The International Workshop on Advanced Information Systems for Enterprises
Constantine April,19-20, 2008

An Ontology and a Schema Description Base for Relational Database Integration

Nadir Salhi, & Amel Benna,
CERIST, University A/Mira Bejaia, Algeria

Zaia Alimazighi
LSI, USTHB Algers, Algeria

Bilal Amrouche & Ferhat Makhloufi
INI, Algeria

Presented by Amel Benna

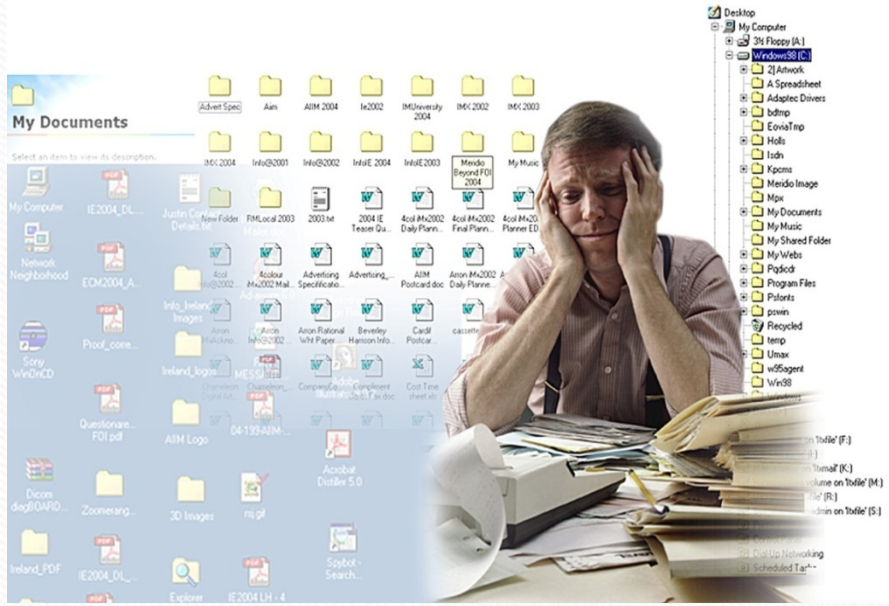
Agenda

1. Background
2. Data Integration issues
3. Our Approach for Data Integration
   - Architecture
   - Schema Description Base
   - Query process
4. Implementation
5. Conclusion & Perspectives

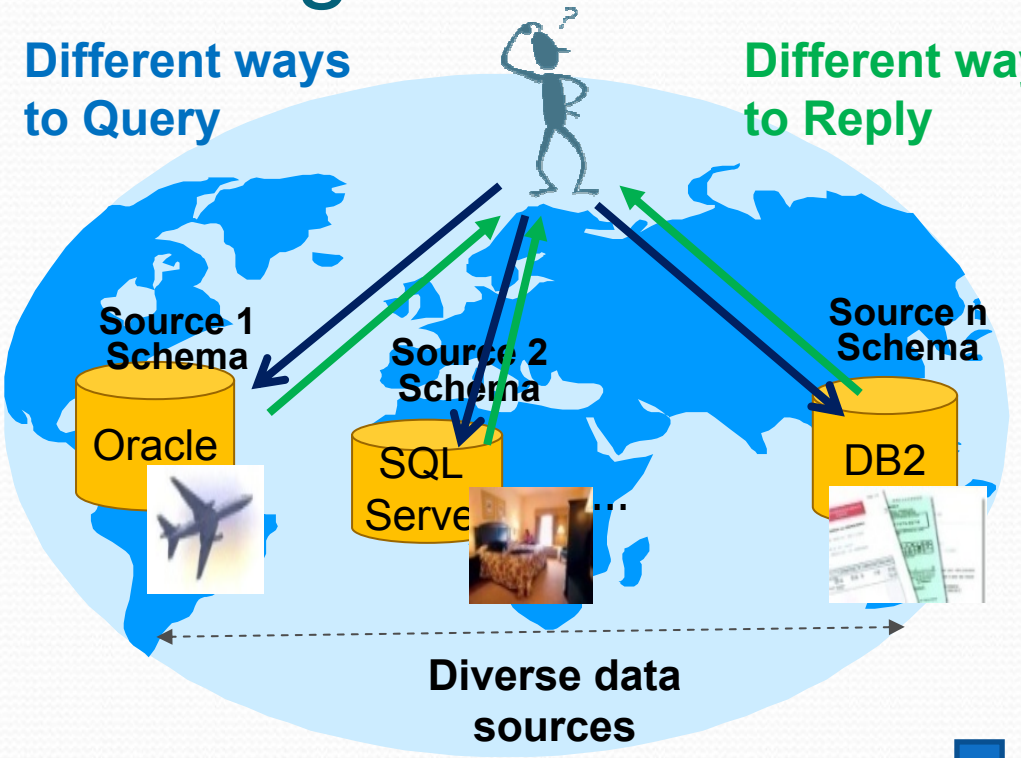
Background

- **Information Systems** are evolving in **heterogeneous** & distributed environments.
- In order to be efficient, companies need to manage and **integrate** all information sources taking into account **semantics**.

