Overcoming Semantic Barriers in Integration of Heterogeneous Enterprises

Guoray Cai
College of Information Sciences & Technology
The Pennsylvania State University
About me

- Associate Professor of IST at Penn State
- Educational background in Electrical Engineering, Geography, and information science
- **Research areas**: geographical information systems (GIS), visualization, human-computer dialogues, semantic modeling, discourse modeling.
- **Teaching**: Databases, visual analytics, Information science theories and methods.
Nature of Large Enterprises

Composed of many sectors of businesses.

Each sector operates in relatively autonomous modes.

Diverse business goals

Adapt and change to avoid risk and capture new opportunities
Business integration means integrating applications, data, and processes within an enterprise or amongst a set of enterprises.

Commonly cited purposes:

1. Reduce redundancies
2. Resolve inconsistencies
3. Capture dependencies and relationships
4. Better modularity and composability
5. Expose functions to novel applications
EA: ecology system as metaphor
EA as theory of how business works

Inside and outside

Inside only

Should EA include people and their activity?
Challenge of EA

Challenge 1: *disjointed collection of hardware and software*

Solutions:

1. Standardize
2. SOA
Challenge 2: Semantic **Heterogeneity**

Signs, Semantics, and Meaning

![Diagram showing the relationship between Concept, Symbol, and Referent]

- Concept activates Symbol
- Symbol stands for Referent

Semantic Model

Real World

Physical Data Stores
1. **Semantic heterogeneity.** Expressions always refer to facts. A group of people may use different expressions to refer to the same Real World fact, or they may use the same expression to refer to different facts.

2. **Syntactic heterogeneity.** Syntactical level of abstract deals with formalizing the grammar of schemata and expressions.

3. **Schematic heterogeneity.** An example of pragmatic investigation is the study of the way people categorize, or classify, and relate Real World facts. The outcome of this process is a schemata of the underlying facts. Naturally different business sections have different schemes of the Real World facts.
Semantic heterogeneity

**Multiple representations**: A Real World fact may have more than one description in the underlying databases to comply with various disciplines, giving as a consequence semantic heterogeneity.
What does Delaware river mean to you?

Interpretation:
- Waterways?
- State boundaries?
- Edge of a parcel?

Naming:
- Watercourse or river?
Different interpretation of the same reality

Pavement management group
Road networks, their number of lanes, their intersections, and their traffic flow.

Marketing group
see road networks as houses and streets as addresses of clients.

In the pavement management and the marketing group streets serve different purposes. In this case there is no common base of definitions of the underlying facts between the two disciplines.
EA Models as Abstraction Mechanisms

Database Architecture

- PAYROLL
  - View 1
- ACCOUNTING
  - View 2
- ADVISING
  - View 3

Conceptual Schema

Physical Schema

Disk
EA Models as Abstraction Mechanisms

Goal-dependent meaning

Cognitive heterogeneity

Schematic heterogeneity

Syntactic heterogeneity

Naming heterogeneity

Contextualization

Ontology mapping

Conceptual model
Schemata integration

Logical model
Schema integration

Thesaurus, metadata
Ontology: formal specification of a shared conceptualization

A knowledge representation language is used to formally describe the concepts of an ontology.
Problems with Ontology-based approach

Agreed upon top-level ontology

Mapping will lose meaning

Difficult to interpret

The result lacks of affordance

Ontology approach is best used in applications where the core problem is the use and management of common representations.

Ontology approach fails to work when no global ontology can be defined for the local ontologies.
Contextualization of Ontologies

Contexts and ontologies have both strengths and weaknesses. It can be argued that the strengths of ontologies are the weaknesses of contexts and vice versa.

Ontologies capture what are common. Shared ontologies define a common understanding of specific terms, and thus make it possible to communicate between systems on a semantic level.

On the weak side, ontologies can be used only as long as consensus about their contents is reached.

