



# Overcoming Ontological Conflicts in Information Integration

Aykut Firat, Stuart Madnick and Benjamin Grosf  
MIT Sloan School of Management  
{aykut,smadnick,bgrosf}@mit.edu

Presentation for

ICIS 2002

Barcelona, Spain

December 15-18

# Death. Taxes. Integration.

## Motivation

- DATEK IT INTEGRATION CHALLENGES AMERITRADE

“...consolidate data centers, develop a single online-trading Web site and set up unified systems...” (CW)

- DEFENDING THE U.S.

“..a sea of unconnected islands of information technology..” (EW)

- A FRANCO-GERMAN PARTNERSHIP IN THE COCKPIT

“EADS: a big challenge to forge an integrated, cross-border group within Europe...” (FT)

- MARKET MAKES IT PRIORITY IN DRUG MERGER

“Competitive pressures make it priority for (Glaxo Wellcome and Smith-Kline Beecham) to combine their systems ...” (CWorld).

# Death. Taxes. Integration.

## Motivation

- DATEK IT

“...consolidated  
Web site and

- DEFENDING

“..a sea of u

- A FRANCO

“EADS: a b  
group withi

- MARKET M

“Competitiv  
and Smith-



- MERITRADE

online-trading

technology..” (EW)

- E COCKPIT

, cross-border

- MERGER

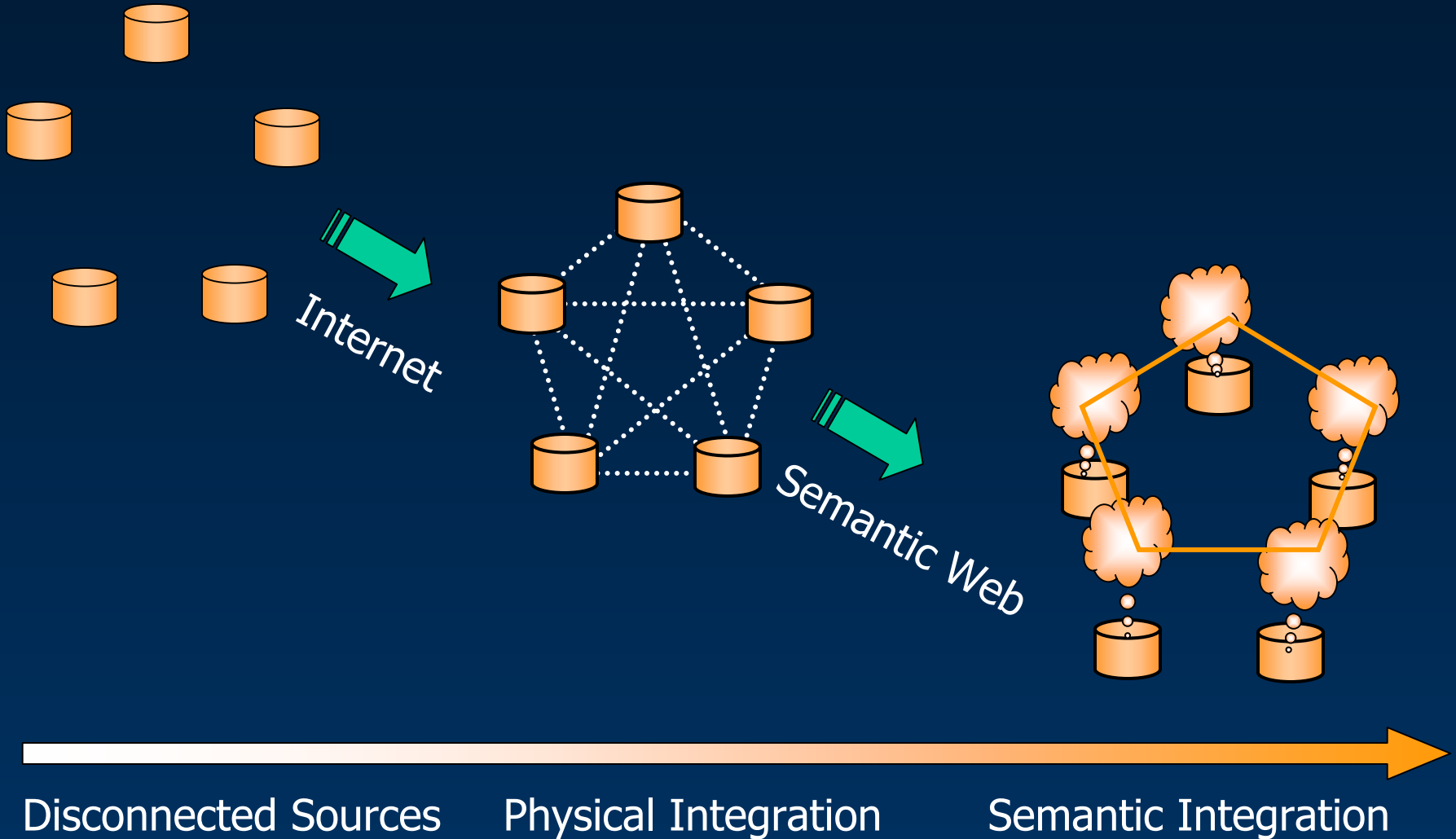
axo Wellcome  
systems ...” (CWorld).