Today… the issue is too many databases, too much information



Background



- How can data sources cooperate?
- How to integrate new sources?
- How to find data semantics?

Data Integration System

- to provide a **uniform** access to heterogeneous source.
- to join partial replies from heterogeneous sources.

Data Integration Issues

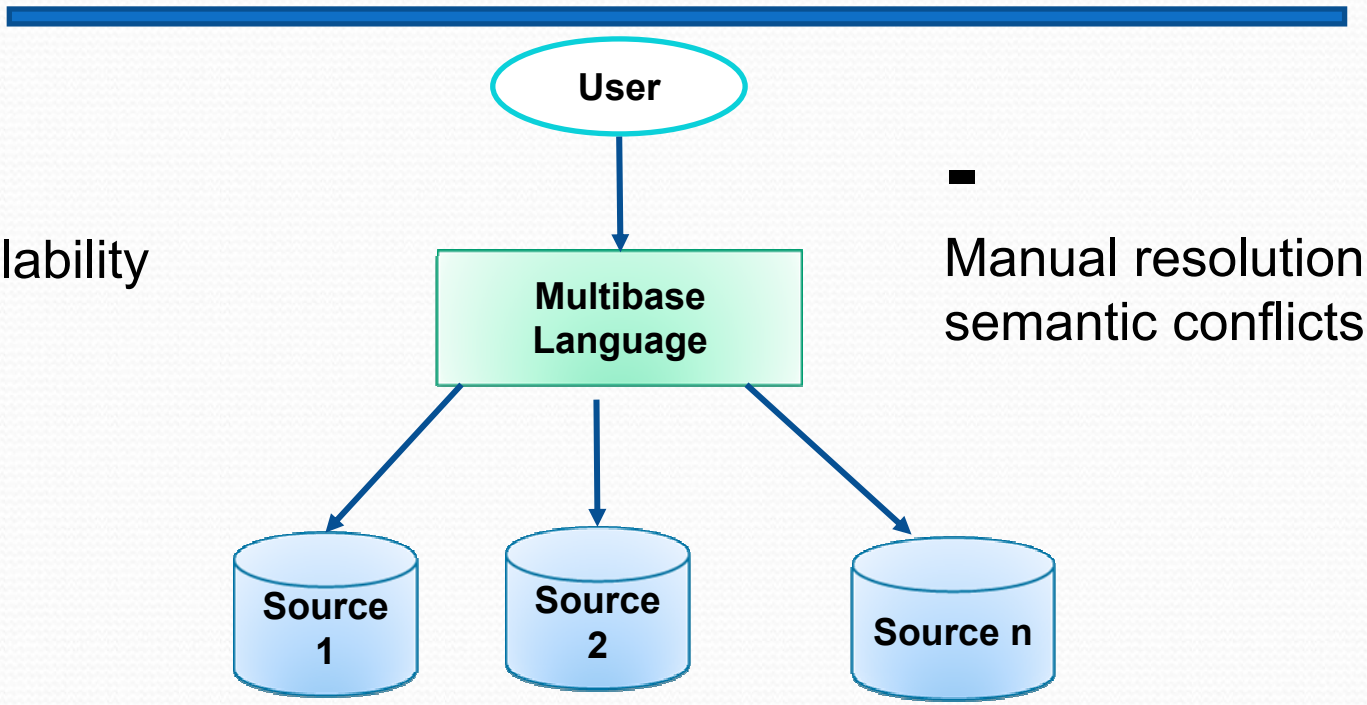
- **Definition:** "*The data integration is the process by which several sources of autonomous data, distributed and under heterogeneous shape are integrated as a* ***unique source*** *represented by a* ***global schema"****.*
- Among Issues to be addressed : **Heterogeneity**
  - Model level: RDBMS, OODBMS, XML, …
  - Structure: Eg. DB1:Book (Title, Author,) ,DB2:Book(Title, ISBN,)
  - Semantics:
    - Names: Eg. Label "NAME" used for Book Title, Author,…
    - Scaling & precision conflicts:

    Eg. Book price in DB1 in Euro with VAT, in DB2 in $ without VAT.

