Interaction of Ontology and Database in the Information System

T. Malyuta, Ph.D.
Computer Systems Technology Department
New York City College of Technology
City University of New York

May 10, 2007

Topics

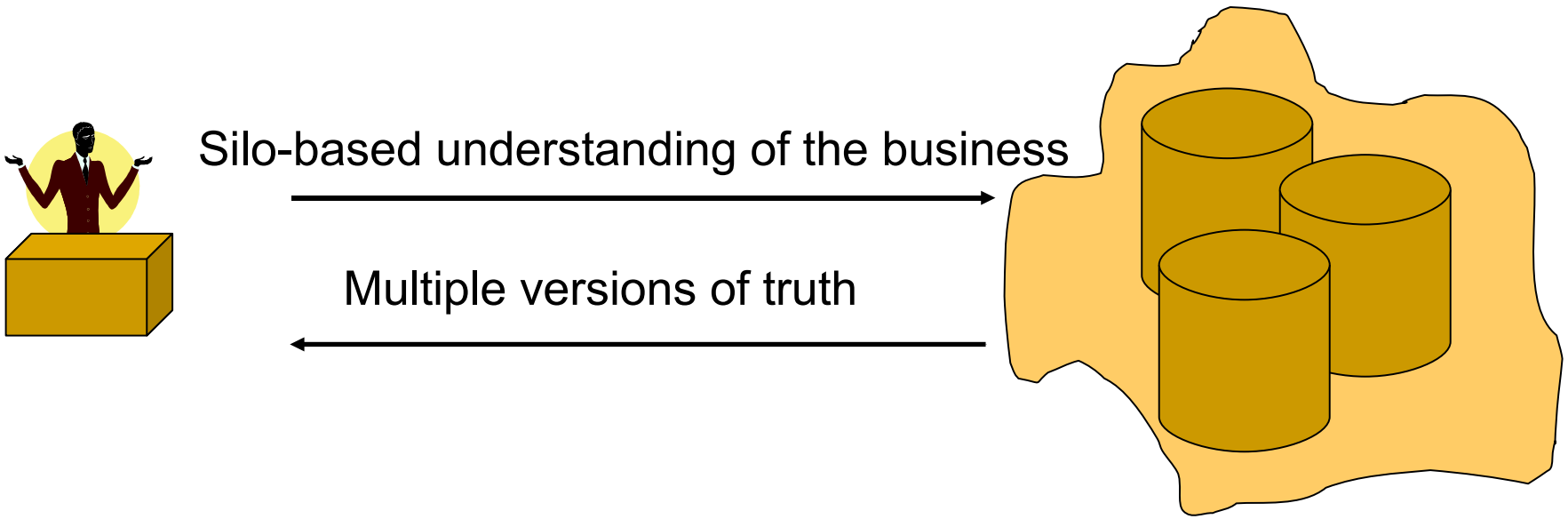
- Information System (IS) challenges and the expected benefits of Ontology
- Ontology and HCI
- Ontology and data utilization
- Symbiotic relationship between Ontology and IS
- Interaction between Ontology and Database
- Database vendors and semantic initiatives

IS Challenges

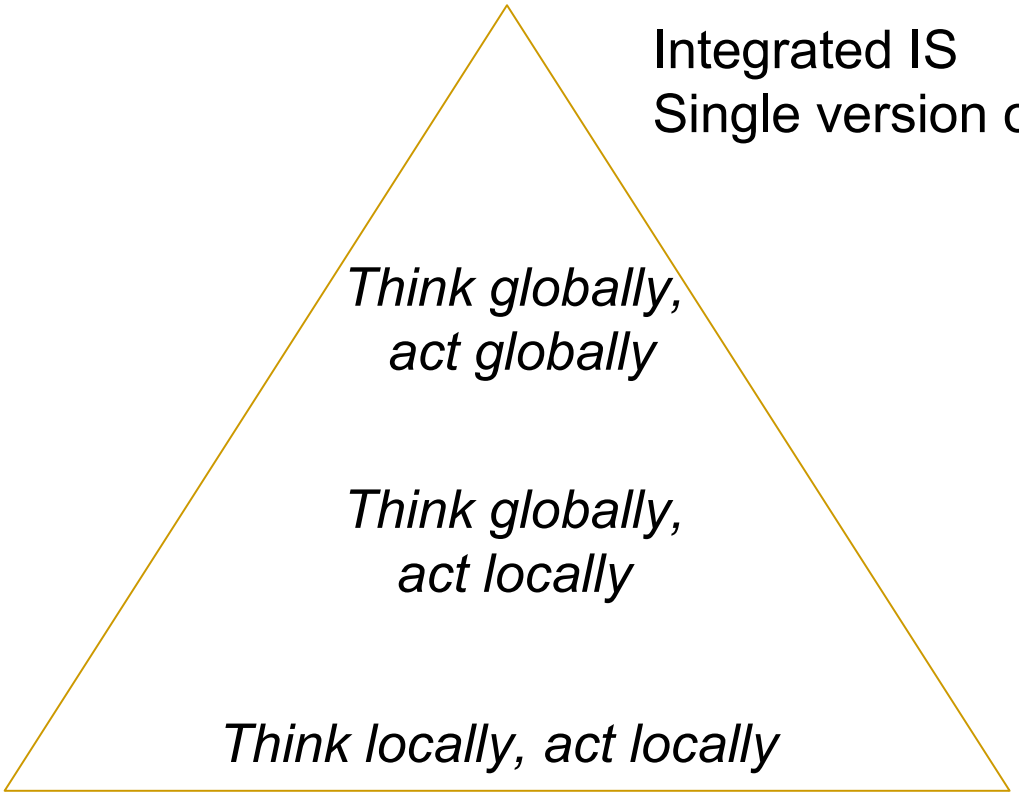
- Sustainability of data solutions in changing business (needs, interfaces, data structures, results) and technological environments
- Shareability, interoperability, and reusability of data within a heterogeneous IS with partial business solutions
- Scalability under the pressure of growing volumes of data and increasing diversity of users' needs
- Complexity and reactive approach to solving the problems

Inability to Fulfill the Ultimate Purpose

Despite growing cost, the IS does not provide users with adequate understanding of the business or single version of truth.



