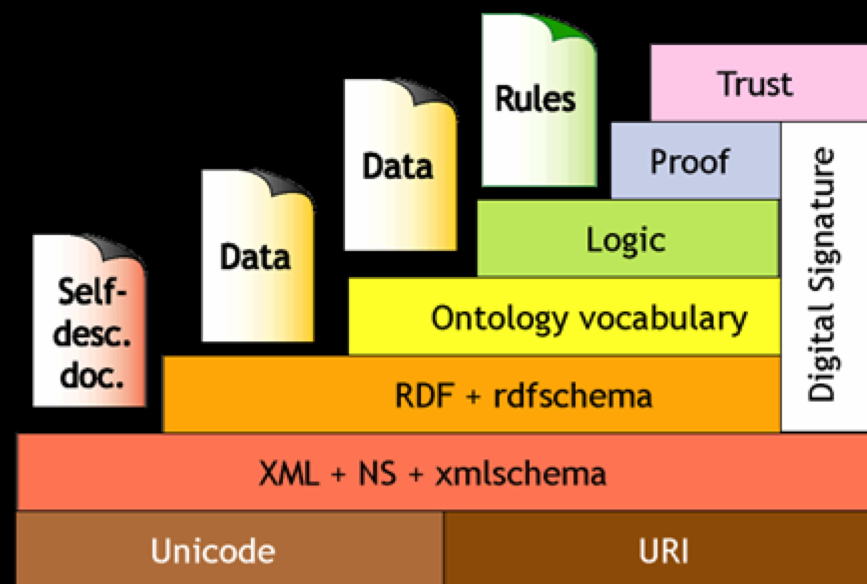
# The Semantic Web

*"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*




*"the bus"*



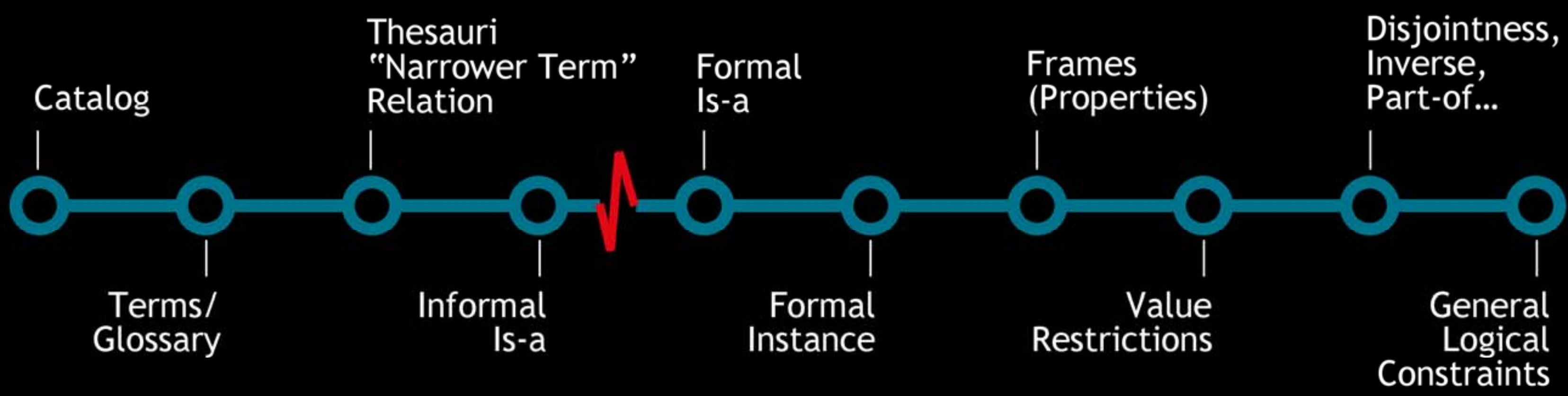

*"the wedding cake"*

# Level Setting

**An ontology** specifies a rich description of the

- Terminology, concepts, nomenclature
- Properties explicitly defining concepts
- Relations among concepts (hierarchical and lattice)
- Rules distinguishing concepts, refining definitions and relations (constraints, restrictions, regular expressions)

relevant to a particular domain or area of interest.