Data Integration Issues

Related Research in Semantic Interoperability for DB is categorized

1. **Query-oriented** (based on declarative languages or extended SQL)



\+

Scalability

\-

Manual resolution of semantic conflicts

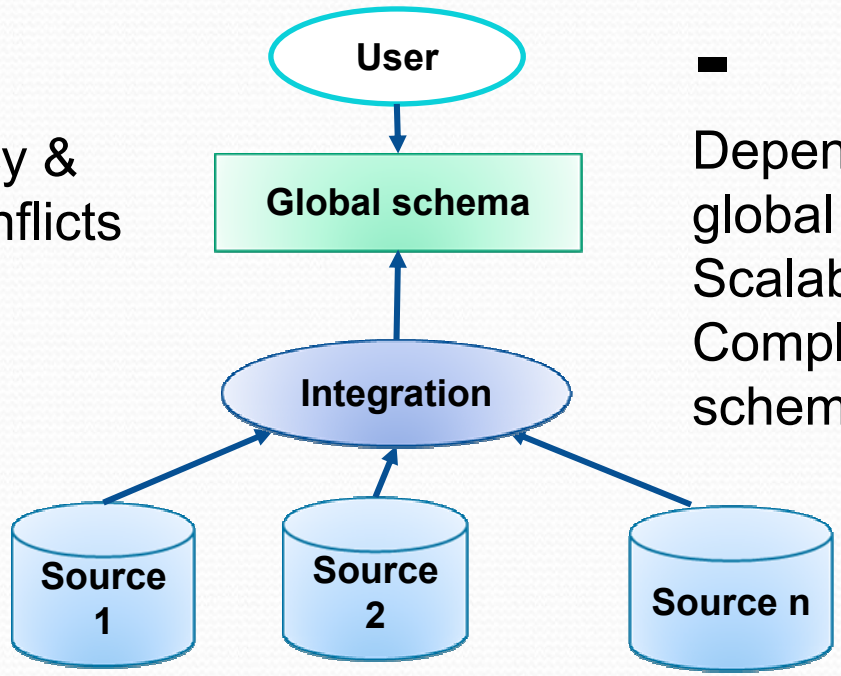
Data Integration Issues

Related Research in Semantic Interoperability for DB is categorized

1. Query-oriented (based on declarative languages or extended SQL)
2. **Mapping-based** (mapping between global & local schemas)

+

Transparency & semantic conflicts resolved



-

Dependancy of particular global schemas
Scalablility
Complexity of building global schema

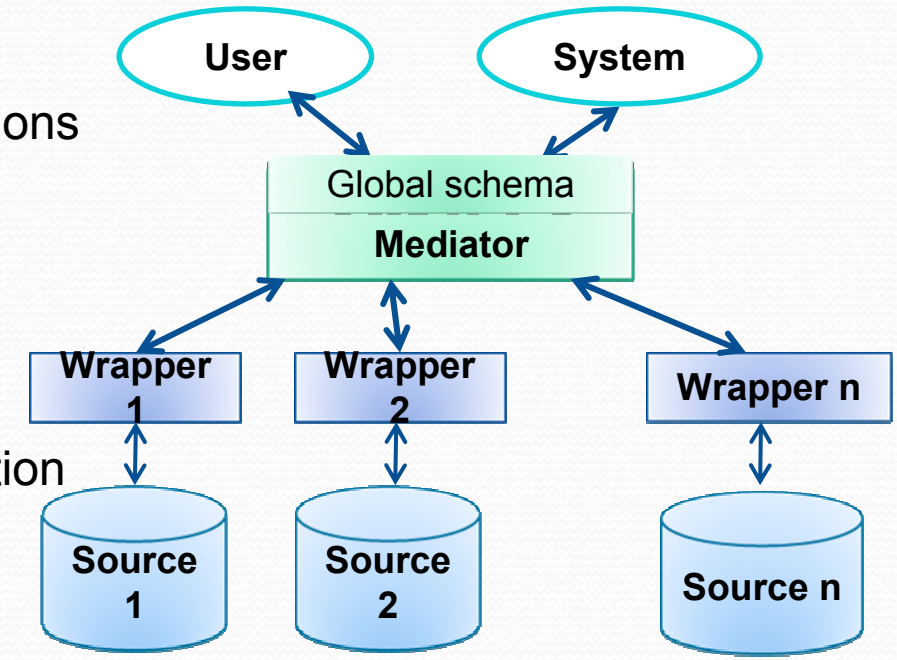
Data Integration Issues

Related Research in Semantic Interoperability for DB is categorized

1. Query-oriented (based on declarative languages or extended SQL)
2. Mapping-based (mapping between global & local schemas)
3. **Intermediary-based** (Mediator-Wrapper)