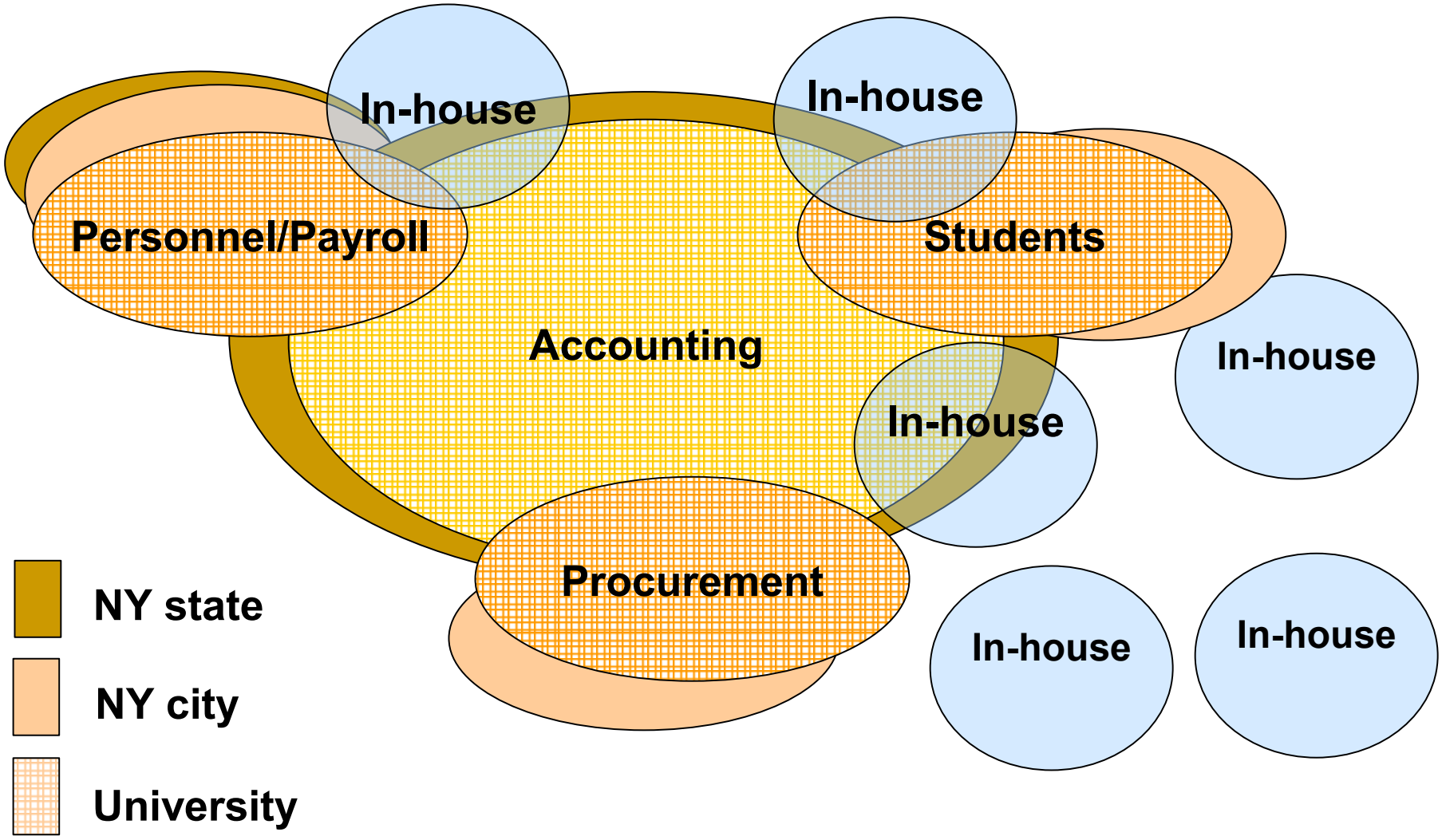
Problems Stem From Approaches



Integrated IS
Single version of truth

Heterogeneous enterprise IS consisting of siloed parts

Heterogeneous IS of NYCCT



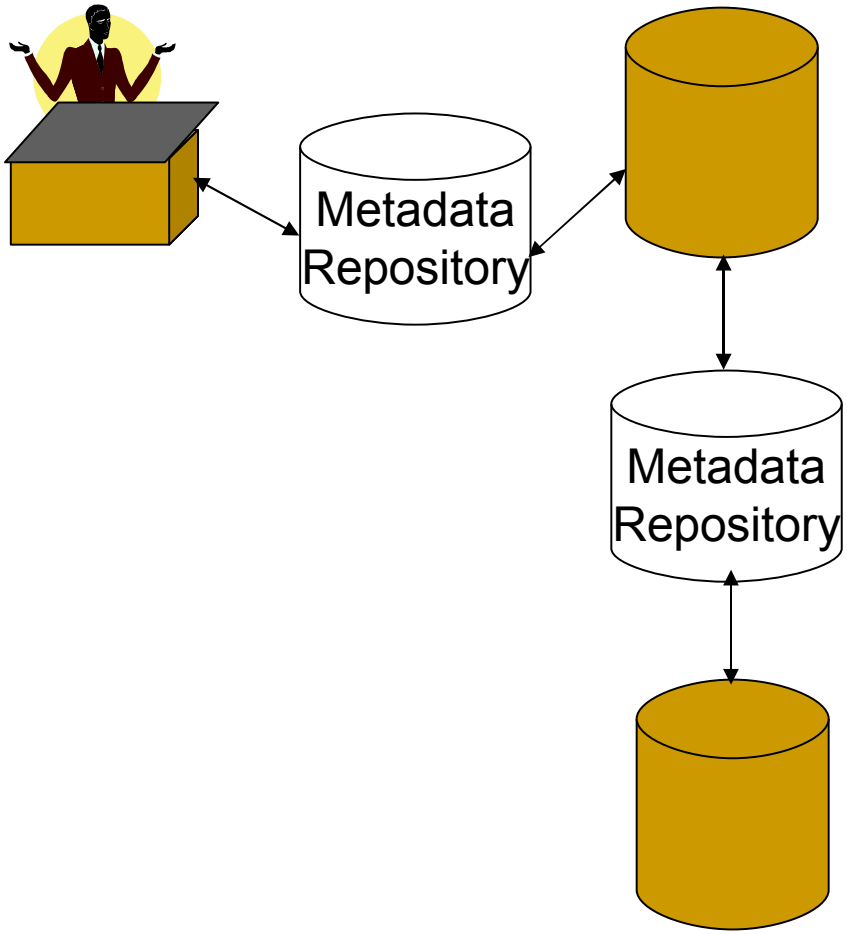
Resolving Problems of HCI and Data Utilization

Search for approaches for :

- Description
- Discovery
- Negotiation
- Mediation
- Composition

of the IS resources that will benefit People, Tools and Applications in the heterogeneous and ambiguous environment

