

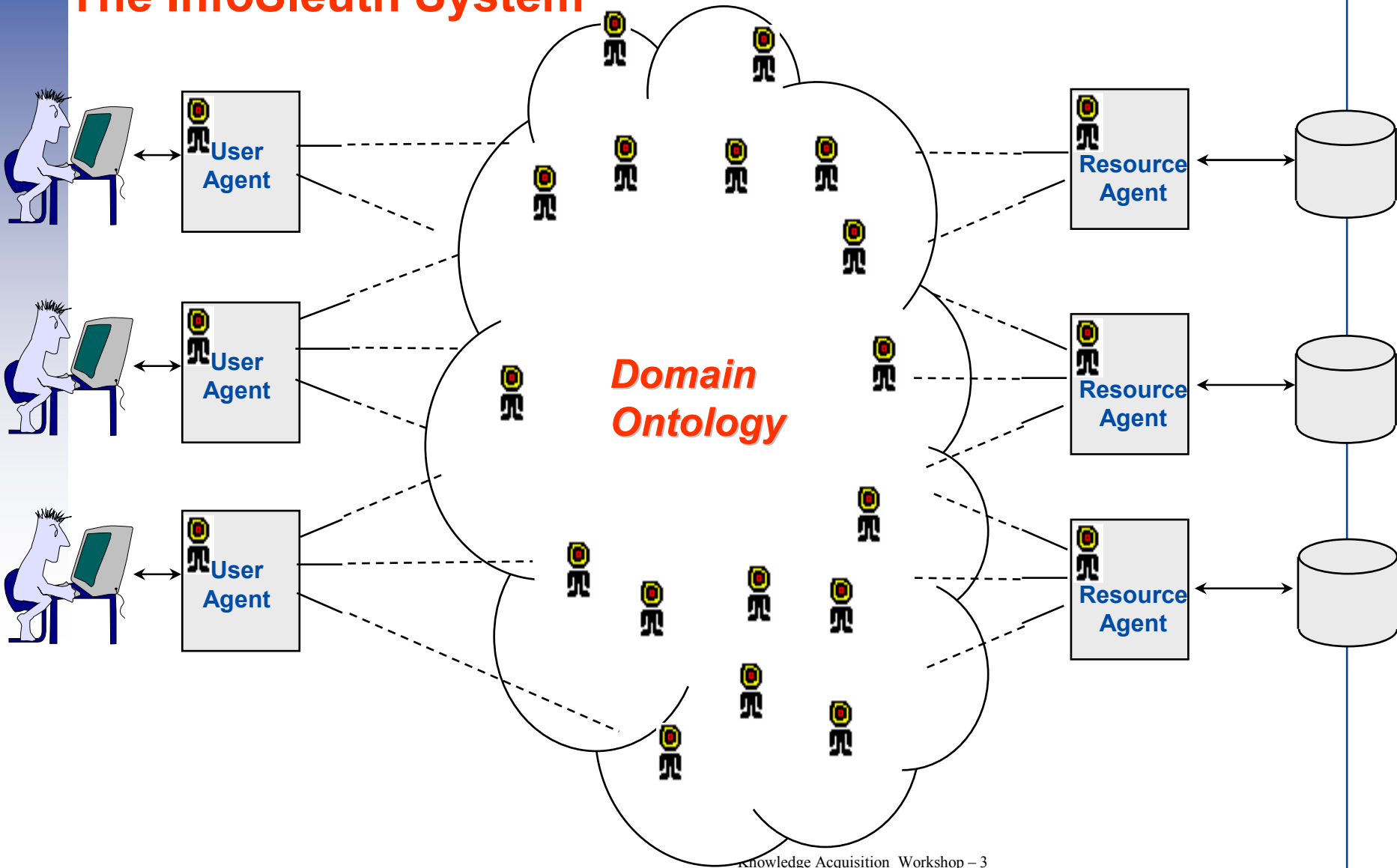
# **Design and Creation of Ontologies for Environmental Information Retrieval**

**Vipul Kashyap**  
**AOS Workshop, Rome, November 2001**  
**[vipul\\_kashyap@yahoo.com](mailto:vipul_kashyap@yahoo.com)**

## Outline

- Ontologies for Information Retrieval: The InfoSleuth System
- Sources for Ontology Construction
- The Ontology Design Process:
  - “Reverse Engineering” from a database schema
  - Ontology refinement based on user queries
- Enhancing the ontology
  - Using a data dictionary
  - Using a Thesaurus
- Conclusions and Future Work