*Based On Aaai '99 Ontologies Panel - Mcguinness, Welty, Ushold, Gruninger, Lehmann

# Ontology-Based Technologies

- Ontologies provide a **common vocabulary** and definition of rules for use by independently developed resources, processes, services
- **Agreements** among companies, organizations sharing common services can be made with regard to their **usage** and the **meaning** of relevant concepts can be **expressed unambiguously**
- By **composing** component ontologies, **mapping** ontologies to one another and **mediating** terminology among participating resources and services, independently developed systems, agents and services can work together to share information and processes consistently, accurately, and completely.
- Ontologies also facilitate conversations among agents to collect, process, fuse, and exchange information.
- Improve search accuracy by enabling contextual search using concept definitions and relations among them instead of/in addition to statistical relevance of keywords.

# Semantic Web Evolution

- Draft specifications for RDF/S & OWL became formal W3C Recommendations in February 2004
- Increasingly attention is on applications and deployment strategies, moving from research to early adoption
- Research from the DAML community and W3C is now focused on
  - Methodology & best practices – representing value spaces, complex relations, engineering methods, applications with richly specified ontologies
  - Query and rule languages
  - Tools and development resources
    - Parsers and Validators
    - Authoring Tools for ontologies and mark-up
    - Ontologies, knowledge bases, and libraries
  - Interaction with SOAP/WSDL to support Semantic Web Services (OWL-S, SWRL)

# Guiding Principles

- Historically, knowledge representation and reasoning systems have operated under *closed world* assumptions
- Uncertainty is magnified under *open-world*, "wild, wild web" conditions, making reasoning more difficult
- Semantic web languages are designed to support less certainty, provide "better" search results, informed answers to questions, not absolutes
- Based on XML, they can assist businesses in leveraging mark-up, content, and data
  - To augment business intelligence/analysis and knowledge mining
  - To support knowledge sharing and collaboration, augment enterprise information integration
  - Enrich web services and other applications
  - Support policy-based applications and ensure compliance with policy

