



Center for Advanced Studies of Rome

# Thinking Ontologies at IBM Center for Advanced Studies, Rome

Guido Vetere  
IBM CAS Rome



**IBM Centers for Advanced Studies**

**Centro Studi Avanzati di Roma**

TONES IAT Workshop  
Edinburgh, 27/05/06

## Agenda

- **Ontologies at IBM**
- **Ontologies at IBM Center for Advanced Studies of Rome**
  - Current works
  - Future works
- **Main industrial topics (in my view)**

# Ontologies at IBM

- **Semantics on AlphaWorks ([www.alphaworks.ibm.com/topics/semantics](http://www.alphaworks.ibm.com/topics/semantics))**
  - **IBM Web Ontology Manager:** A Web-based system for managing Web Ontology Language (OWL) ontologies.
  - **Semantic Tools for Web Services:** A set of Eclipse plug-ins for semantic matching and composition of Web services.
  - **Multimedia Analysis and Retrieval Engine:** An automated desktop indexing and search system for digital images.
  - **Web Metadata Extractor** - A tool for extracting relevant information from well-formed Web pages by using semi-automatically generated templates.
  - **Unstructured Information Management Architecture SDK:** A Java SDK that supports the implementation, composition, and deployment of applications working with unstructured information.
  - **IBM Integrated Ontology Development Toolkit:** A toolkit designed for storage, manipulation, query, and inference of ontologies and corresponding instances.
  - **Ontology-based Web Services for Business Integration:** A semantic Web services proof-of-concept demonstration for the industrial sector that shows service discovery, composition, and business process transformation.
  - **ETTK for Web Services and Autonomic Computing:** A software development kit for designing, developing, and executing emerging autonomic technologies and Web services.
  - **Agent Building and Learning Environment:** A Java framework, component library, and productivity tool kit for building intelligent agents using machine learning and reasoning.
  - **ToXgene:** A template-based generator for complex, semantically-correlated collections of synthetic XML documents.

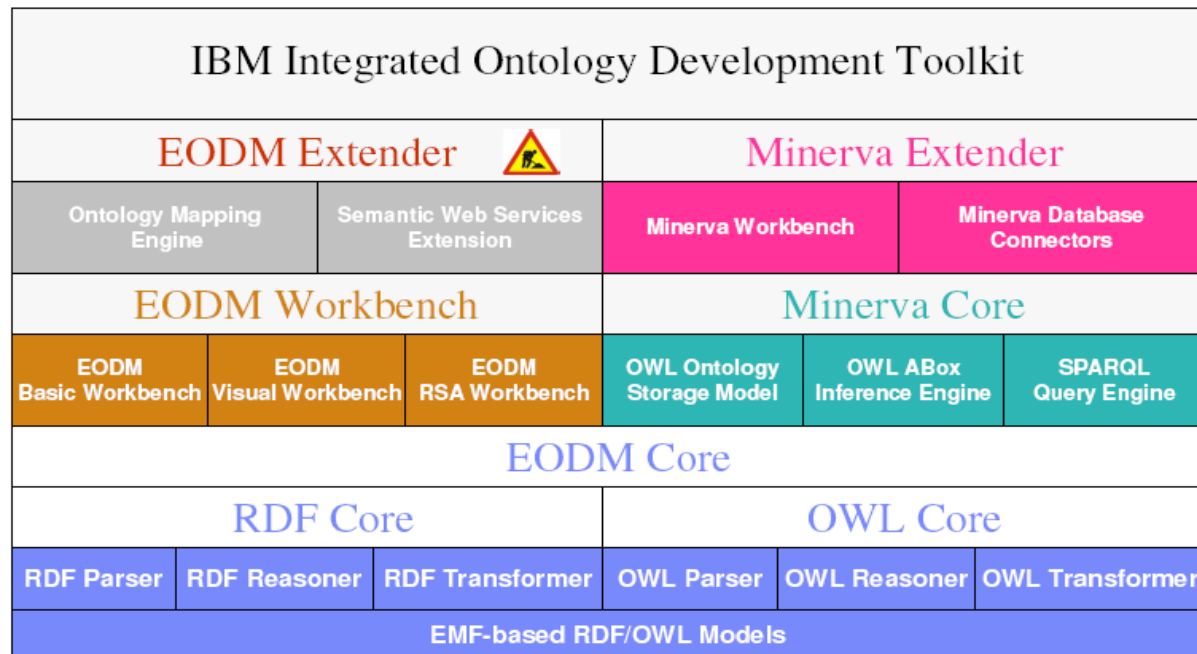
## Ontologies at IBM (2)