# Roadmap

- Key Concepts
- Case Study
- Solution Methodology
- Test Example
- Concluding Remarks

# Semantic Integration

## Key Concepts



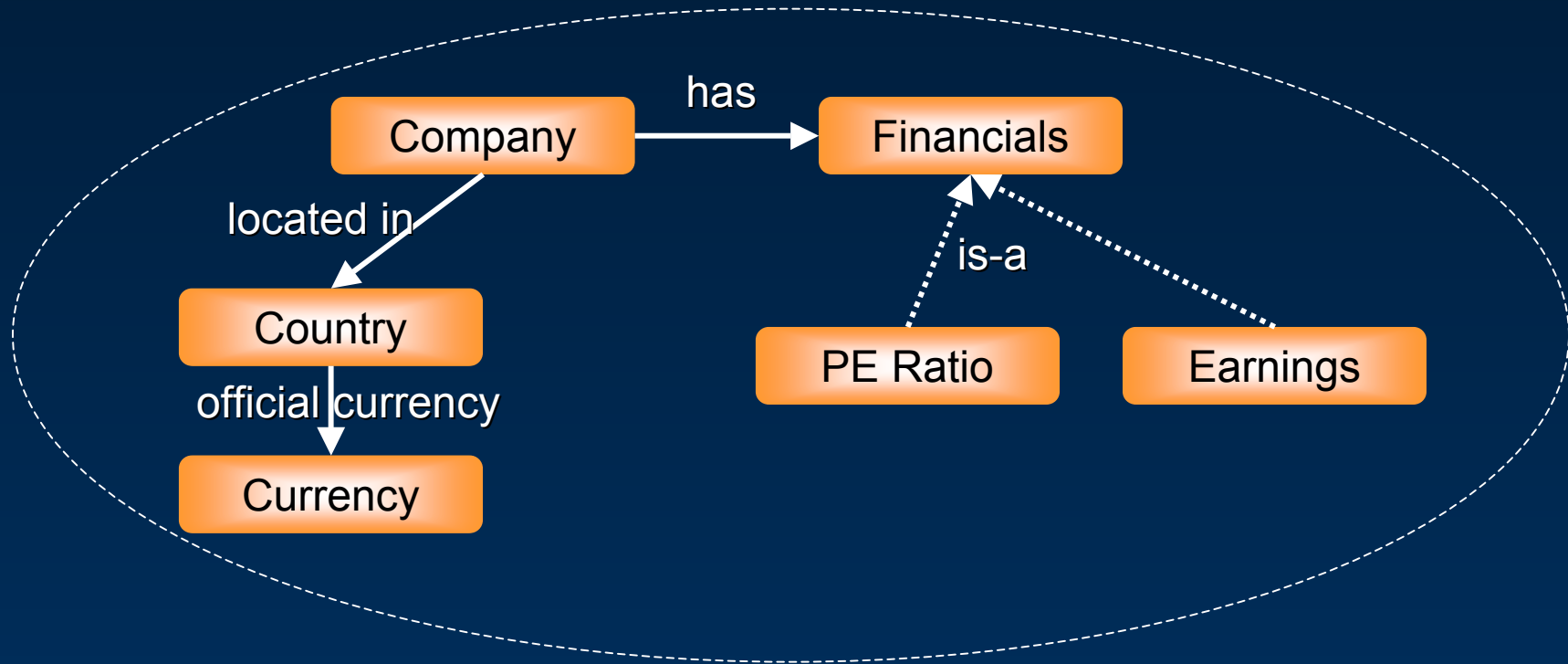
Disconnected Sources

Physical Integration

Semantic Integration

# Ontology

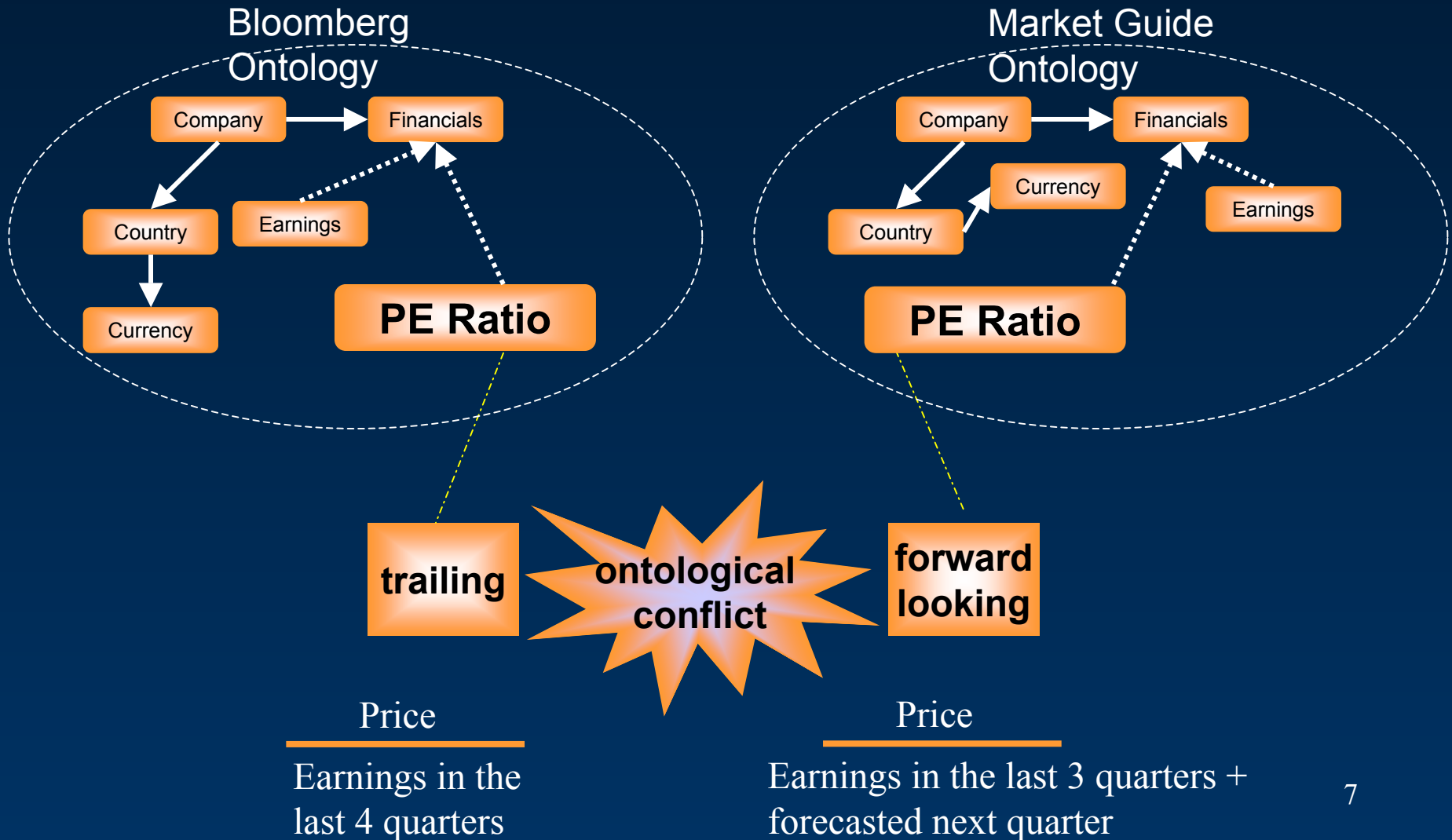
- “Specification of a conceptualization”