Metadata

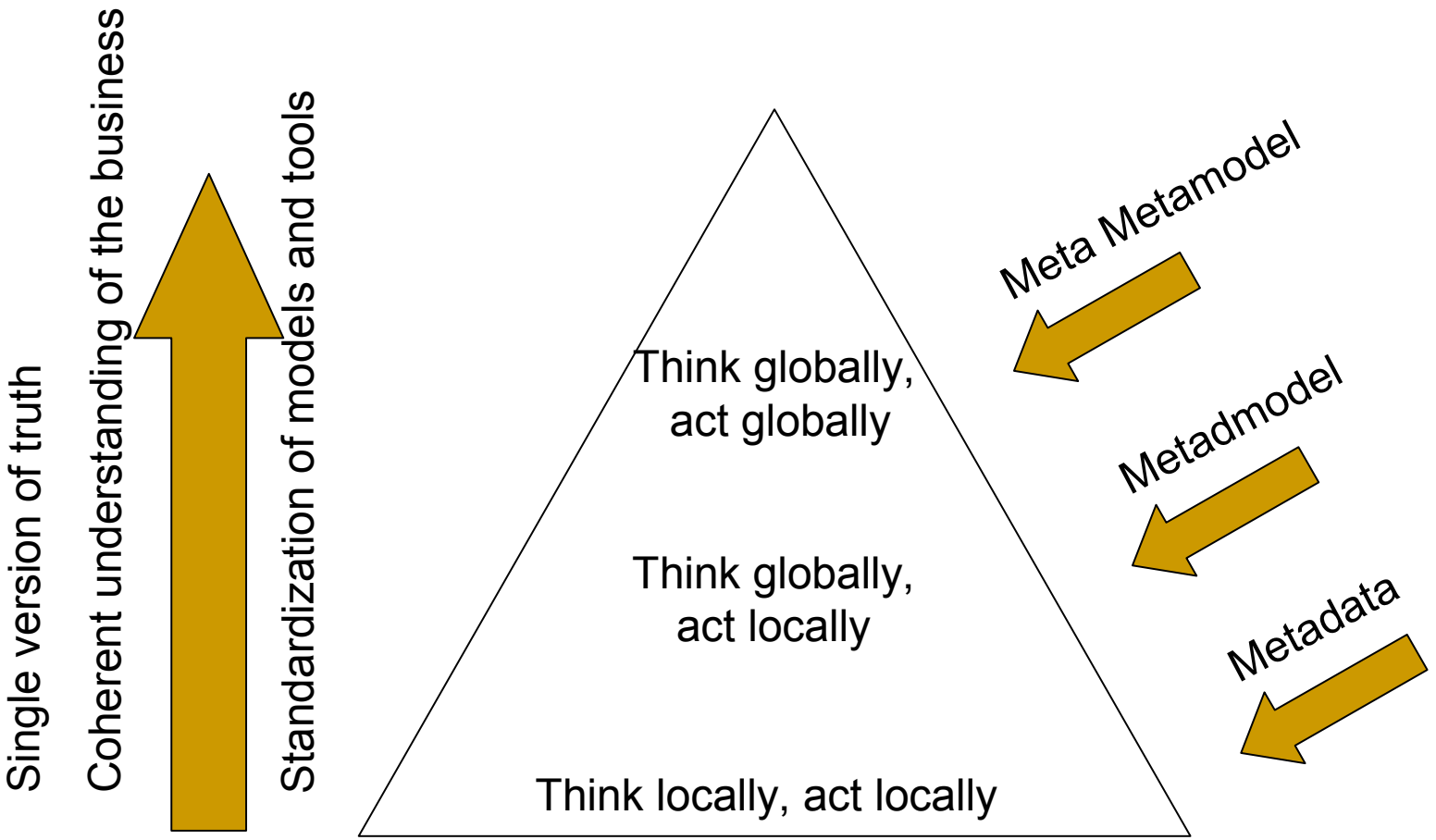


Non-standardized approach leads in metadata to what we tried to escape:

- Duplication
- Ambiguity
- Inconsistency
- Redundancy

Such approach is still local thinking and acting

Towards Global Enterprise Thinking



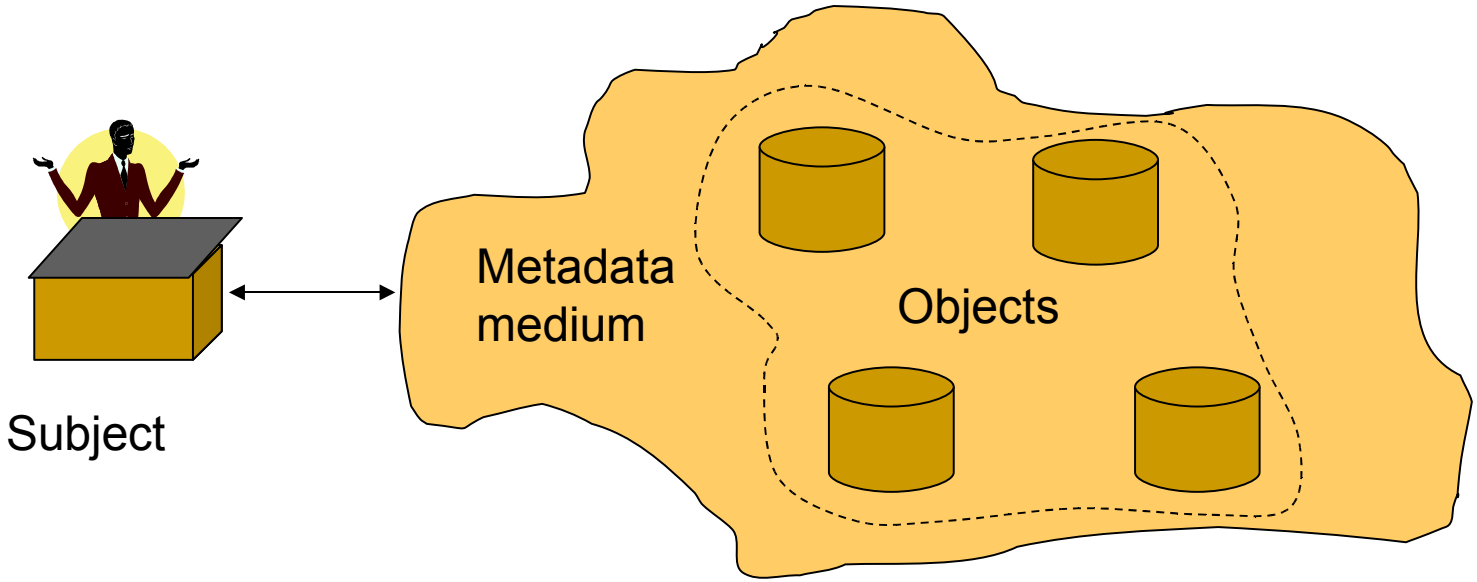
Metadata and Models Standardization Initiatives

- Object Management Group (OMG)—modeling standards, including:
  - Common Warehouse Model describes metadata interchange among data warehousing, business intelligence, knowledge management and portal technologies.
  - Meta-Object Facility bridges the gap between dissimilar meta-models by providing a common basis for meta-models.
- SOA--Services Repository, UDDI.

M. West “From Ontology to Data Model”, 11-16-06.
Reference Data Model, ISO-15926-2 data model.

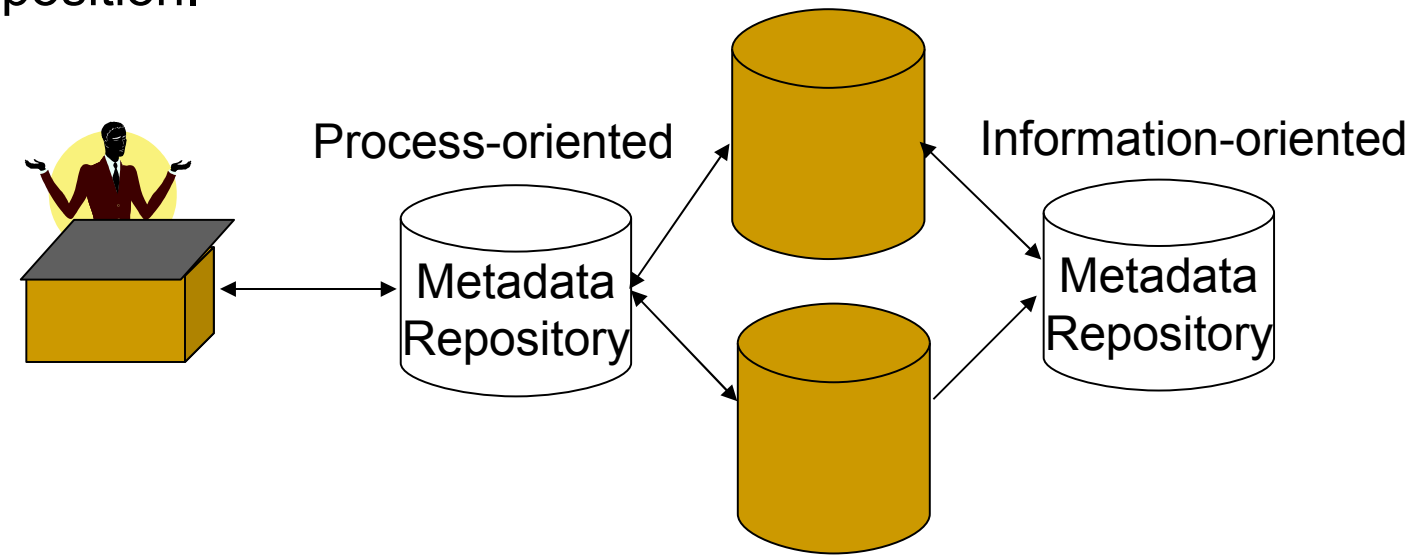
Data and Metadata in IS

Ideally, databases—traditionally the object of data processing—are immersed in the metadata medium that complies with the standards, and defines and enforces the subject of data processing. This medium is still the object of processing for the human/external application.