- **IBM Internal Open Source Bazaar**
  - Semantic Layer Research Platform
    - A platform for building Semantic applications that use RDF, LSID and other Semantic Web technologies. The platform includes several components such as an RDF server with collections, acls, replication and transactions, client and web development kits including an Eclipse suite of plugins for RDF consumption. This project also contains several prototypes such as Life Sciences Workbench and Unified Activity Management.
  - Semantic Super Computing
    - The Semantic Super Computing project performs research on large amount of data or large amount of computation. The aim is to explore innovative way to deal with the scalability issues and user interface problems of such projects.
- **Other relevant research projects**
  - SHER (Scalable Highly Expressive Reasoner) - China Lab, Watson
    - Semantic retrieval using ontologies
    - Uses OWL DL over RDF secondary storage

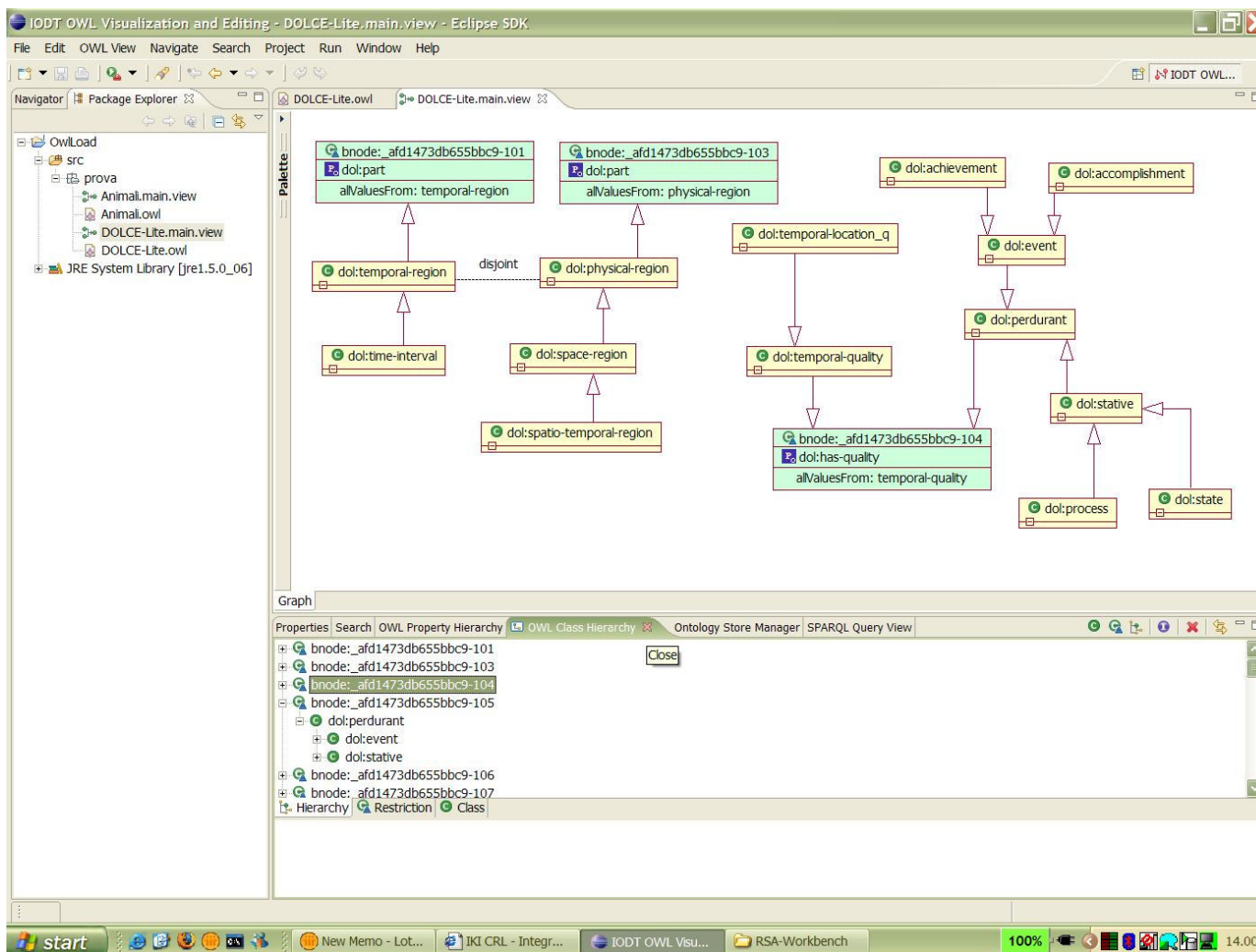
## Integrated Ontology Development Toolkit (China Labs)

### ■ Integrated Ontology Development Toolkit

- <http://www.alphaworks.ibm.com/tech/semanticstk>
- Both a core technology for managing semantic information and an API and toolkit for developing and using the technology
- Based on ODM (Ontology Definition Metamodel, OMG) and EMF (Eclipse Modeling Framework (MOF))