A snapshot from a financial ontology

# Ontological Heterogeneity

## Key Concepts



# Information Chain

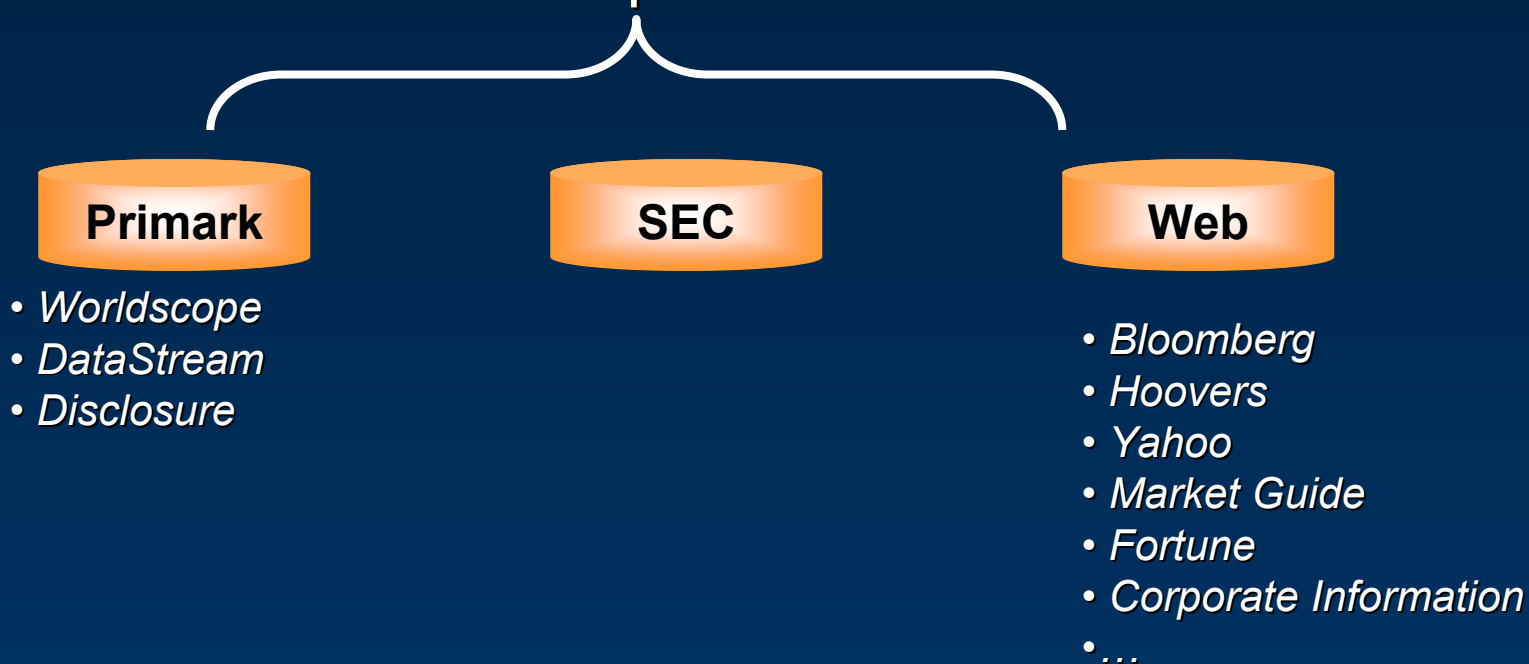
## Case Study





### Found Three Different Types of Heterogeneities

- Data Level
- Ontological
- Temporal

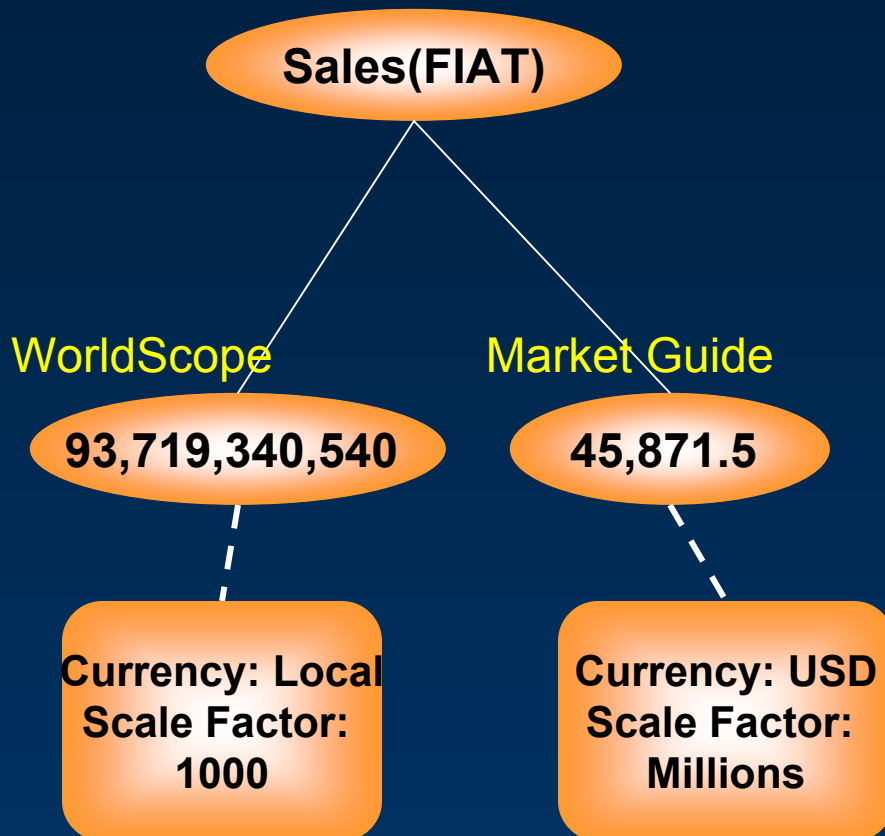


# Data Level Heterogeneities

## Case Study

Definition: Same entity different representations