# Ontologies for Information Retrieval: The InfoSleuth System



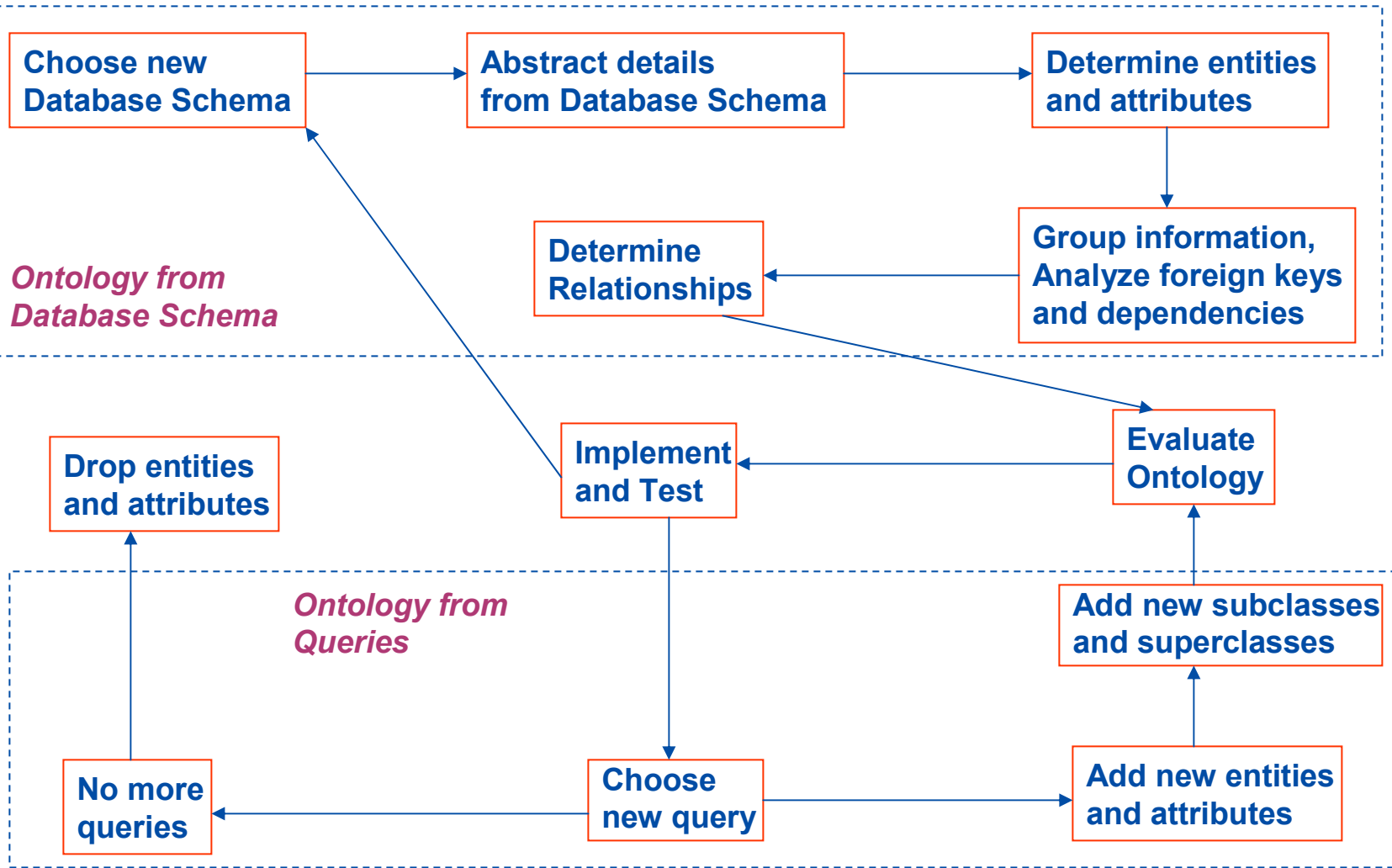
# Ontologies for Information Retrieval

- Provide a concise, uniform, declarative description of semantic information
- Independent of syntactic representations, conceptual models of the underlying information bases
- Domain models provide wider access by supporting multiple world views on the same underlying data
- EDEN ontology defined in the context of the InfoSleuth system:
  - important and crucial to capture elements of environmental information

# Sources for Ontology construction

- Pre-existing Database Schemas
  - data directed component
- Collection of representative set of queries possibly parameterized based on application user interface
  - application directed component
- Thesauri and Vocabularies (e.g., EEA Thesaurus)
  - knowledge directed component
- Ontology = knowledge-based middle ground  
between applications and data !!!