Context capture local and unique meaning.
Proposal: Contextualization

Multiple ontologies which contain information that *should not be integrated* should be *contextualized*. An example would be concepts which are mutually inconsistent.
Method: Context-Mediated Semantic Integration

1. Contexts should be explicitly represented; contextual knowledge should be associated with context representations; and be used to guide all facets of an agent’s behavior.

2. A semantic model of EA should have multiple ontologies ranging from top-level generic ontology to application specific ontologies. Each ontology be associated with a context that ‘wraps’ around it.

3. Ontologies are related through the generalization/specialization relationship of their contexts, as well as through explicit ‘lifting’ rules.

Contextualization hides the heterogeneity of data at the ontology level, just like ontologies effectively hide the heterogeneity of data at the syntax level.
Method: Context-Mediated Semantic Integration

Semantic models of EA have the following six components:

1. A context space which is a set of contexts \{C_i | i=1,..,N\}, where N is the total number of contexts
2. An ontology space which is a family of ontologies \{O_i | i=1,..,N\}
3. A set of inter-ontology bridging rules \{\Phi_{i,j} | (i, j \in \{1,.., N\}) and (i \neq j) \}. Each \Phi_{i,j} is a set of rules that specify how elements of ontology \(O_i\) relates to elements in ontology \(O_j\), if any relationship exists.
4. A set of inter-context bridging rules \{\Psi_{i,j} | (i, j \in \{1,.., N\}) and (i \neq j) \} that specify how context \(C_i\) relates to context \(C_j\), if any relationship exists.
5. A set of rules governing context coordination.
6. A set of rules governing ontology coordination.
Context, ontology, and meaning

Nested context space
What are the benefits?

1. **Contexts provide the necessary metadata knowledge that help establish bridges between two different ontologies.** One of the important components of context is the joint activity between cooperating agents.

2. **Contexts narrow down the search scope for discovering common semantic anchors.** When a shared ontology is impossible, contexts (when shared) provide a constrained domain where local interpretation of foreign concepts is discovered.

3. **Contexts create opportunity for emergent data semantics.** Shared contexts allow semantic agreements to be established ‘on-the-fly’ (during the time of interaction) rather than relying on pre-existing consensus.
Research questions

1. **What is the nature of context?**
   Is it an application, a domain, a team, a person, an organization?
   We propose the notion of activity as the central construct of contexts.
   We propose to do empirical study of activity-based contexts (ABC)

2. **How to represent context?**
   We explore the use of predicate logic to represent contexts and their state.

3. **How does the approach help agility of EA?**
   We propose to use the context/ontology abstraction hierarchies to model both functions and data of EA. Working with the sponsoring businesses, we would like to demonstrate and measure the impact of our approach to business agility.

4. We propose to study the degree to which our architecture model improve the affordance of business decisions/actions.
Research question 1: Activity-centered contexts

Based on well established cognitive theories (Situated actions [Suchman, 1987] and Activity Theory [Nardi, 1996]). They are cognitive basis for modeling business operations.

[Definition] An activity is a coordinated set of goals together with the necessary set of mental states (beliefs, intentions, and commitments) of the interacting agents and environmental conditions that ensure the success of the shared goals.
Research question 1: Activity-centered contexts

We represent an activity by a SharedPlan {Grosz, 1996}.

The progression of an activity corresponds to the process of evolving the SharedPlan from a partial one towards a full SharedPlan (Lochbaum, 1998).
Research question 1: Activity-centered contexts
Research question 1: Activity-centered contexts

1. Computational representation of activity-based context hierarchies for semantic abstraction.
2. Contextualizing data architecture
3. Contextualizing function architecture
4. Context recognition
5. Context-mediated Business-IT alignment
Research question 2: Contextualizing EA for agility and diversity

**WHY?**

Adding a new business initiative is likely to require less effort EA update, such changes are contained local.

- Using context to capture the unique aspects of semantics, and
- Use ontology to capture the shared semantics.
Questions /Discussions