at a lower cost with higher potential ROI than traditional computing methods

# MDA from the KR Perspective

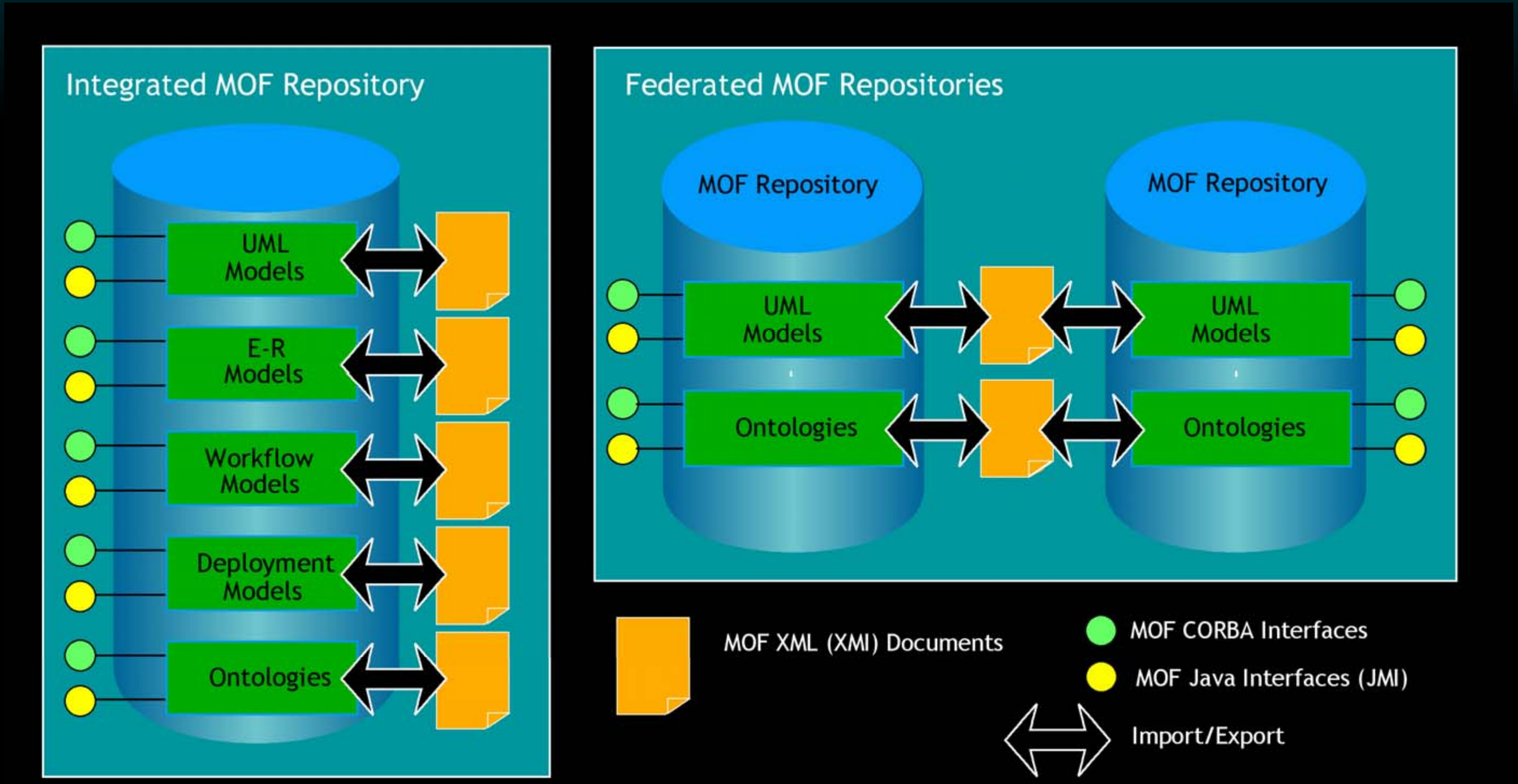
- EAI & EII solutions rely on strict adherence to agreements based on common information models that take weeks or months to build
- Modifications to the interchange agreements are costly and time consuming
- Today, the analysis and reasoning required to align multiple parties' information models has to be done by people
- Machines display only *syntactic* information models and informal text describing the semantics of the models
- Without formal *semantics*, machines cannot aid the alignment process
- Translations from each party's syntactic format to the agreed-upon common format have to be hand-coded by programmers
- MOF® and MDA® provide the basis for automating the syntactic transformations

# MOF and KR Together

- MOF technology streamlines the *mechanics* of managing models as XML documents, Java objects, CORBA objects
- Knowledge Representation supports *reasoning* about resources
  - Supports semantic alignment among differing vocabularies and nomenclatures
  - Enables consistency checking and model validation, business rule analysis
  - Allows us to ask questions over multiple resources that we could not answer previously
  - Enables policy-driven applications to leverage existing knowledge and policies to solve business problems
    - Detect inconsistent financial transactions
    - Support business policy enforcement
    - Facilitate next generation network management and security applications

  while integrating with existing RDBMS and OLAP data stores
- MOF provides no help with reasoning
- KR is not focused on the mechanics of managing models or metadata
- Complementary technologies – despite some overlap

# Metadata Management Scenarios



# Cost of Terminological Issues

- Terminology is one of the most urgent issues in health care today – by some accounts, the second most important problem we are facing
  - Health data is not comparable
  - Health systems cannot interchange data
  - Secondary uses (research, efficiency) are not possible
  - Links to decision support systems are not possible
- In Drugs and Pharmacy
  - Historically proprietary information systems
  - No publicly available ontology; many formularies with overlapping, spotty, encoded knowledge
  - No consistent organizational paradigm
  - No reliable cross-reference between trade names and generic names in machine readable form
- Genomics and Bioinformatics is compounding the problem
  - Inconsistent naming of genes, sequences, SNPs, or proteins
  - Numerous Genomic nomenclatures

*Derived from Presentation by Dr. Chris Chute, Chair of Medical Informatics, Mayo Clinic – At NIST eHealth Conference, June 2003*