Metadata for HCI and Data Utilization

This metadata medium supports HCI and inter-database communications, including data integration, by providing descriptions of data and relationships, and the rules of data access and processing. It explicitly represents users' understanding of IS functionality and data, and, therefore, enforces HCI and data utilization, including data discovery, mediation, and composition.



Role and Goals of Ontology

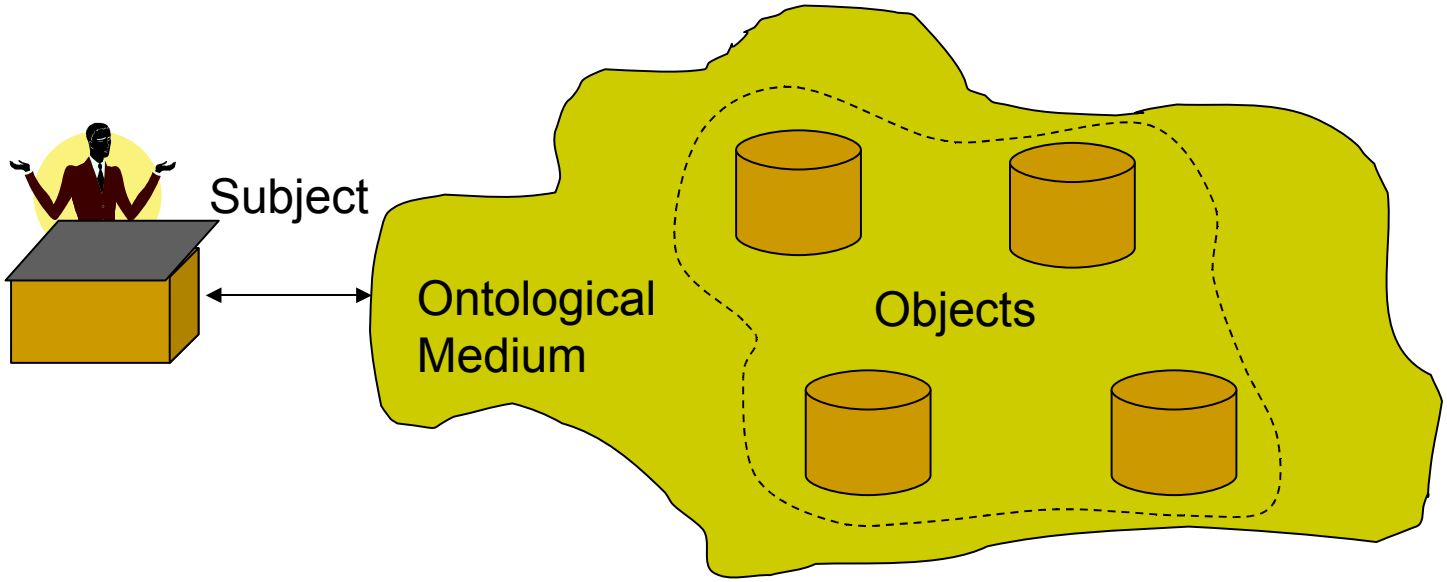
As a resource, ontologies are:

- About global thinking and understanding of the business, and global, general, technology independent presentation of this understanding, which intended to be shared.
- Relatively independent of particular applications, consisting of relatively generic knowledge that can be reused by different kinds of applications/tasks. They are not task-oriented and implementation-dependent.

The ontological data model and supporting technologies enforce the role and goals

Ontology vs. Metadata

Ontology can be this data medium in which data and other resources exist, and which describes resources, their relationships, and functionality. The ontological data model suits much better to this role than existing metadata models.



Advantages of Ontological Medium (OM)

- Publishing models as Web resources—global thinking and acting *beyond* the enterprise
- Querying the models
- Resiliency to changes
- Flexibility of integration
- Application of inference and rules

Expected Benefits of OM

Bringing to a new level benefits of metadata medium with less human involvement:

- Integrating IS into a comprehensive and continuous environment
- Providing a single version of truth for business decisions
- Reducing reliance upon documentation
- Reducing maintenance cost
- Reducing duplication of effort
- Improving efficiency of IS resources utilization, including data processing and retrieval

Pre-requisites of Ontology Success in IS

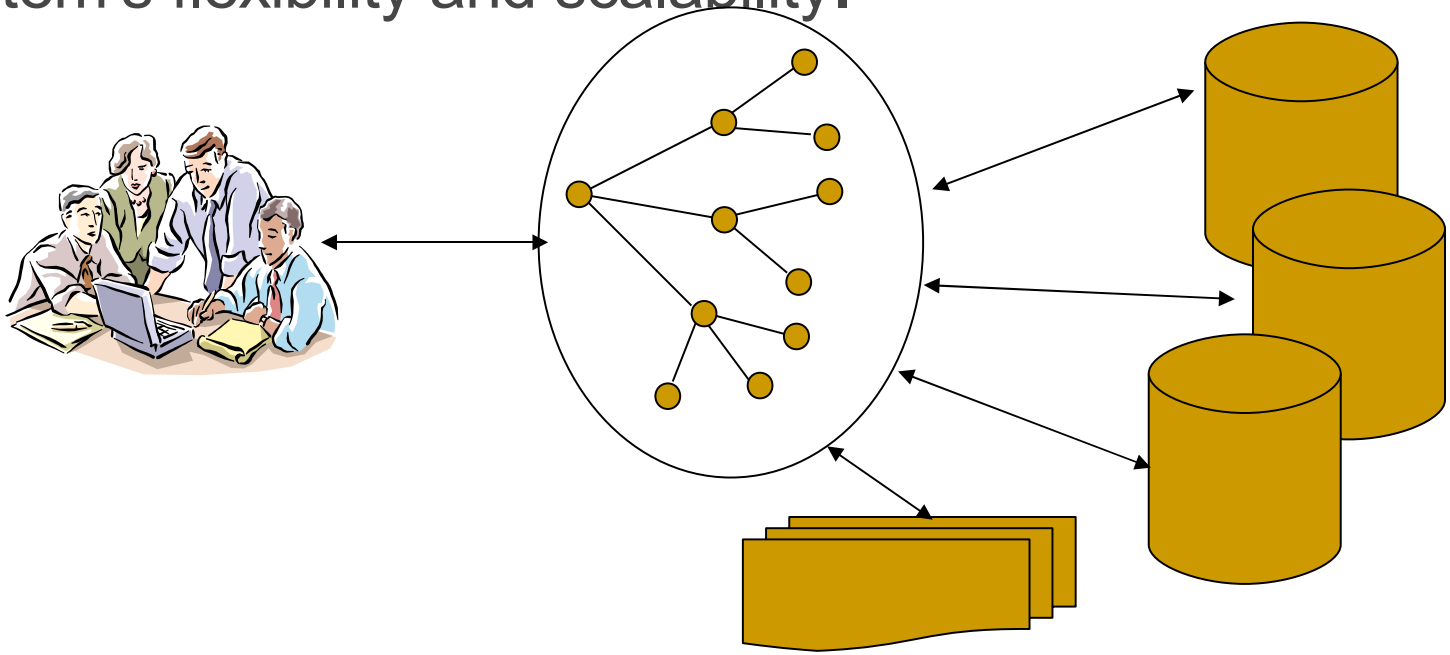
Utilizing data and metadata models

- OMG “Ontologies can’t ignore this wealth of knowledge, tools, expertise and industry momentum.” (about data and metadata models)
- M. West “Data Models are a key component for getting ontologies used in Information Systems.”