### Sales Data



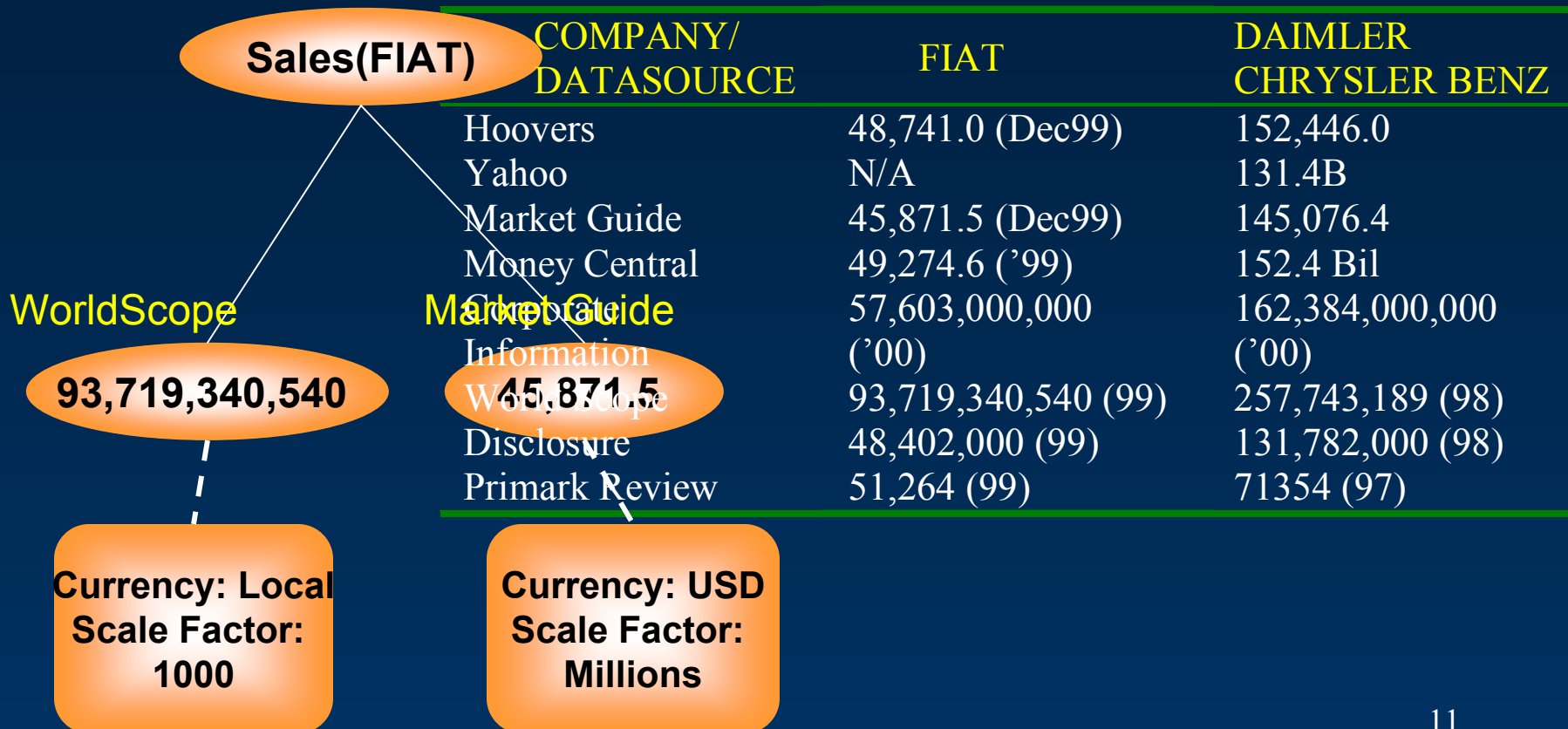
COMPANY/ DATASOURCE	FIAT
Hoovers	48,741.0 (Dec99)
Yahoo	N/A
Market Guide	45,871.5 (Dec99)
Money Central Corporate Information	49,274.6 ('99)
World Scope	57,603,000,000 ('00)
Disclosure	93,719,340,540 (99)
Primark Review	48,402,000 (99)
	51,264 (99)