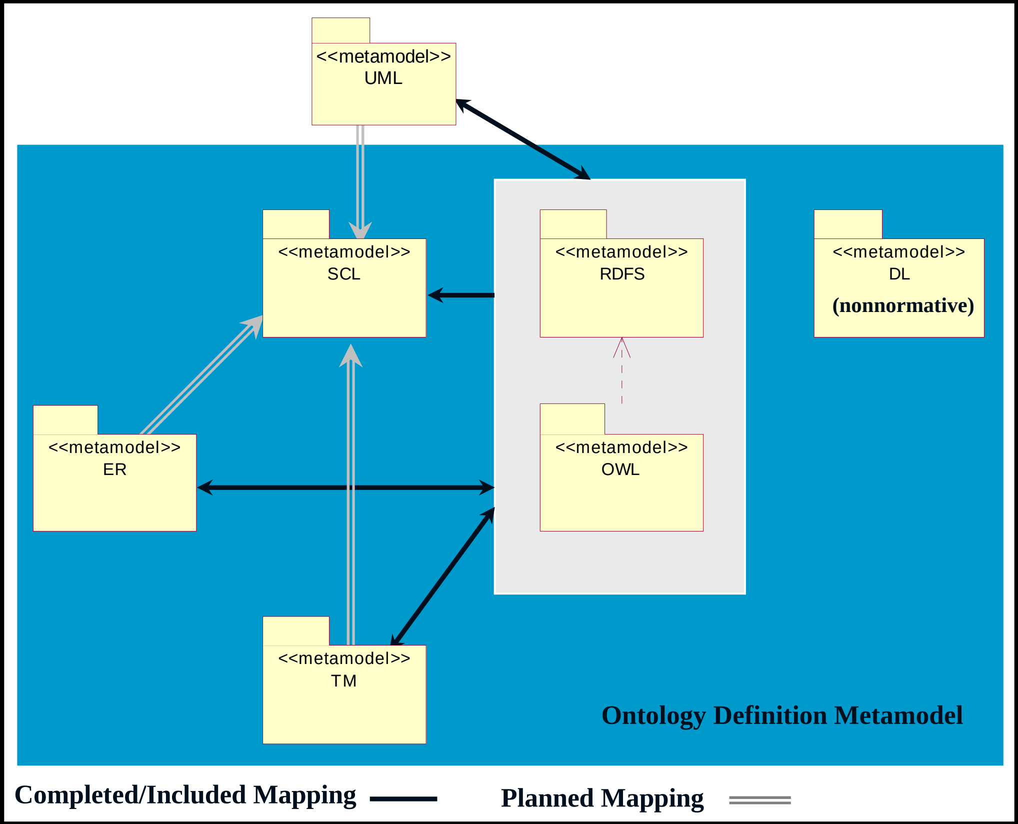
# Cost of Knowledge Capture

- Most tools are oriented towards taxonomy or limited ontology formation vs. knowledge base development, terminology reconciliation, or alignment
- They require significant knowledge of formal logic and domain analysis from an ontological perspective
- Few are graphical and/or standards based
- Vulcan estimates the cost of encoding 50 pages of basic high school chemistry textbook knowledge at *$10,000 per page**
- Tools to assist subject matter experts (SMEs) in encoding knowledge, including increasing automation in developing the starting points for ontology and KB development are essential
- Combining MOF® and MDA® technologies with KR dramatically reduces cost, increases likelihood of success, increases availability to a much broader community of potential users