E. Kendall “The Ontology Definition Metamodel”, 01-04-07.
OMG Initiative on bridging metamodels and ontology.

Ontology in the Heart of Knowledge-Centric Organization

Enterprise Ontology can realize the accumulated potential of global thinking by reducing the cost and improving the quality of utilization and management of information assets, and system's flexibility and scalability.



Ontology and HCI—Beneficiaries

Beneficiaries of IS Ontology:

- Insiders—designers, developers, technical support.
- Outsiders—users of the IS, consumers of its resources.

Ontology and HCI—Insiders

Insiders:

- Unified, integrated, consistent, IS-embedded presentation of knowledge about the functionality, tools, parts, etc.
- Guidance and automated support for design, development, and maintenance of IS resources through the IS life cycle
- Standardized, meaningful, machine-readable interfaces designed, implemented, and sustained in an automated way

Ontology and HCI—Outsiders

- Description of functionality, processes, relationships, data for:
  - Discovery and navigation of resources
  - Disambiguation of functionality and products, and avoiding conceptual and terminological confusion
  - Negotiation, mediation, and composition
  - Information extraction with improved precision and recall
- Standardized, meaningful, machine-readable interfaces with adjustment to users' profiles and access patterns

HCI and Ontology—New Possibilities

Richer semantics of ontology:

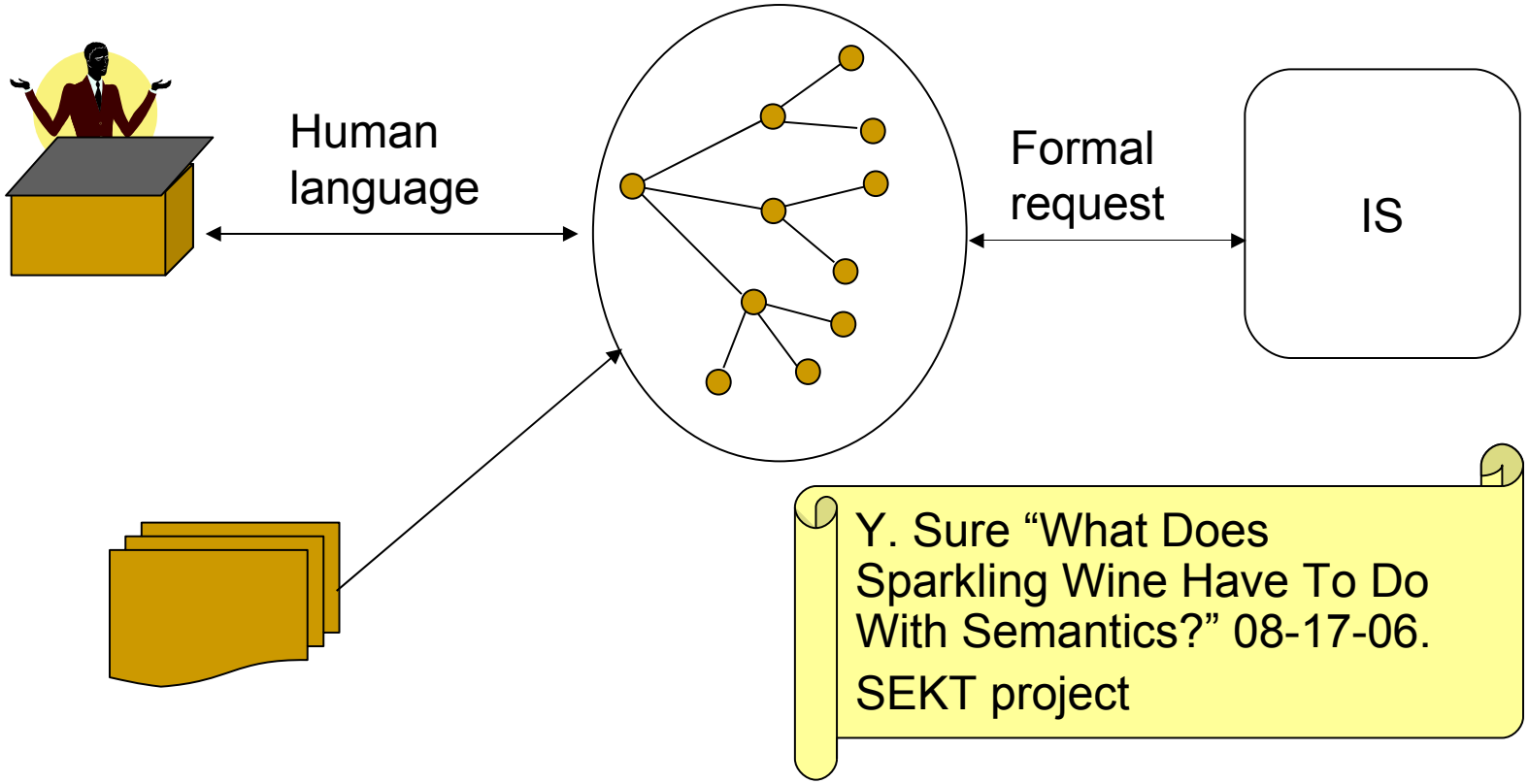
- Narrowing the gap between the human language and computer interface
- Processing of unstructured data, natural language documents and requests
- Information extraction vs. information retrieval
- Human Language Technology, ontology-based language processing
- Natural language generation

A. Walker “Usability, Databases and Ontologies” 03-08-07.

Executable English.