Mediator :

• Integrates data from different representations (mapping using GAV or LAV)
• Decompose the query
• Re-compose the replies



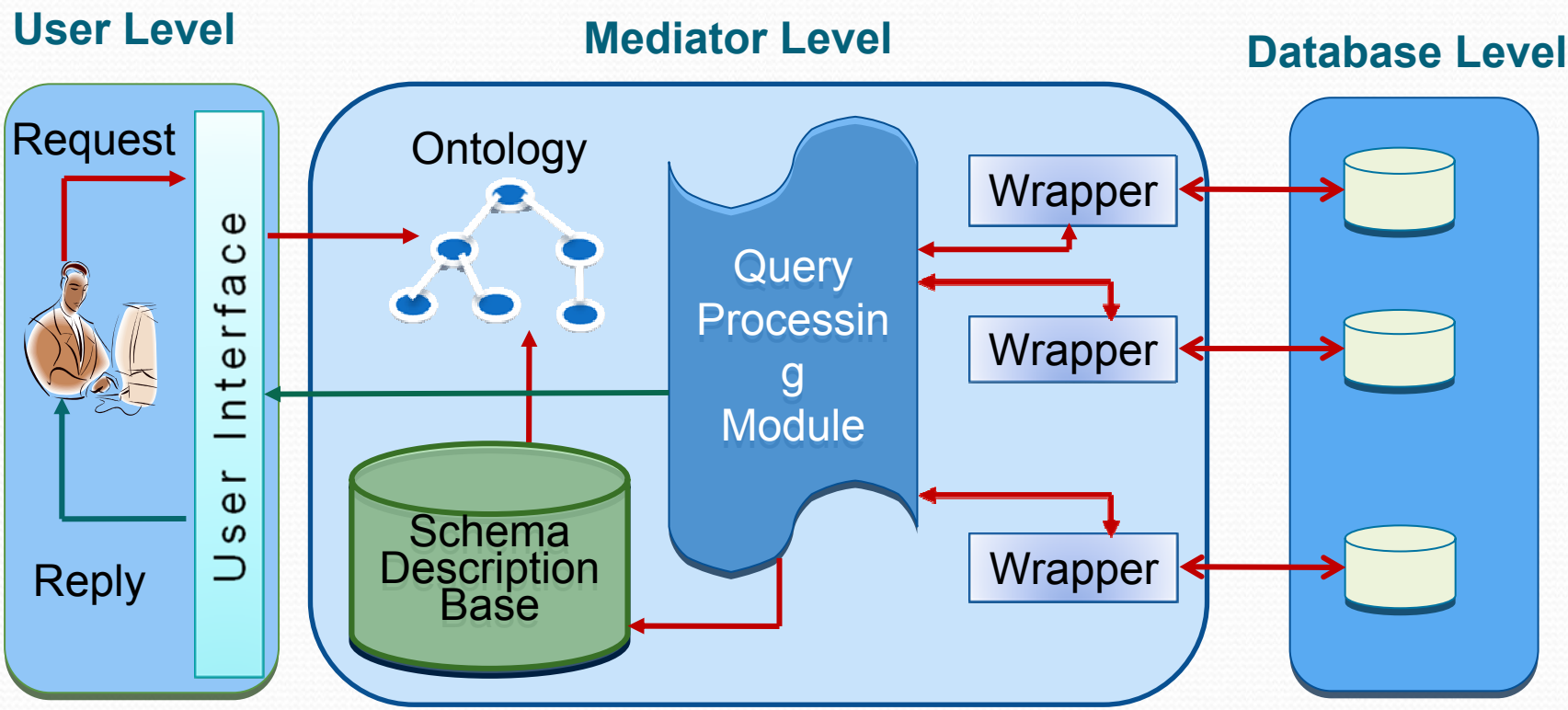
Wrappers convert to common representation Query from mediator & Reply from source.

Our Approach for Data Integration

1. **Intermediary-based** approach (Mediator-Wrapper)
2. Use domain **ontology** to resolve semantic conflicts
3. We have defined "**Schema Description Base"** to store and manage mappings between ontology and sources
4. A user **Query Format** based on ontology concept and similar to SQL.
5. **Algorithms** for localization of the sources, decomposition of the query, re-composition of the replies.