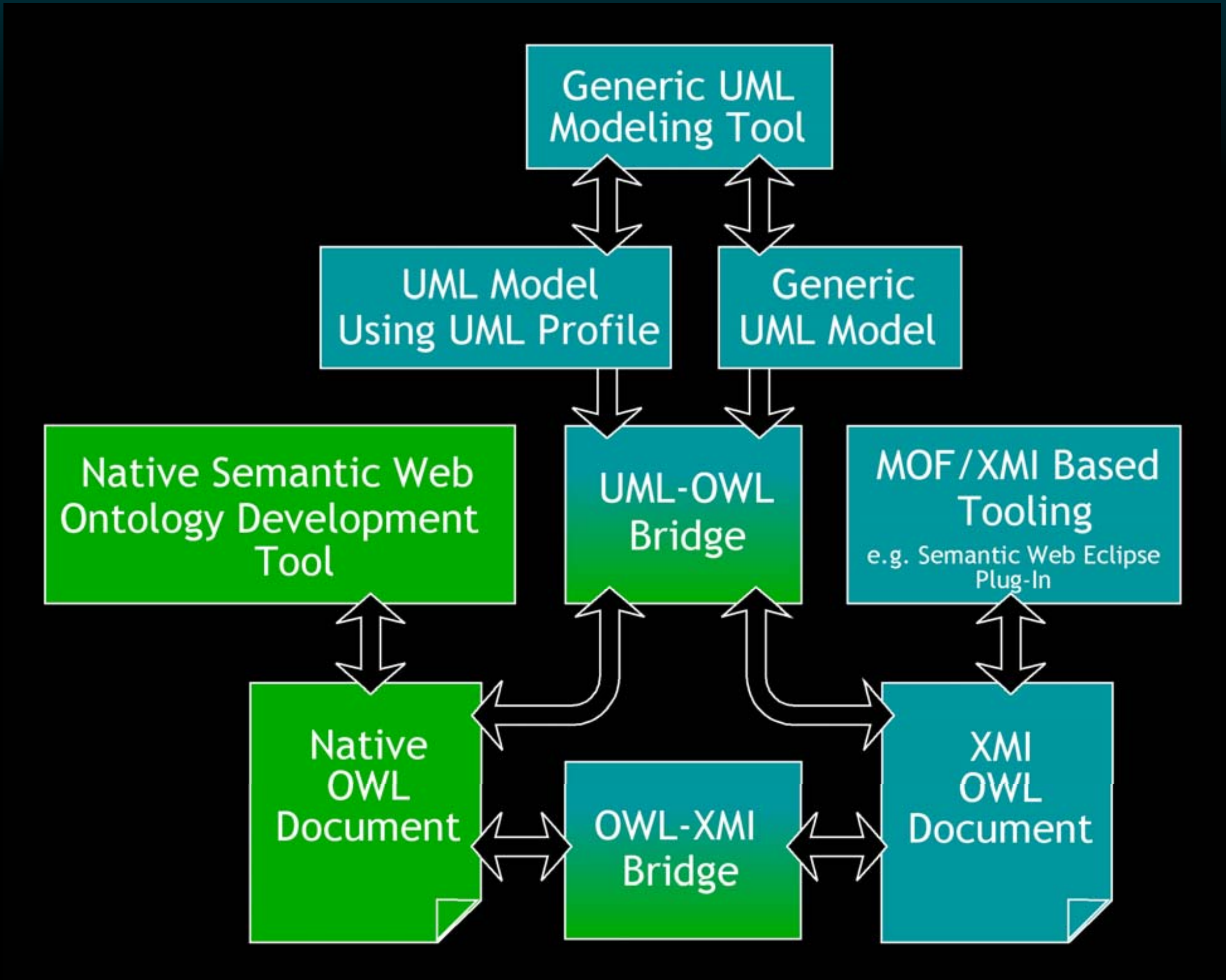
** See Vulcan, Inc. – Project Halo, at http://www.projecthalo.com/*