# Data Level Heterogeneities

## Case Study

Definition: Same entity different representations

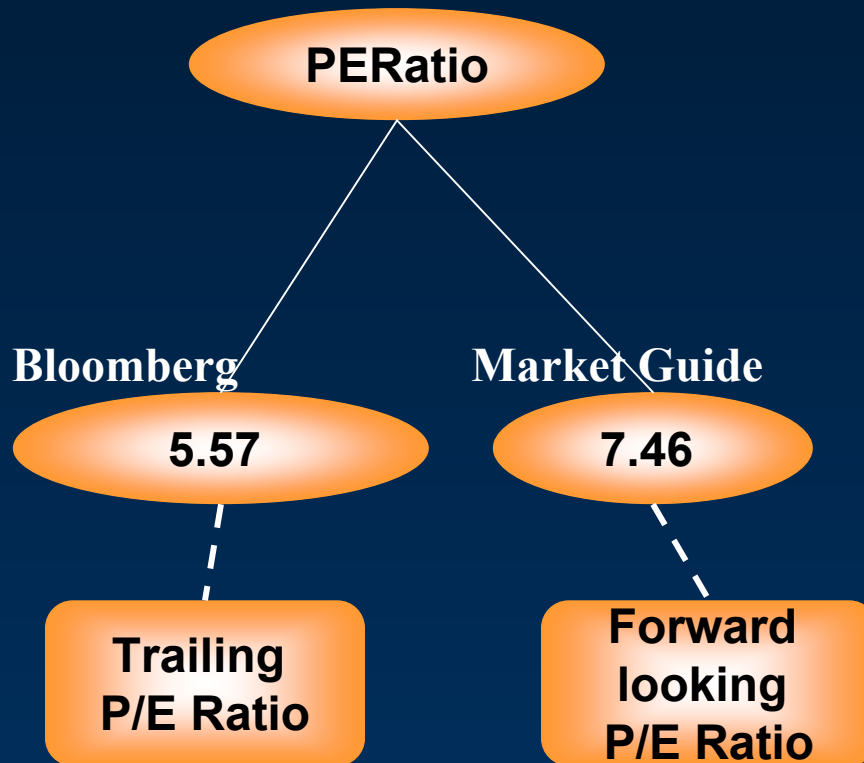
### Sales Data



# Ontological Heterogeneities

## Case Study

*Definition: Different entity type definitions and/or relationships*



SOURCE	P/E RATIO
ABC	11.6
Bloomberg	5.57
DBC	19.19
MarketGuide	7.46

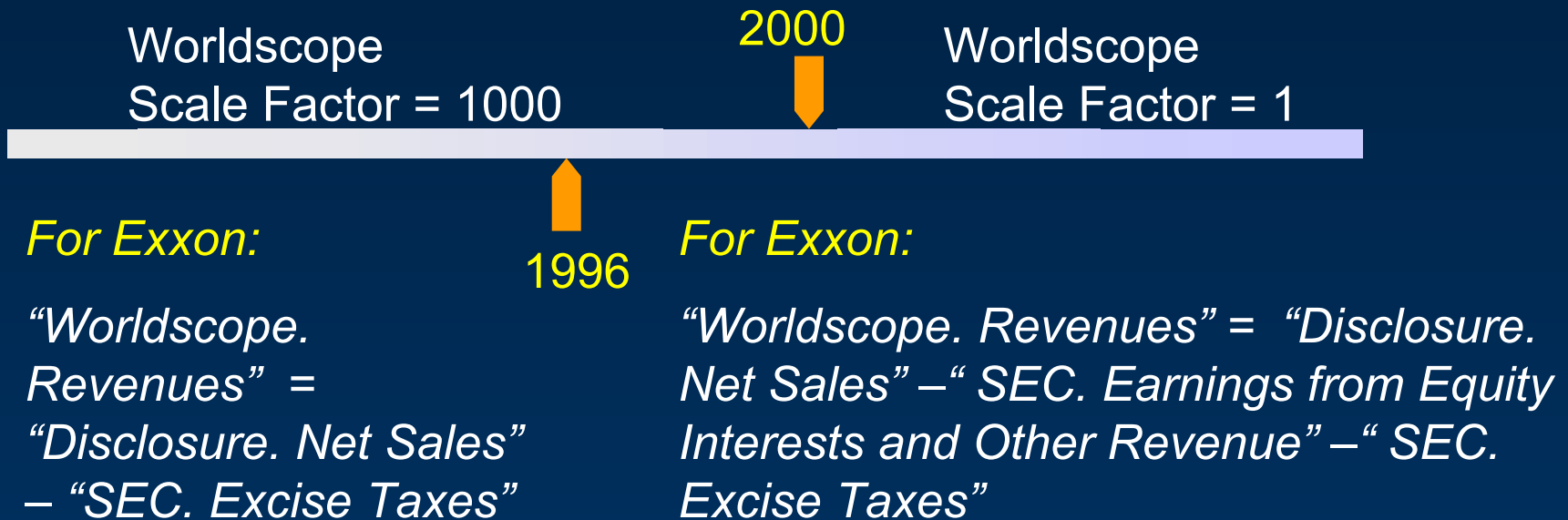
$Price/PE_{Trailing} =$

$Price/PE_{Forward} - Forecasted\_Quarterly\_Sales(t+1) + Quarterly\_Sales(t-3)$

# Temporal Heterogeneities

*Definition: Entity values or definitions belong to different times, or time intervals.*

## Data Level



## Ontological

# Our Challenge

How to represent and reason with semantic heterogeneities?

Why challenging?