Focused on relational data bases as data sources**.**

Architecture



Architecture

1. User Level

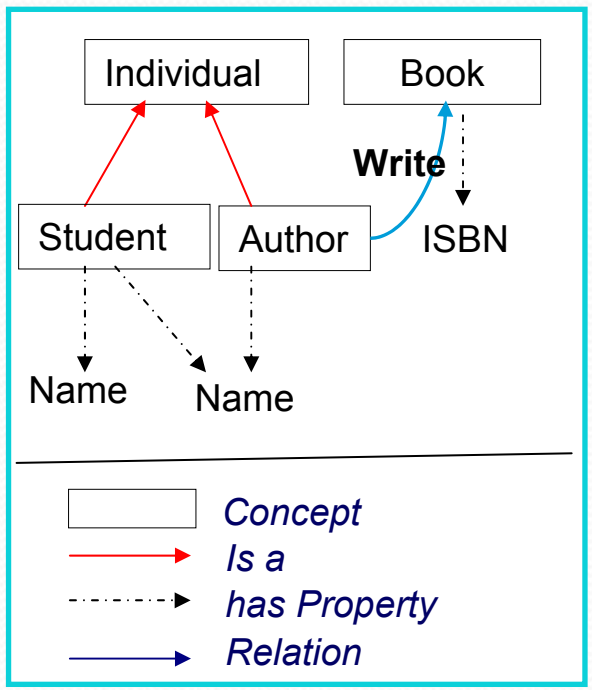
- The user has an interface allowing him to write his requests using ontology concepts.
- The ontology is described with OWL language: concepts, properties and relations.
- The user's request is written in the format:

SELECT [List of properties]

FROM [List of concepts | relation between concepts]

WHERE [List of conditions]

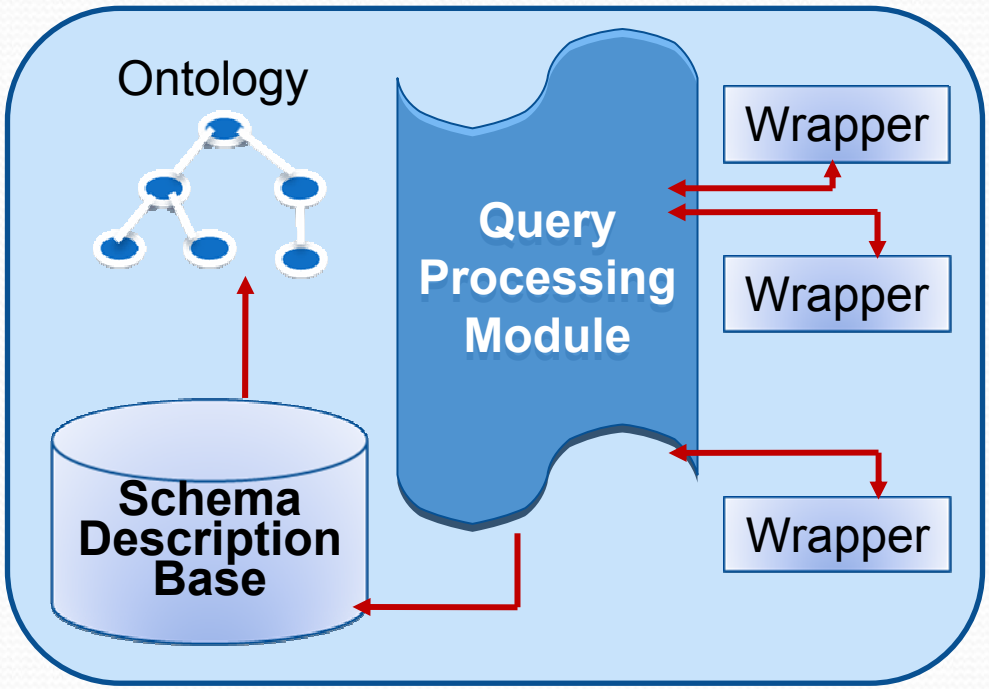
Eg.: SELECT BOOK.ISBN, Author.Name
FROM Book, Author, Write(Book, Author)
WHERE Book.price<100