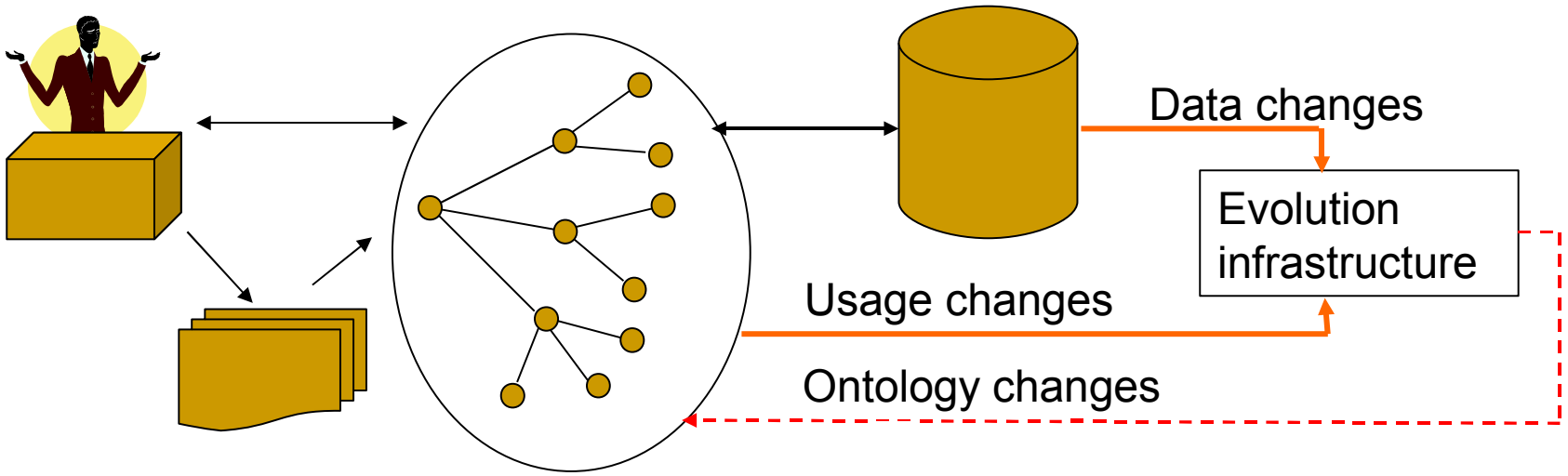
Human Language Technology

- Ontology-based HLT bridges HL with formal models that define operating the IS



Ontology, HCI, and Database—Symbiosis

- Ontologies, as mediators between the users and information resources have to reflect changes in user requirements,needs in data, access patterns, as well as changes in data sources.



HCI, Database—Influence on Ontology

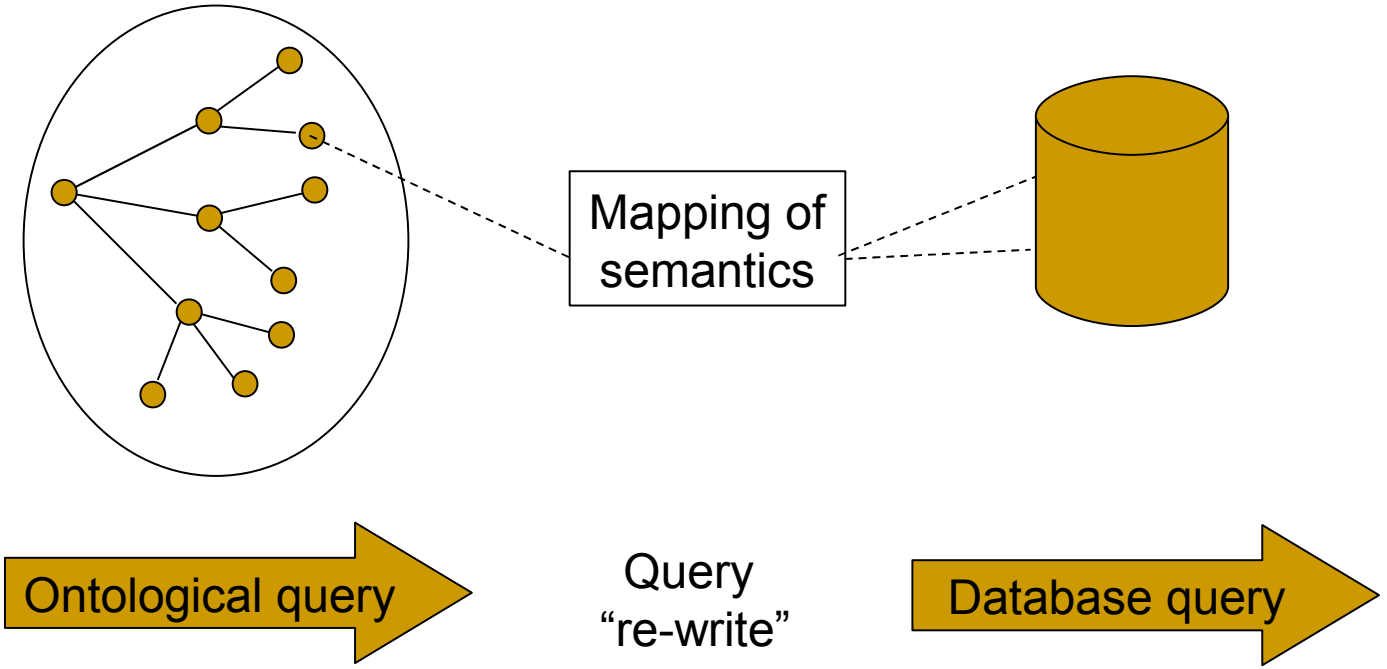
- HCI- and data-driven changes:
  - Re-classification to accommodate users' access patterns
  - Modifications in attributes, concepts, relationships in response to changed users' needs and/or data patterns

Database and Ontology: Responsibilities

- Database stores data that is the primary object of data processing:
  - Within the model that is the most beneficial for performance and ease of maintenance on large data volumes
  - Reflects conceptual peculiarities of a particular domain and is task-oriented
  - Defines physical characteristics of data implementation
  - Executes queries
- Ontology stores data that features the subject of data processing and IS environment:
  - Within the model that provides rich semantics
  - Reflects semantic characteristics of data that are partially reflected in the database semantics
  - Is separated from “physical gloss” (M. West) of data
  - Analyzes requests and transforms them in the queries

Interaction of Ontology and Database

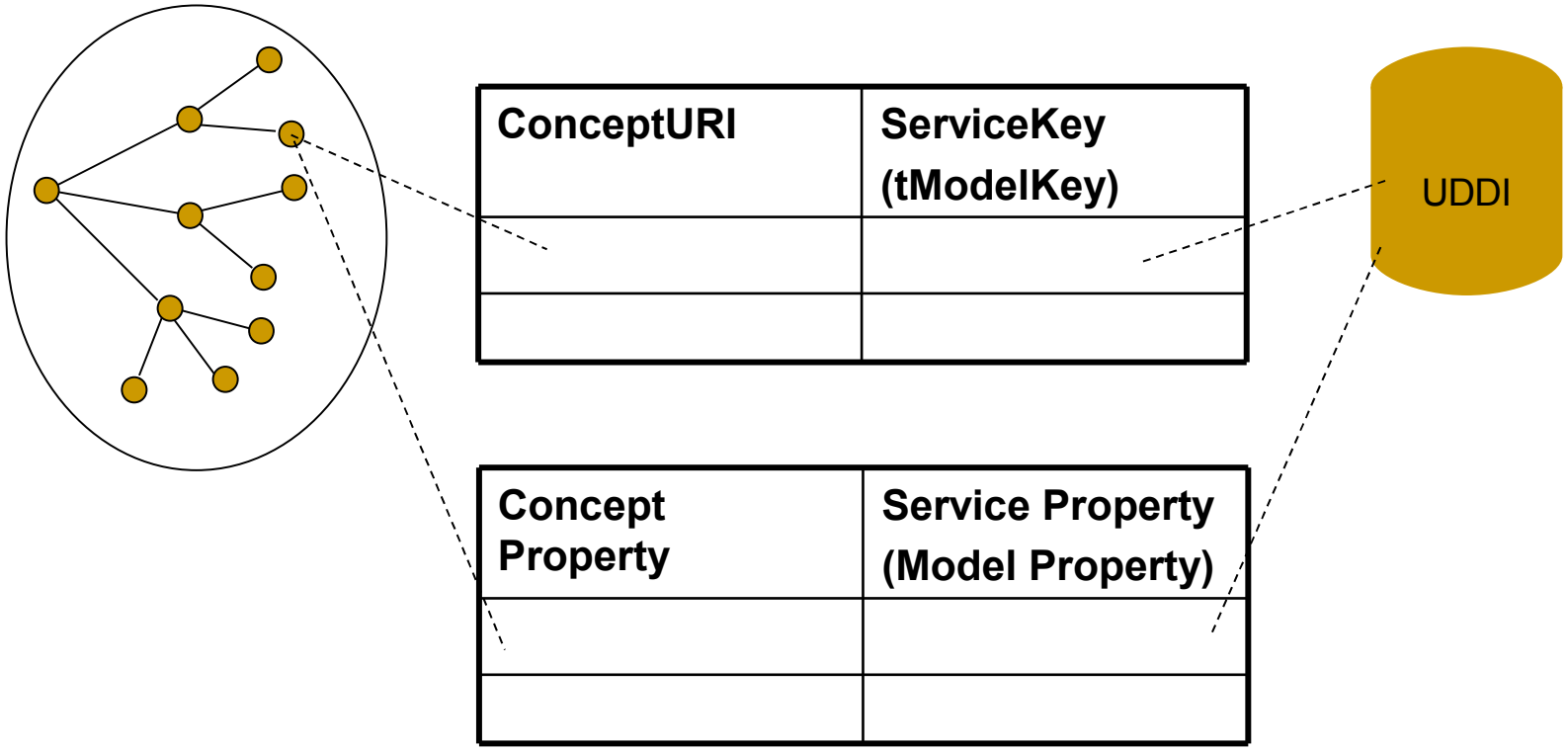
- Bridging semantics of ontology and database