# The Ontology Design Process



# Reverse Engineering from a Database Schema

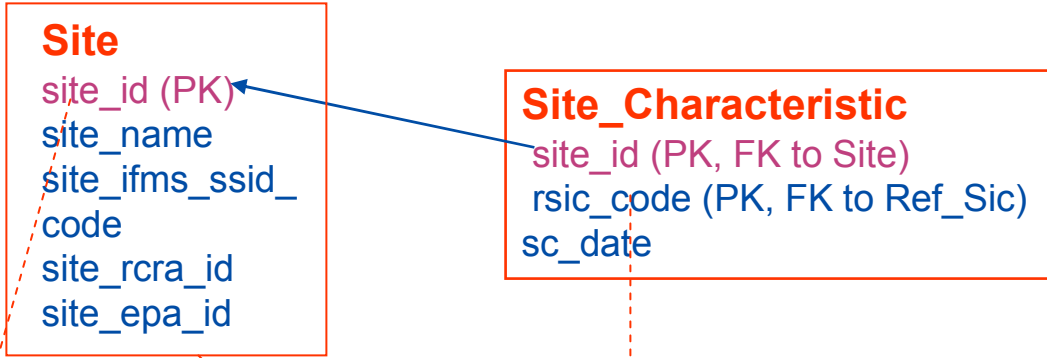
- Abstraction of details related to:
  - data organization
  - local keys
- Grouping information in multiple tables
- Identifying Relationships
- Incorporating new concepts suggested by new schema

# Environmental Databases

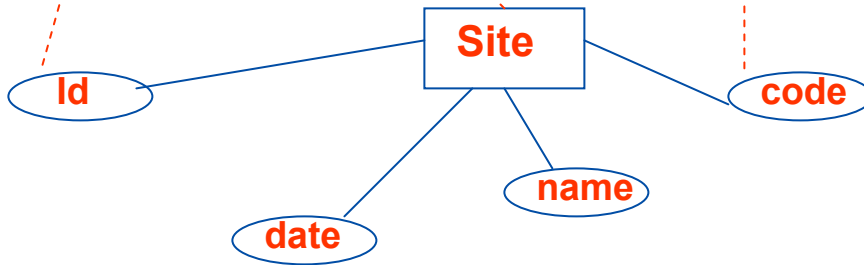
- CERCLIS 3
  - [http://www.epa.gov/enviro/html/cerclis/cerclis\\_overview.html](http://www.epa.gov/enviro/html/cerclis/cerclis_overview.html)
- ITT
- HAZDAT
  - <http://www.atsdr.cdc.gov/hazdat.html>
- ERPIMS
  - <http://www.resdyn.com/erpims>
- Basel Convention Database
  - <http://www.unep.ch/basel>