Example : Domain Ontology

Architecture

Mediator Level

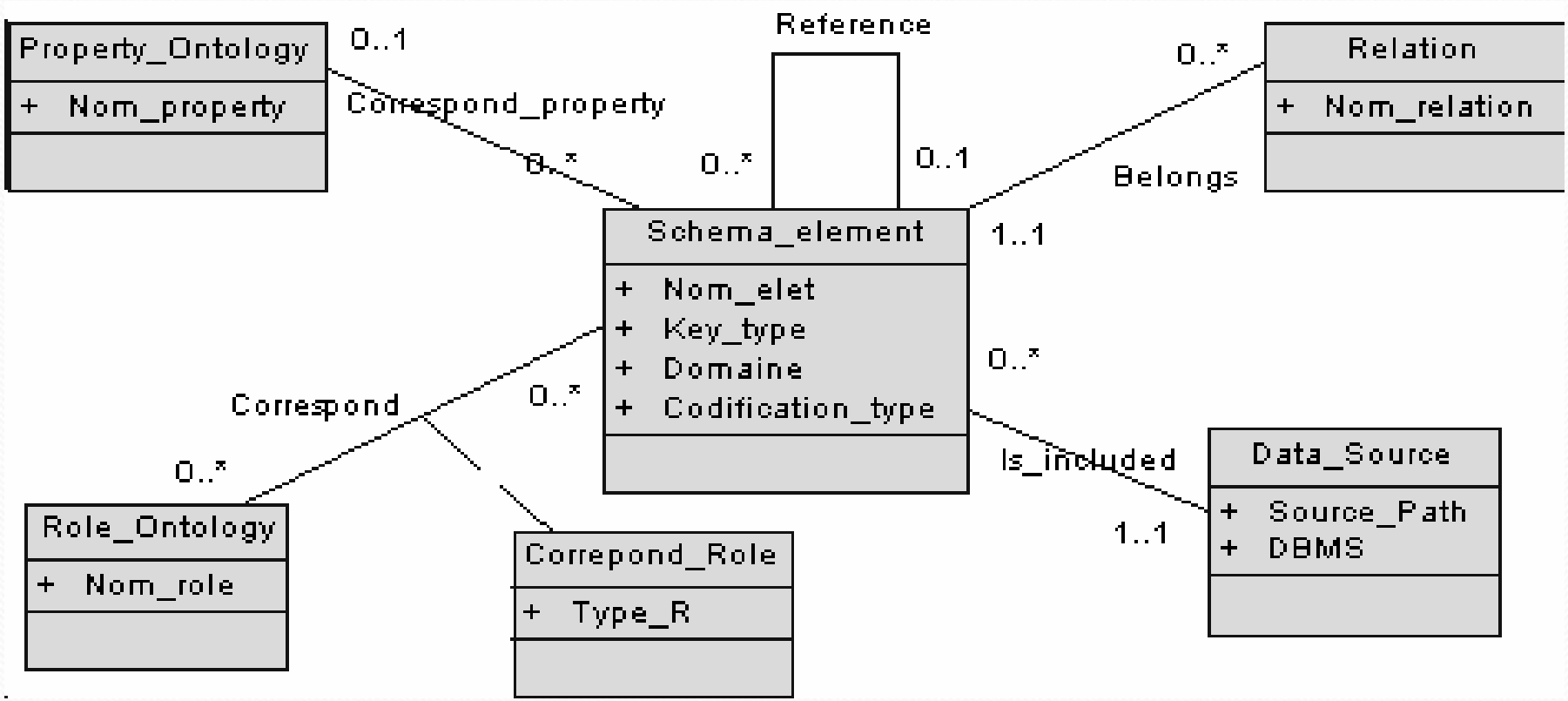


Schema Description Base: Mapping Ontology - Source

- The Schema Description Base is a database that store mappings between ontology and sources.

- In our case, this is done manually by the DBA of each source.
- Our mapping is based on the methodology of building an ontology from a relational DB.
- This mapping can be defined as follows :
  - Every attribute of a schema source can be associate to a property or to a Concept.
  - Every foreign key can be associated to an ontology relation