Improved data precision and recall

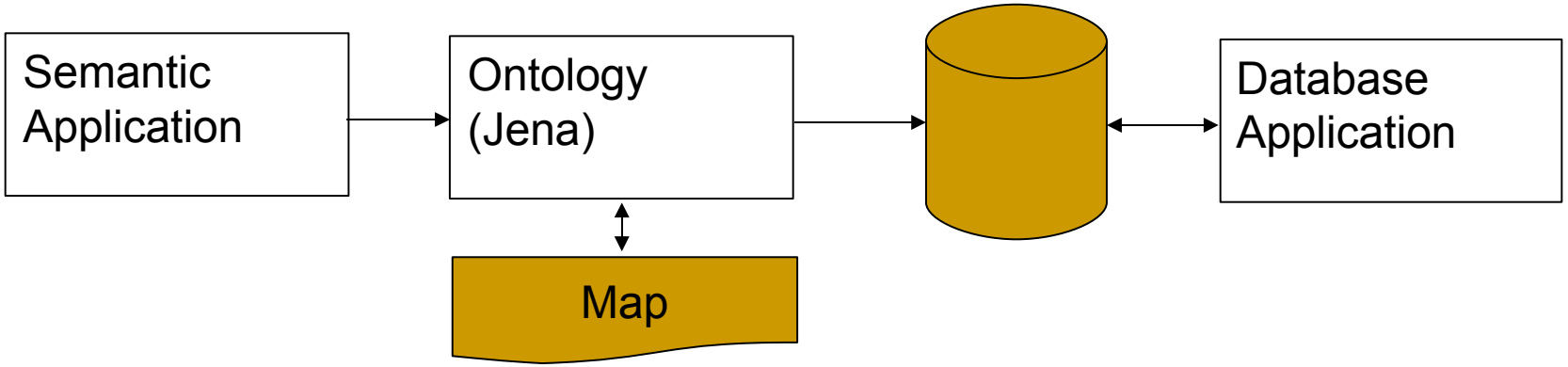
Examples—Ontology and UDDI

Multiple initiatives for Web-services discovery with the help of Ontology



Examples: D2RQ, SWRL

D2RQ maps database into a virtual RDF graph

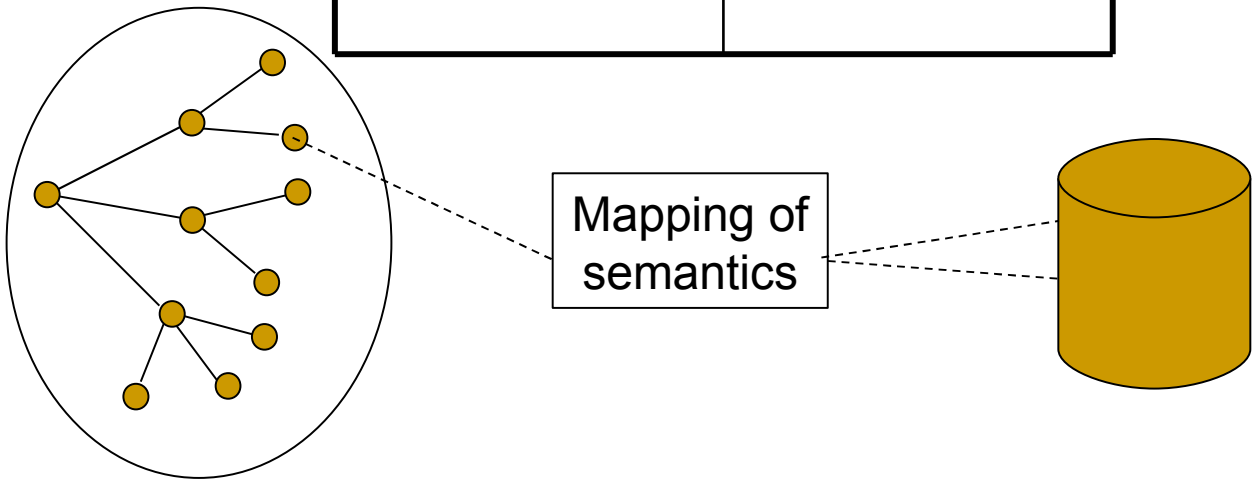


M. O'Connor "Efficiently Querying Relational Databases using OWL and SWRL" 02-08-07.

Bridging: Database a Passive Partner

- Bridging approaches do not utilize enough the database semantics and result in a rather tight-coupling of Ontology and Database

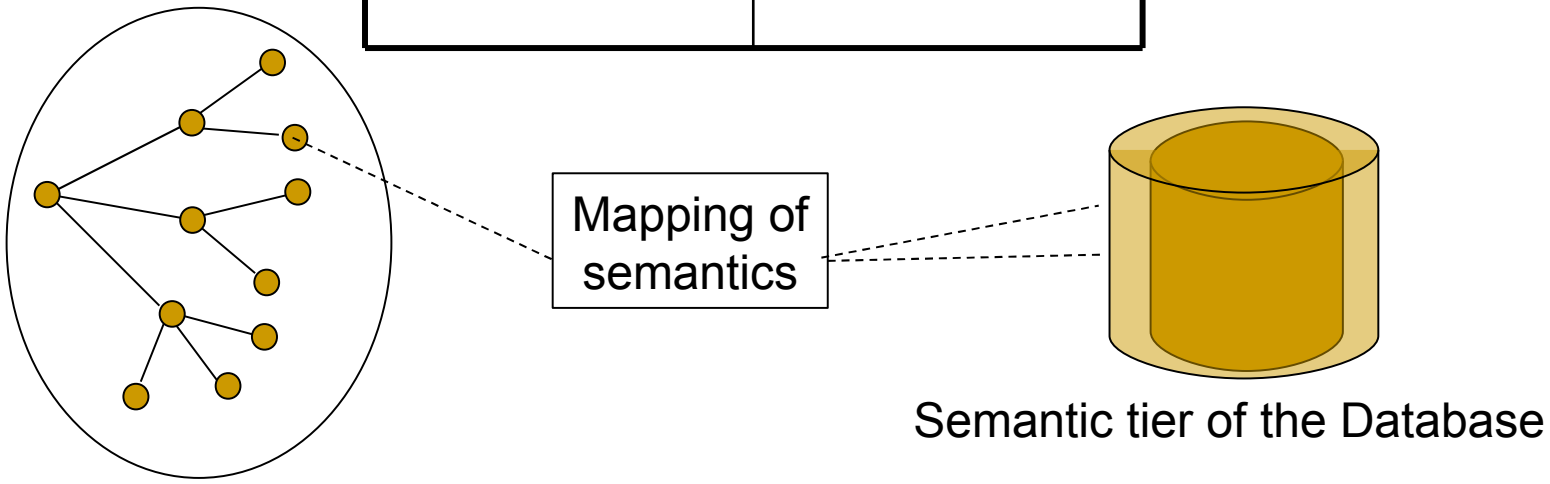
Concept URI	Data Source
Employee URI	Table Person Join Table Department