# Abstracting out details related to local keys

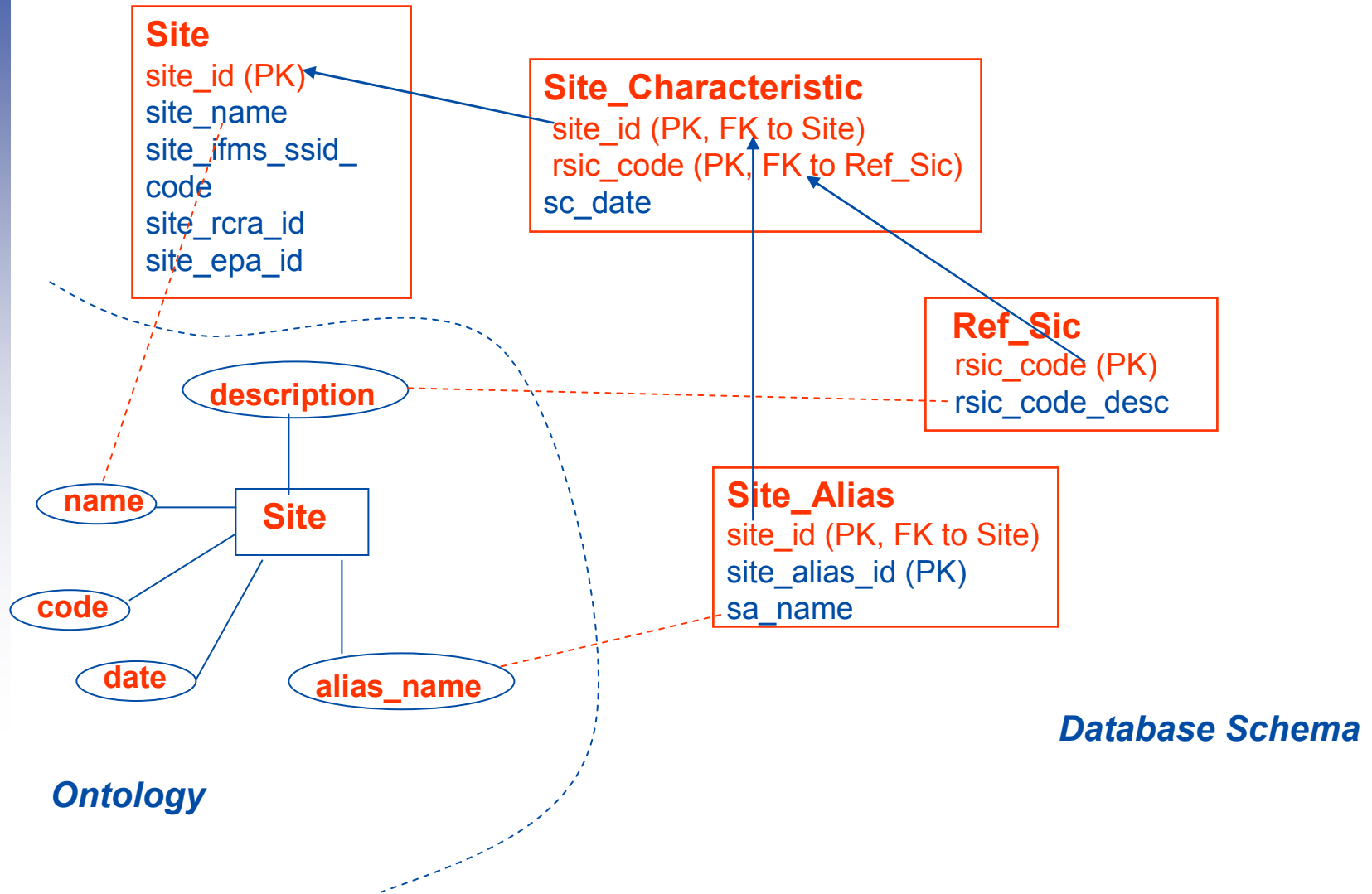
## Database Schema