Schema Description Base



Query Process

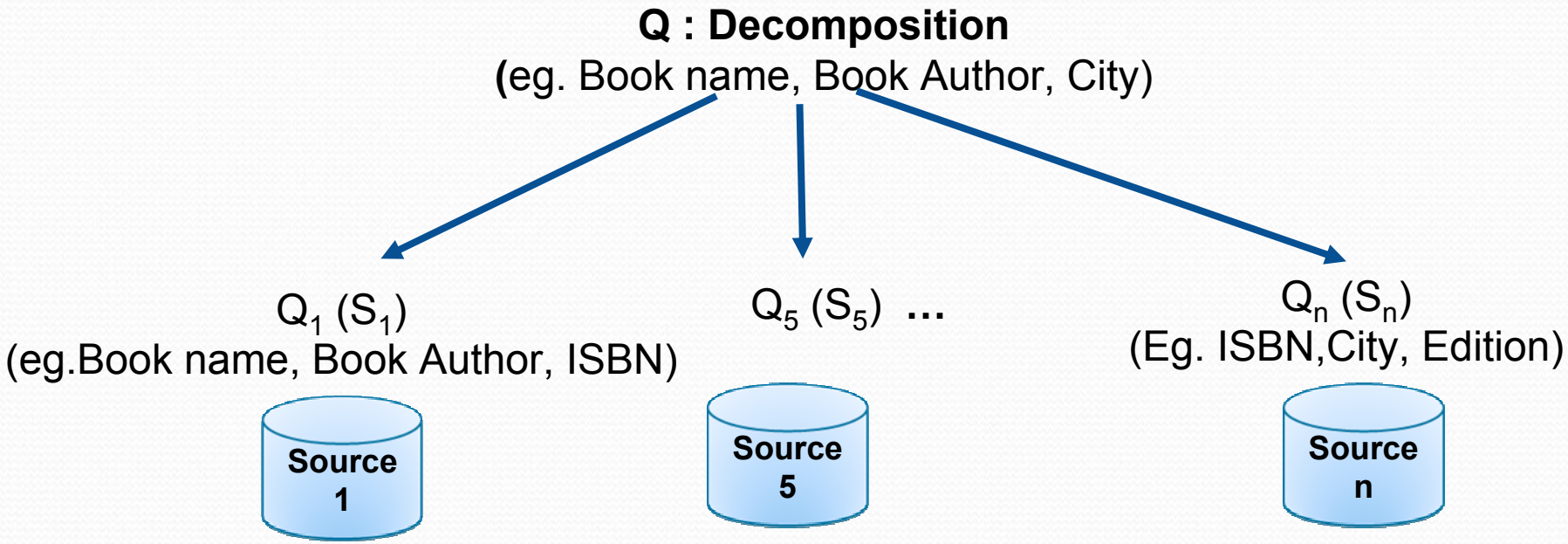
Query Process

1. **Analysis of the global request:**
   - Extracting the different components of the global request
   - Finding equivalent elements in the sources.

2. **Localization of the sources :**
   - Select from the Schema Description Base the sources that provide a partial or complete answer to the global request. Relevant source contain:
     - All the attributes equivalent to the elements of the global request.
     - Partial properties of the global request that can be joined with other attributes of other sources.
     - Some of the properties of the global request.