Bridging: Database's Move to Ontology

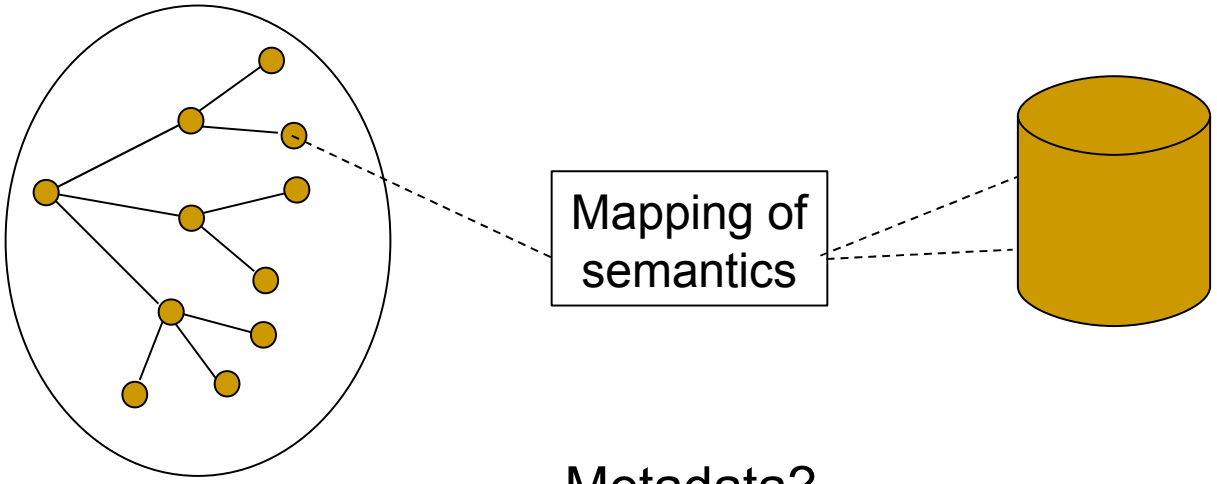
- Database has to volunteer to participate in the relationship with Ontology by creating an additional semantic tier, e.g. views, procedures, to hide the "physical gloss"

Concept URI	Data Source
Employee URI	View Employee



Bridging: Standardization?

- Needed the model (ontology?!) of mapping



Metadata?
Metamodel??
Ontology???

Ontology in a Database

- Databases are much more than just data. They implement data support infrastructure:
  - Security
  - Reliability and availability
  - Query optimization
  - Transaction management
- Ontological data can benefit from this infrastructure
  - Storing ontologies in a database—see “Database Vendors and Ontology” below

Ontology and DBMS

- Will Semantic Repository replace DBMS?
- In the near future unlikely and in general impractical because of different roles and responsibilities, and different types of data processing.
- However, even today DBMS acquire some ontological features, and the core of any DBMS —data dictionary—which is actually the database's metadata, becomes more semantic.

DBMS With Ontological Core?

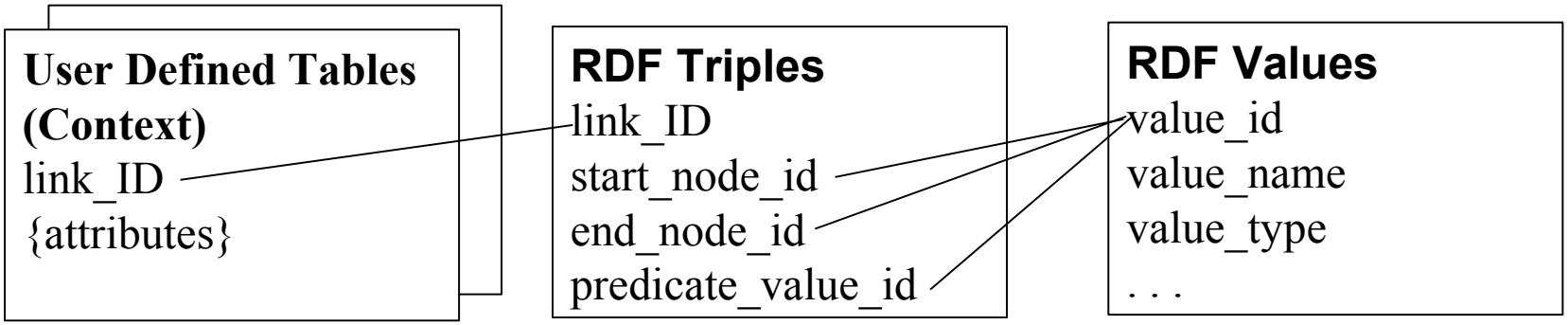
- Fascinating advancement of DBMS and more intelligent handling of data support:
  - Efficient query processing due to semantic analysis and utilization of special metadata (e.g. cost-based optimization, query re-write in Oracle)
  - Content management (IBM's UIMA based on object types)
  - Content-dependent security management
- Data dictionary includes additional metadata describing data resources, dependencies between data elements, and data and users.
- From metadata dictionary to ontological dictionary?

Database Vendors and Ontology

- Support
- Storage
  - Currently DBMS offer rather straightforward storage of ontological triples
- Implementation of inference mechanisms
- Implementation of rules
- Querying ontological data
  - Query processing benefits relational data model

Oracle and Ontology

- RDF Store: no support, triples storage, inference (RDFS and RDF entailment rules) and user-defined rules



Rules
rule_name
antecedents
filter
consequents

Coming 06-28-07:
M. Annamalai from Oracle

IBM and Ontology

Integrated Ontology Development Toolkit (IODT) Includes Ontology Definition Metamodel (EODM), EODM Workbench, and an OWL Ontology Repository (named Minerva):

- EODM is derived from the OMG's ODM and implemented in Eclipse Modeling Framework (EMF).
- In order to facilitate software development and execution, EODM includes RDFS/OWL parsing and serialization, reasoning, and transformation between RDFS/OWL and other data-modeling languages. These functions can be invoked from the EODM Workbench or Minerva and called by the applications that use ontologies.
- EODM Workbench is an Eclipse-based, integrated, ontology-engineering environment that supports ontology building, management, and visualization.
- Minerva is a high-performance OWL ontology storage, inference, and query system based on RDBMS

Future of Semantic and Database Technologies

- Recent involvement of database power players is promising
- Conceptual and technological merge can help to tame data tsunami