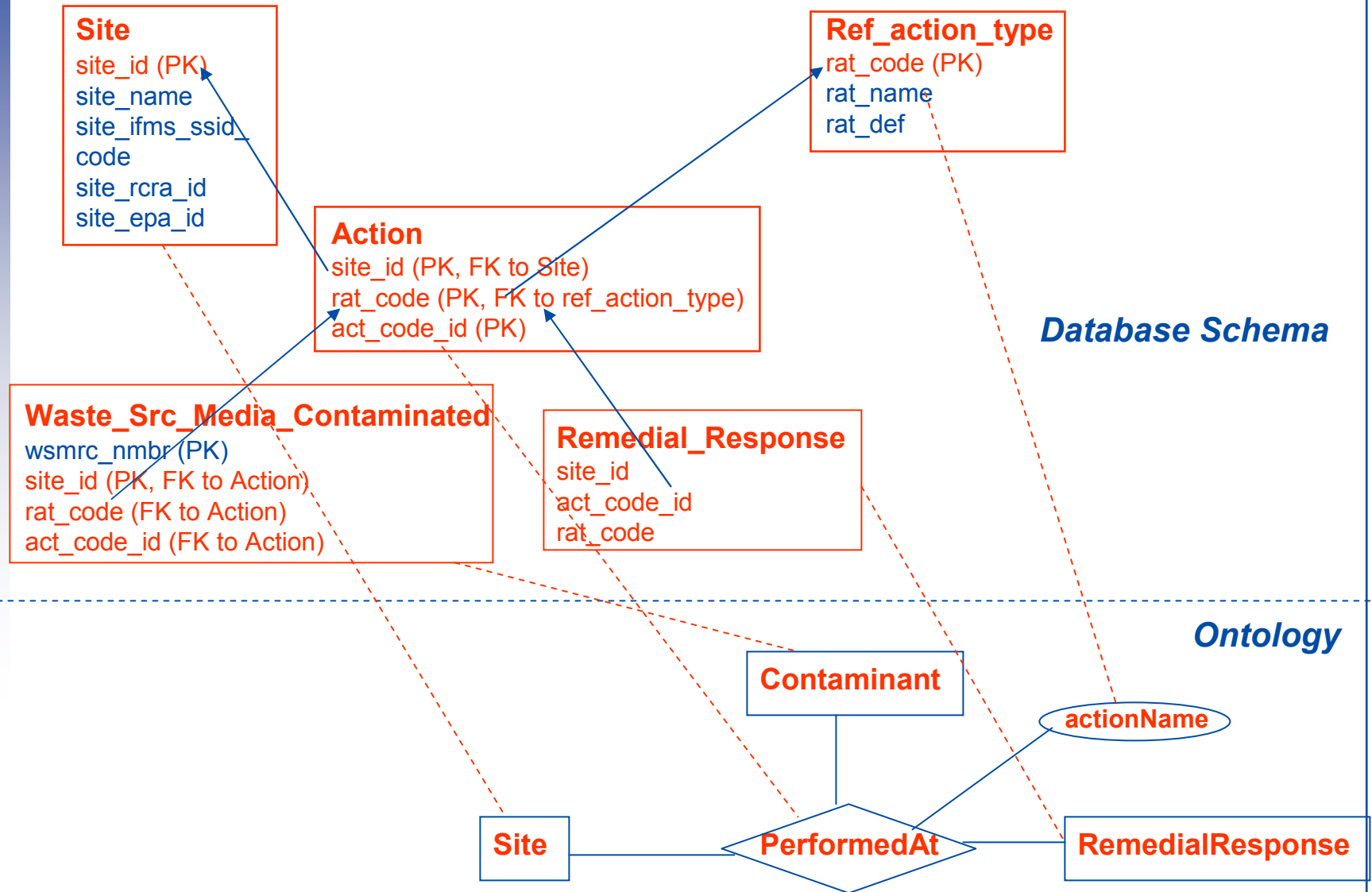
## Ontology



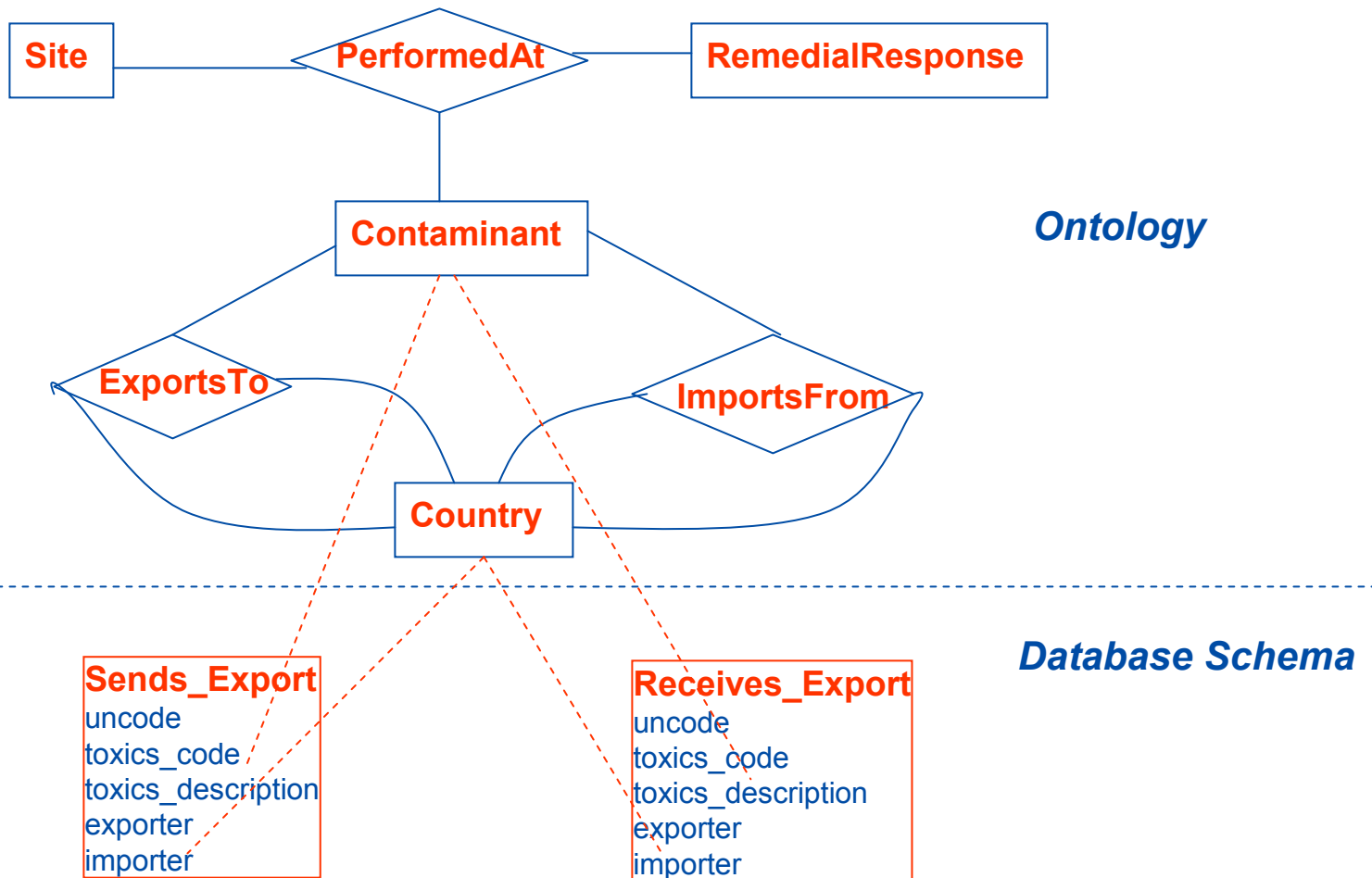
# Grouping Information in Multiple Tables