- Machine to machine communication: AI
- Declarative, scalable, extendible solution required
- Modeling: combination of art and science
- Includes NP-Complete problems (e.g. Source Selection)

# Approach

## Context Interchange (COIN): Data Level

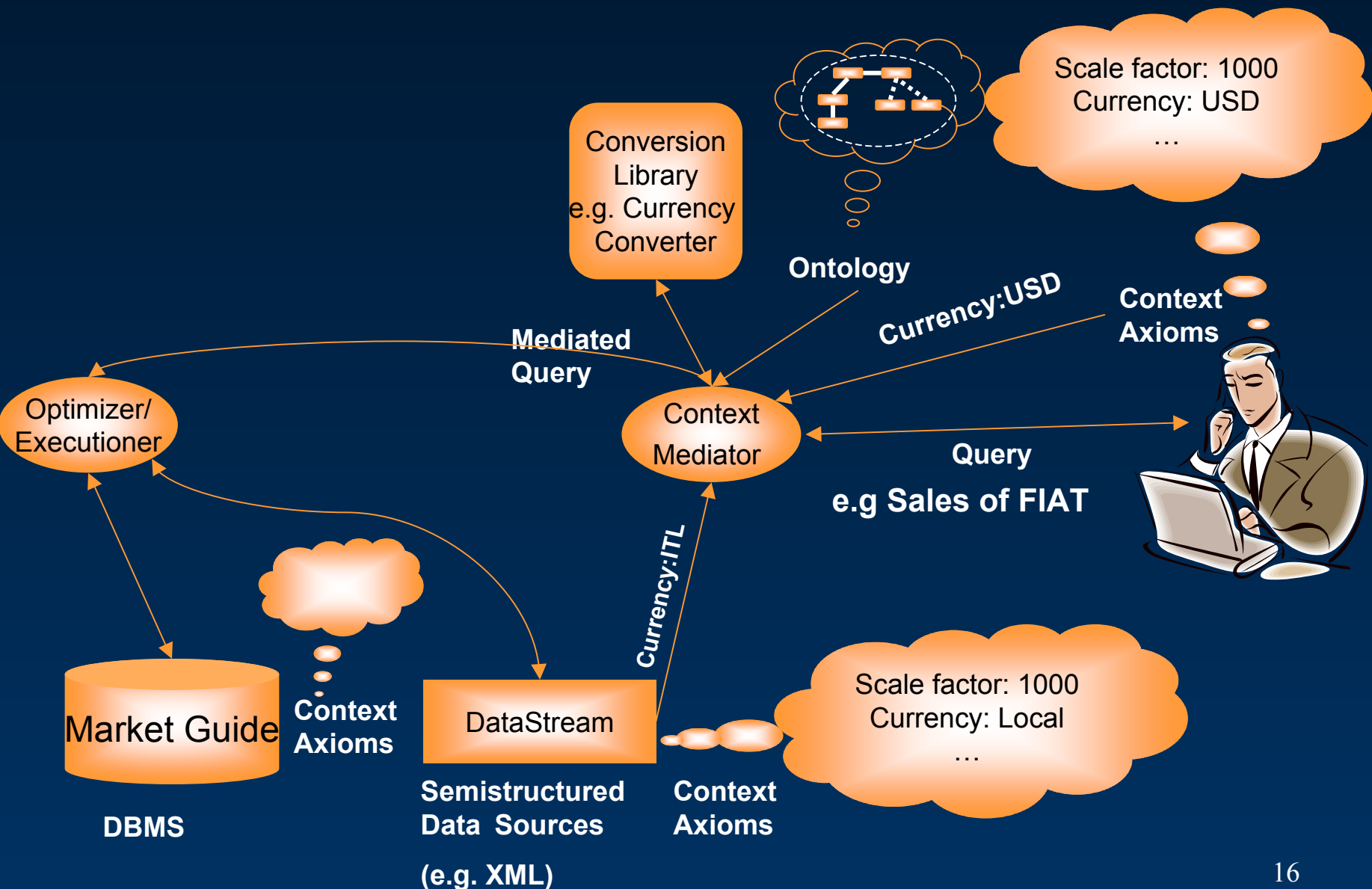
- Research undertaken by Sloan IT Ph.D. Cheng-Goh
- Logical Framework: COIN Data Model + Abductive Logic Programming
- Loosely coupled approach to semantic integration

## Extended COIN: Ontological

- Extended Data Model + Symbolic Equation Solving Techniques
- Based on Constraint Logic Programming
- Source Selection
- Ontology Merging