# Towards a Model Driven Semantic Web – ODM

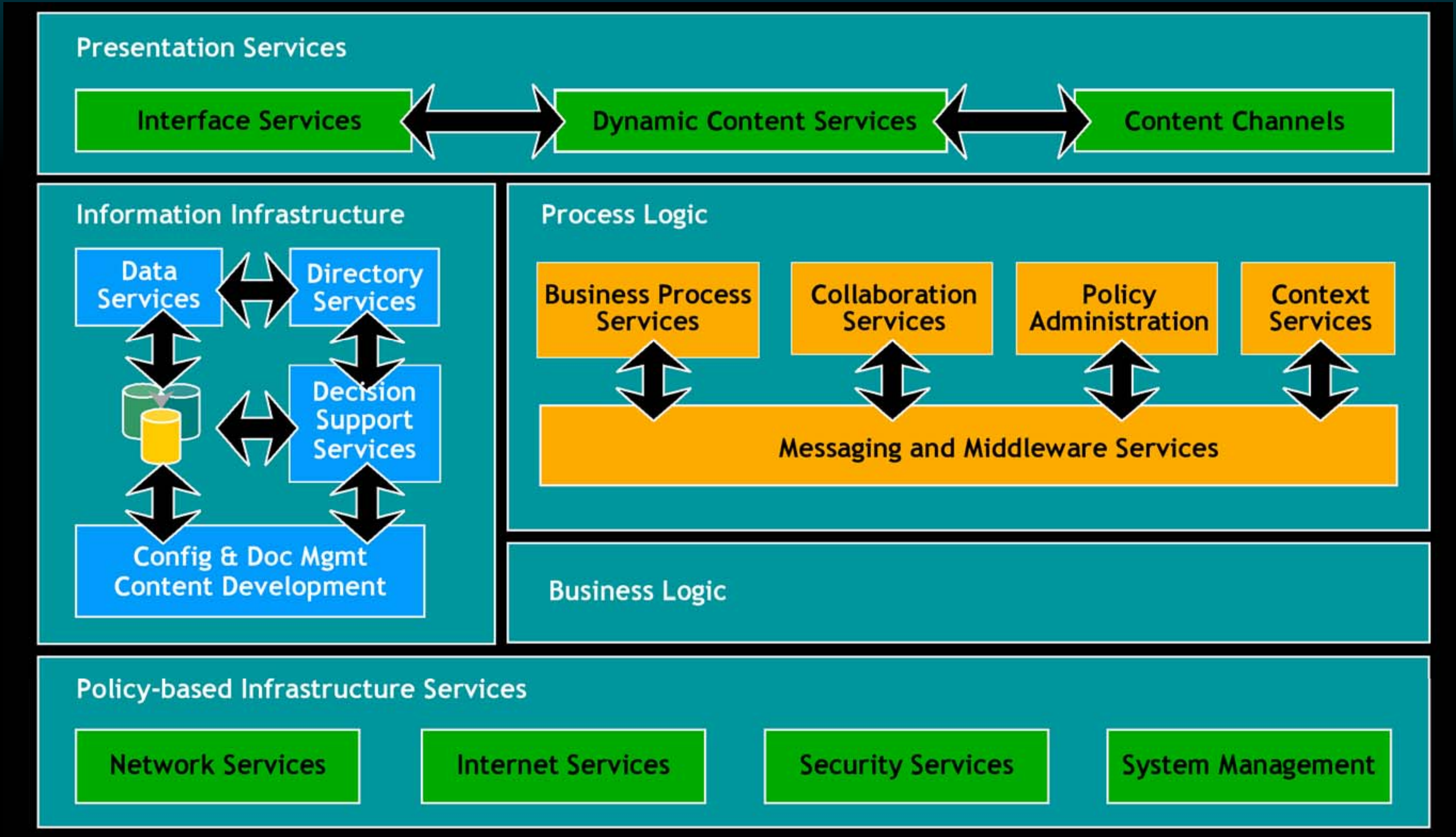
- Six EMOF platform independent metamodels, (PIMs), five normative
- Mappings (currently tables, QVT planned)
- UML2 Profiles
  - RDFS & OWL
  - TM
  - SCL planned
- Collateral
  - XMI
  - Java APIs
  - Tool support
- Conformance
  - RDFS & OWL
  - All else optional