# Identifying Relationships



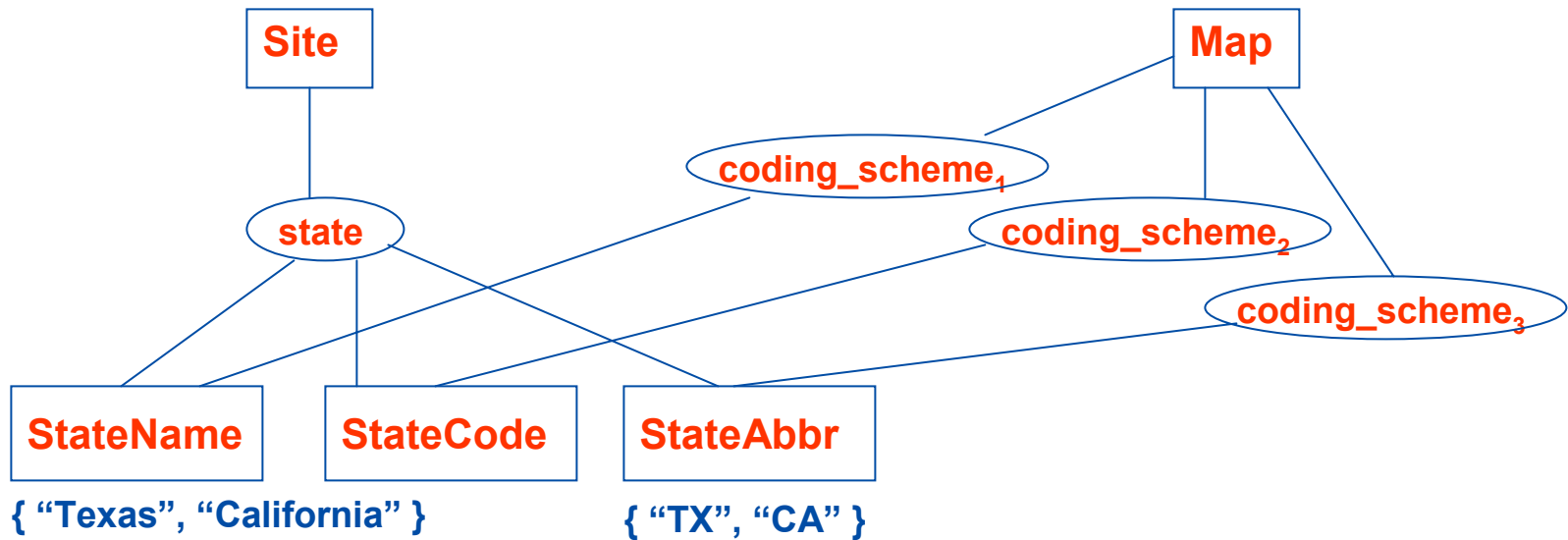
# Incorporation of new concepts from a different database schema