# COIN Architecture

## Solution Methodology



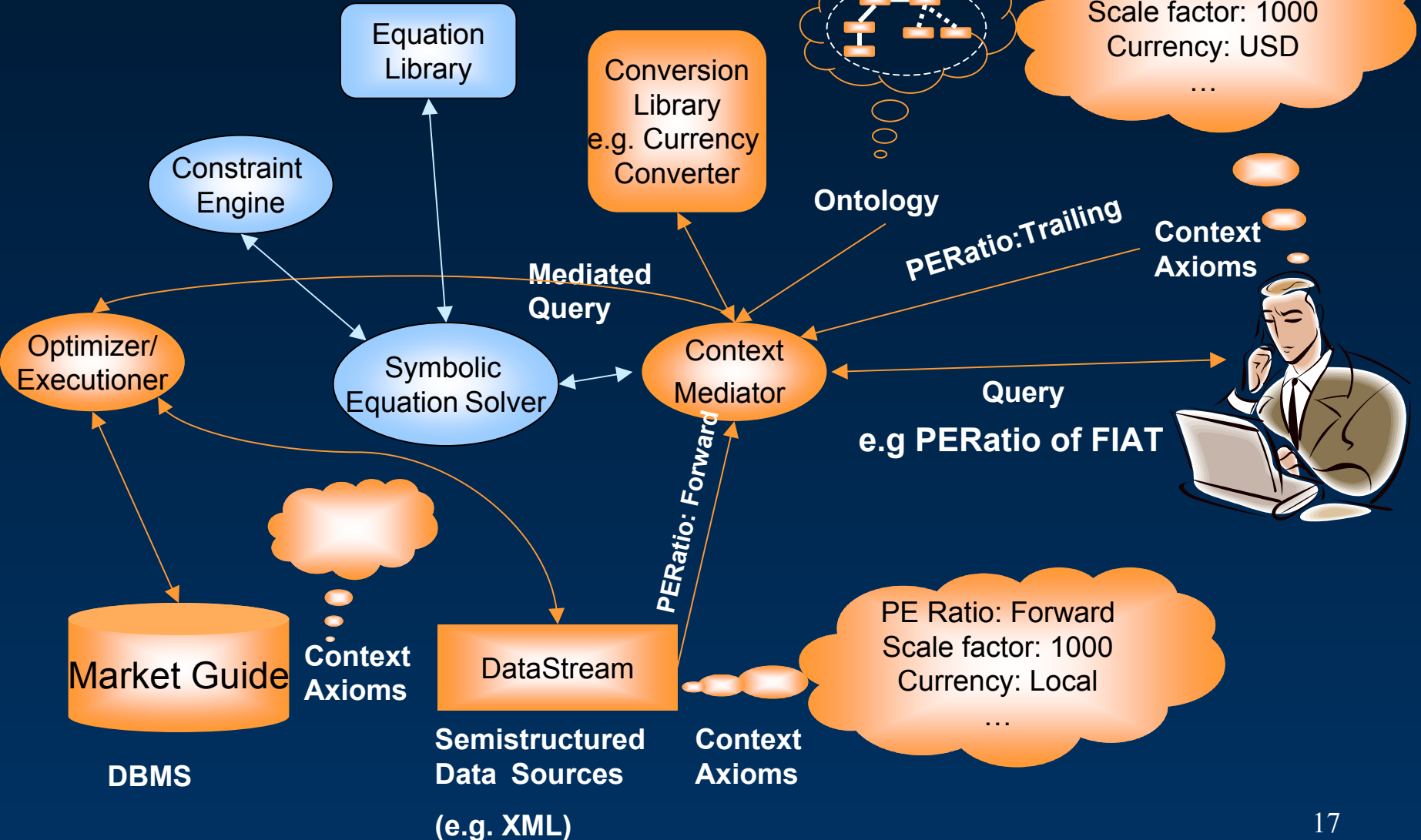


# ECOIN Architecture

## Solution Methodology

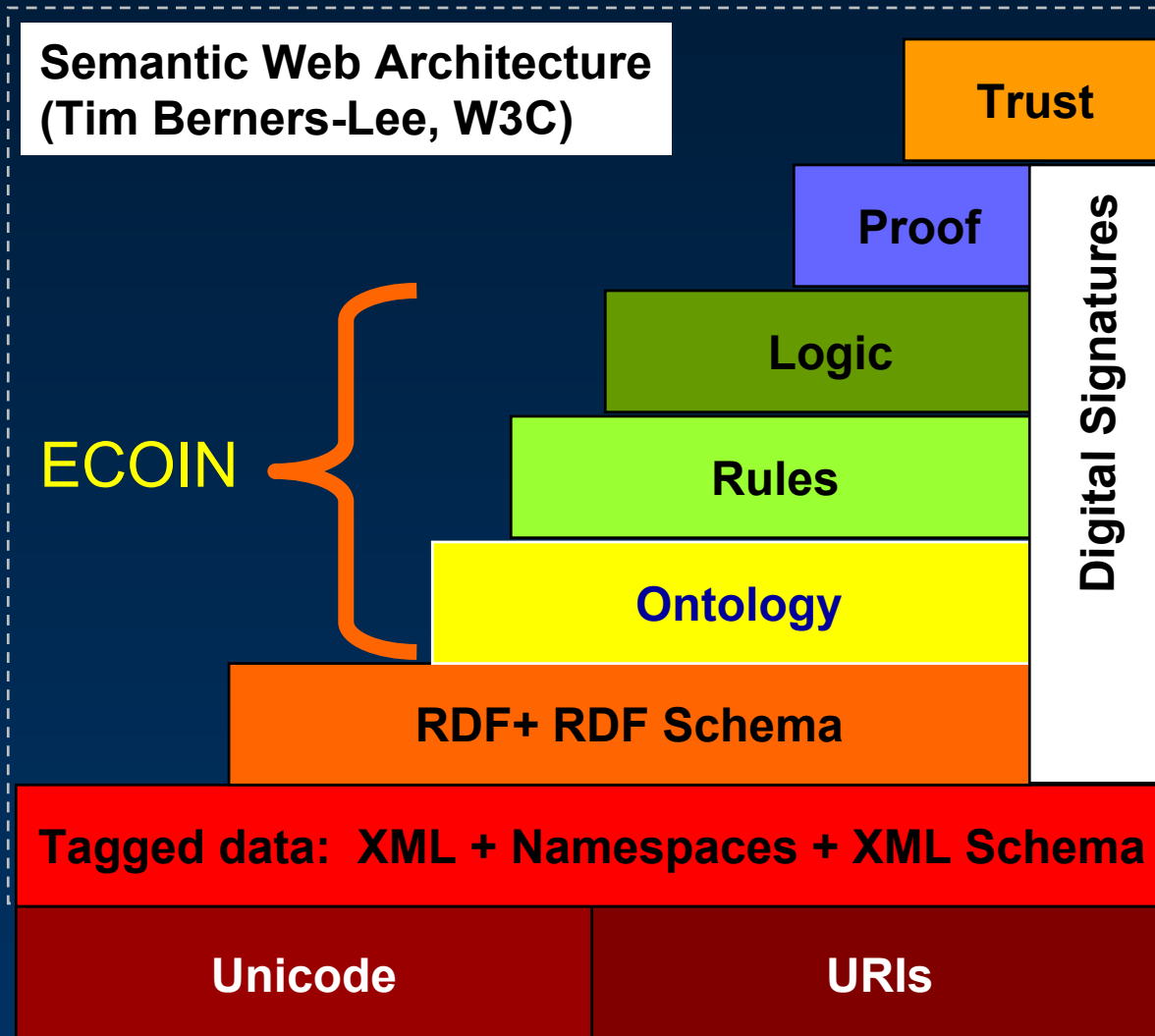
$Price/PE_{Trailing} =$

$Price/PE_{Forward} - Forecasted\_Quarterly\_Sales(t+1) + Quarterly\_Sales(t-3)$



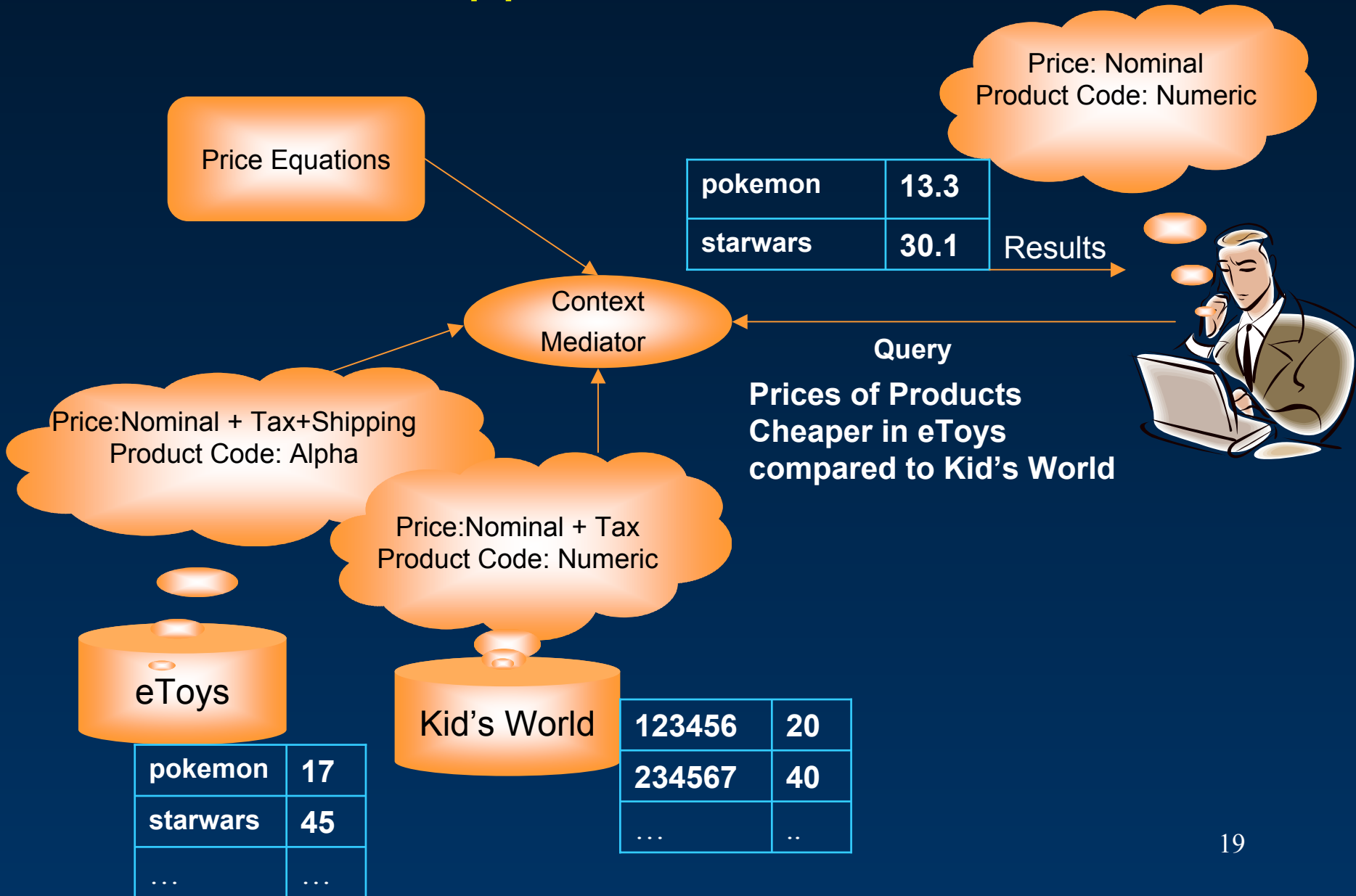
# Key Link: ECOIN & Semantic Web

Solution Methodology



# E-Business Application

## Test Example



# Concluding Remarks

- We developed ECOIN
  - Symbolic Equation Solving Techniques
  - Proof of the concept prototype
- First system to deal with both data level and ontological heterogeneities
- Temporal in our agenda next
- Contribute to the global semantic interoperability efforts

# Industry solutions?

## Problems addressed

Many Experts in **Physical** and **Tightly-Coupled** Integration

- Content Aggregation: **Yodlee,..**
- Workflow Integration: **IBM, Microsoft,..**
- Enterprise Application Integration: **SAP, WebMethods, Siebel**
- B2B: **Ariba, CommerceOne,..**
- Portals: **Vignette, Broadvision,..**

Our focus is on **Semantic** and **Loosely-Coupled** Integration