# IODT Workbench (Eclipse)



# IBM Center for Advanced Studies of Rome

- **IBM CAS Network ([www.ibm.com/ibm/cas](http://www.ibm.com/ibm/cas))**
  - Mission: Innovation through collaboration
    - Facilitate the exchange of academic research knowledge and real world industry challenges towards enhancing IBM products, processes & services
    - Establish IBM as the partner and employer of choice for top students as they learn and develop skills to create the technology of the future
    - Build and foster relationships among researchers, funding agencies, IBM, and customers
    - Expose IBM developers to current research directions, and identify new and emerging technology issues for academic research
- **IBM CAS Rome**
  - Mission: Innovation through collaboration with Italian research
    - Services and Data Integration
    - Autonomic Computing
    - Information Insight
  - Partners
    - University of Rome (I-III)
    - University of Pisa
    - CNR

# IBM CAS Rome, Current Works

Peer to Peer Data Integration (partnership with University of Rome 'La Sapienza' - DIS)

- **Projects**

- Hyper (IBM SUR GRANT 2003, 2005) – P2P integrated Data Grid (OGSA-DAI)
- WORKPAD (STREP 6° FP) – “Extreme Integration” to support emergency scenarios

- **Architecture**

- Peers are independent information sources with integration capabilities
- Each peer exposes its ontology
- Local mappings specify how data sources that are local to the peer are linked to the peer ontology
- Peer-to-peer mappings specify how views over one peer ontology are related to views over other peer ontologies (GLAV)
- Each peer specifies its mappings to other peers autonomously

- **Semantics**

- No commitments to any global ontology (relativism)
- Mapping is viewed as *translation* (vs. *intepretation*)
- Query answering is given an epistemic semantics (certain answers are those true in all “possible worlds”)

- **Issues**

- Support a variety of ontology languages
- Support individual mapping
- Encode and evaluate “propositional attitudes”
- Perform distributed query processing
- Tune up epistemic semantics



## IBM CAS Rome, Future Works

- **Italian e-Government Ontologies**
  - Nationwide metadata standardization initiative (CNIPA – Assinform)
    - Set up the appropriate metalevel
    - Identify (or develop) a core ontology
    - Identify specific domains (taxes, social insurance, etc)
    - Design a cooperative development workflow
  - Ontology Management System for ‘Sistema Pubblico di Connettività e Cooperazione Applicativa’ (Public Connectivity and Cooperation System)
    - Centralized ontology repository for Italian governing bodies
    - Should be able to
      - Store, check, and retrieve (part of) ontologies
      - Support a variety of syntactic renderings (RDFS, OWL, XSD, XMI,...)
      - Browse, edit, map, allow cooperative authoring
      - Support lexicalization
    - Possible developments
      - Support complex intensional queries (e.g. relatedness)
      - Generate pairwise transformation rules through mappings

## IBM CAS Rome, Future works

- **Computational Cooperative Lexicon of Italian Language (Senso Comune)**
  - Joint open initiative of University of Roma “La Sapienza” (Computer Science, Linguistics), CNR (Laboratory for Applied Ontology), and IBM Italia, hosted by Fondazione IBM Italia.
  - Build a Linguistic Ontology for Italian à la Wiki
  - DL (Lite) to encode “semantic frames” (Fillmore)
  - Mapping with similar resources (WordNet)
  - Easy and yet formal user interface (but how?)
  - Reasoning repository of 100-500K lexicalized concepts

## Main industrial topics

- **Model Driven Development**
  - Reasoning to enhance model-driven development
  - Main issues
    - Clarify and extend formal semantics of modelling languages
    - Make models explicit and consistent
    - Support model transformations
  - Directions
    - Extend UML and supply formal semantics (e.g. ODM)
    - Integrate reasoning services through standard interfaces (e.g. DIG)
    - Support rules

## Main industrial topics (2)

### ▪ Information Integration

- Reasoning to integrate information coming from heterogeneous, independent, and distributed sources
- Main issues
  - Efficient (logspace) distributed query answering
  - ABoxes on (legacy) DBMS
  - At Web scale (Semantic Web)
    - Position wrt different ‘philosophic flavors’ (realism? anti-realism?)
    - Deal with conflicting ‘possible worlds’
    - Formalize ‘attitudes’ such as *trust* and *understanding*
    - Deal with uncontrolled naming policies
    - Deal with low data quality
    - ...
- Directions
  - Set the right expressiveness (e.g. DL-Lite)
  - Focus on mapping (theory, languages, and tools)
  - Drive the adoption of naming conventions and high quality ontologies

## Main industrial topics (3)

### ■ **Information Extraction**

- Reasoning to extract meaning from linguistic sources
- Main issues
  - Named entities recognition
  - Disambiguation
  - Frame recognition
- Directions
  - Efficient use of huge linguistic ontologies
  - Support connectivity-related queries (e.g. role paths)
  - Support abductive reasoning

Thanks for the attention

- Q&A



Guido Vetere, [gvetere@it.ibm.com](mailto:gvetere@it.ibm.com)