Query Process

3. **Decomposition and Re-writing of the global request into sub-queries**

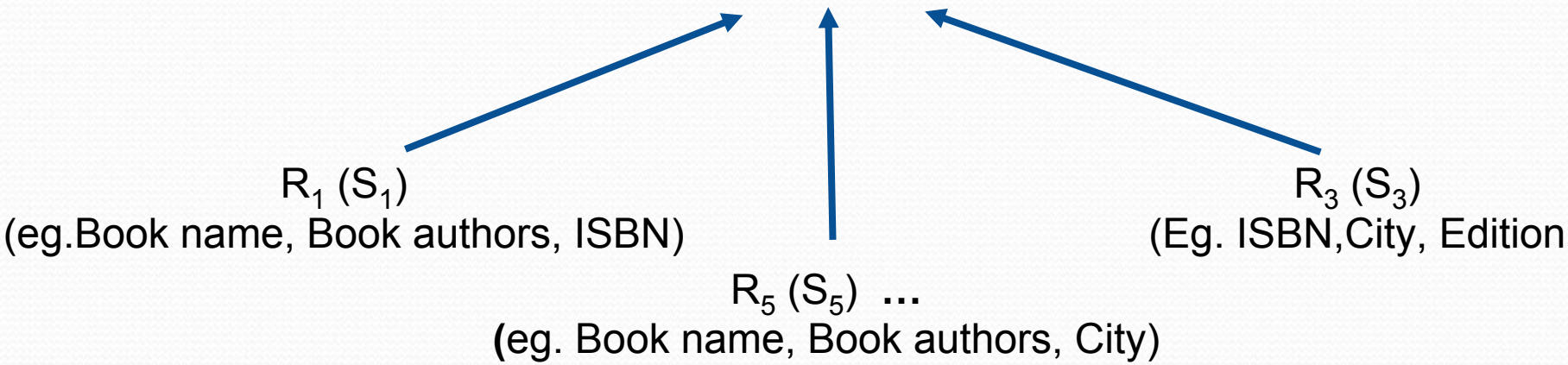


Query Process

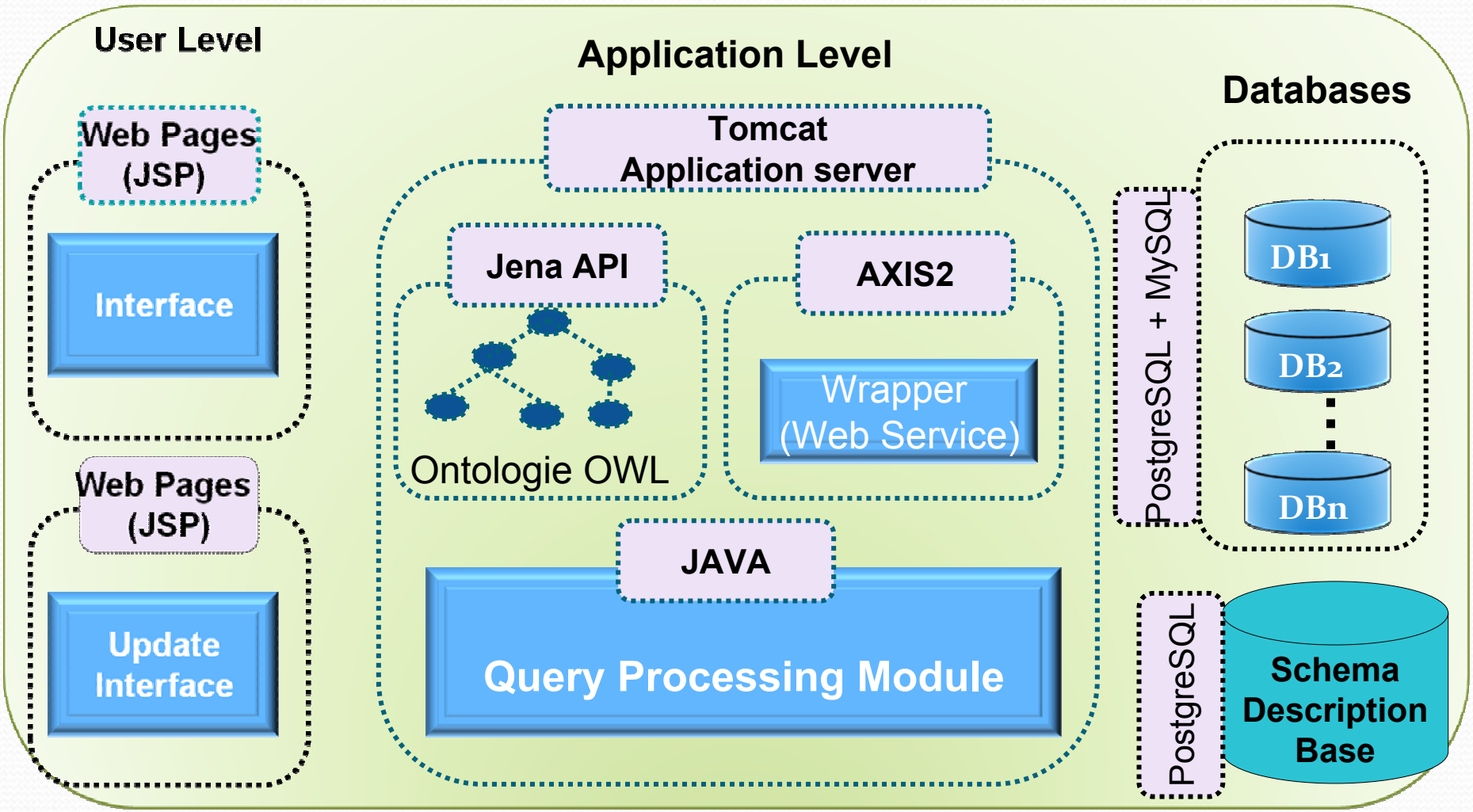
4. **Execution of sub-query**
   - Each sub query is run by each of the local DBMS
   - Wrapper translates the replies generated from the DBMS into a common format for the mediator.

5. **Re-composition of the replies:**

R: Recomposition $R_5 (S_5) \cup (R_1 (S_1) \cap R_3 (S_3))$
(eg. Book name, Book authors, City)



Implementation



Conclusion & Perspectives

- Our approach is based on:
  - **Intermediary-based** approach (Mediator-Wrapper)
  - A shared **ontology** that respects the autonomy of every relational source, and resolve some semantic conflicts.
  - A newly defined concept of "**Schema Description Base"** to find relevant sources.
  - A user **Query Format** based on ontology concept and similar to SQL.
  - **Specific Algorithms** for the Query Processing Module.
- Prototype **Implemented**.

Conclusion & Perspectives

- In our solution, the mapping is done manually for every relational source.
- Our future work, is about :
  - Automating management of mappings
  - Define other criteria for joining sources.
  - Optimize the query process.

Thank You