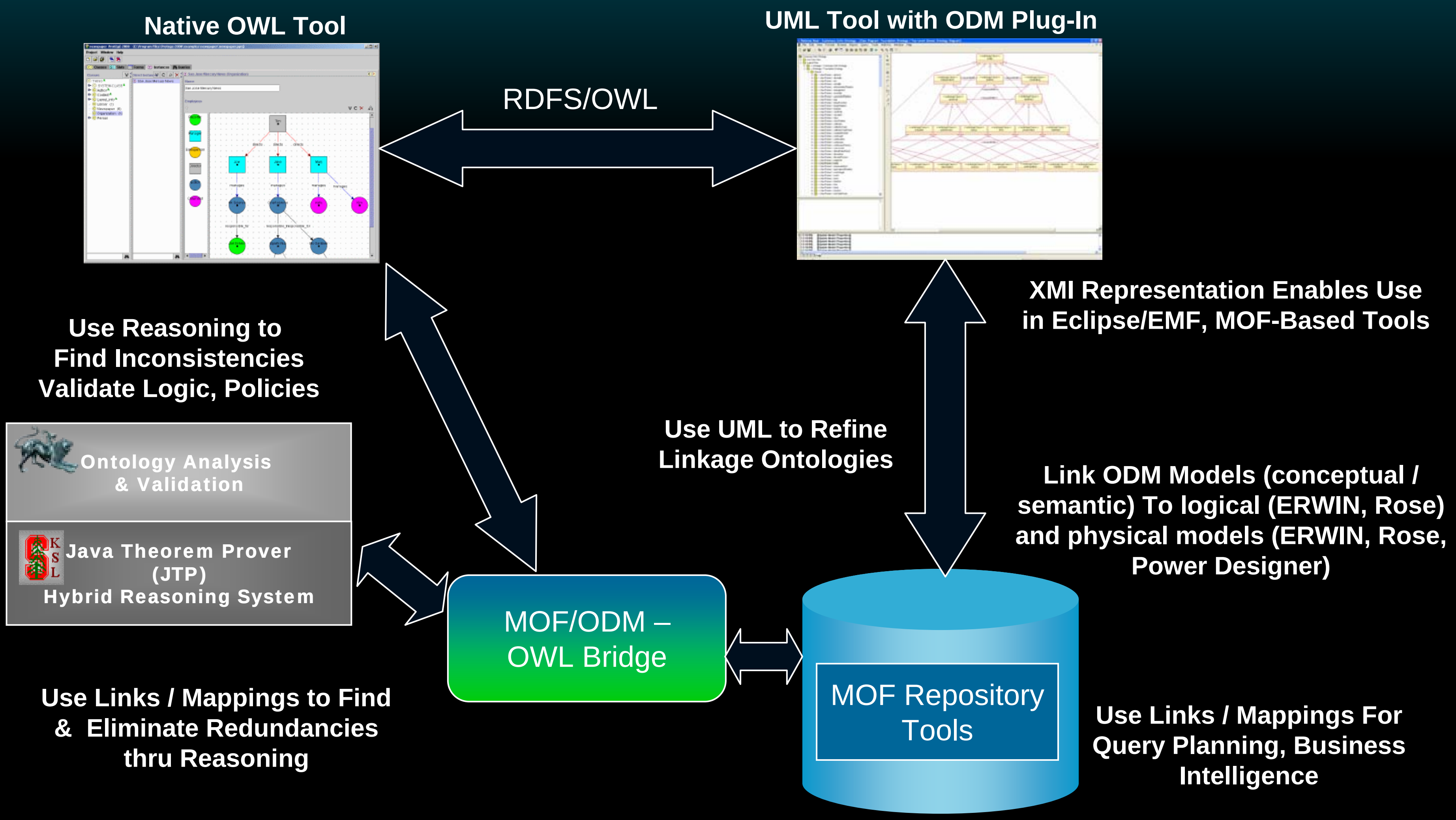
# Bridging KR and MDA



# Technology Architecture



# Implementation Strategies



# Business Integration

- Semantic Web Services standards are converging (OWL-S and SWSL)
- OMG RFC/RFPs forthcoming for extensions to ODM to support
  - Additional mappings and profile for Common Logic (CL)
  - Semantic Web Services
  - ISO/EXPRESS
  - Emerging Semantic Web Rule languages
- Leverage mapping from UML for BPEL to ODM extensions
- Strategy:
  - Link business process models through MOF environment
  - Generate OWL for the linkage
  - Use linkage as basis for mediating business process semantics

# A Framework for Next Generation Interoperability

- MOF's model management facilities and KR capabilities for machine interpretable semantics and reasoning are separate, complementary concerns
- The ability of reasoners to find discrepancies in invariant rules, preconditions, and post conditions, can add scalability to MDA's use of Design-by-Contract (DBC)
- UML profiles can serve as graphical notations for Semantic Web languages, dramatically increasing ease of use
- The combination of MDA and SW technologies promises to
  - Address the missing link in business process automation
  - Enable true information interoperability and continuity
  - Support next generation policy-based applications development

# The Model-Driven Semantic Web

- Knowledge acquisition, developing the semantics is the bottleneck
- Leveraging existing assets breaks that bottleneck
- Correlation through reasoning provides the utility
  - Multi-dimensional, cross organizational tailored semantic views
  - “Virtual” repository approach enables elimination of redundancy
  - Reasoning supports quality initiatives through inconsistency discovery, model and content validation
- MDA and MOF coupled with Semantic Web technologies are the key