# Ontology refinement based on user queries

- Addition of New Attributes
  - *At NPL sites with a **land use category** of INDUSTRIAL, what is the cleanup level range for LEAD ....*
  - Add an attribute **landUseCategory** to the entity **Site** in the ontology
- Addition of new Relationships
  - *What is the range of concentrations for ARSENIC is a **contaminant of concern** in the SURFACE SOIL at NPL sites*
  - Add a relationship **HasContaminant** between the entities **Site** and **Contaminant** in the ontology
- Addition of class-subclass relationships and new entities
  - *How many **Super fund sites** are in Edison County, New Jersey ?*
  - Add an entity **SuperFundSite** as a subclass of **Site** in the ontology

## Using a data dictionary (EDR) to enhance the ontology



- `select * from Site where state = 'TX' or state = 'California'`
- `select coding_scheme1 from Map where coding_scheme3 = 'TX'`

# Enhancing the Ontology by using a Thesaurus

abandoned site

**THEME**

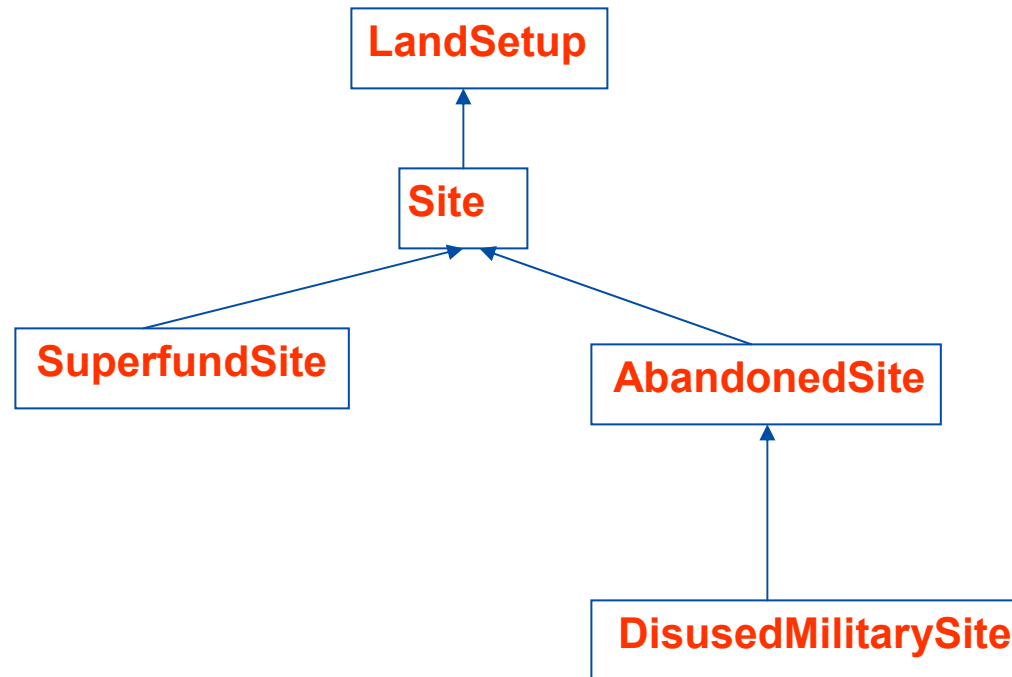
**BT**

**NT**

**POLLUTION**

land setup

disused military site



## Conclusions and Future Work

- Role of semantic content in handling data/information overload
  - **Domain Specific ontologies: an approach for capturing semantic content**
- Design and construction of domain ontologies
  - **labor intensive, time consuming, difficult endeavor**
- Re-use readily information: schemas, queries, data dictionaries, thesauri
  - **minimize the involvement of the domain expert**
- **Extrapolate this technique into other domains:**
  - **telecommunication**
  - **IP networks (use of CIM information model by DMTF)**
- **Apply these techniques to Knowledge Management and Acquisition**