DB Group @ unimo

The MOMIS approach for Information Integration

Sonia Bergamaschi

www.dbgroup.unimo.it

Dipartimento di Ingegneria "Enzo Ferrari"

Università di Modena e Reggio Emilia, via Vignolese 905, 41100 Modena

www.datariver.it

Spin-off presso Università di Modena e Reggio Emilia

DataRiver

open source data integration

Data Integration

- Modern enterprises are often organized as “virtual networks”, where enterprises operate through inter-enterprise cooperative processes.

- To manage inter-enterprise processes and data exchange a key issue is to mediate among the hetereogeneity of different information systems:

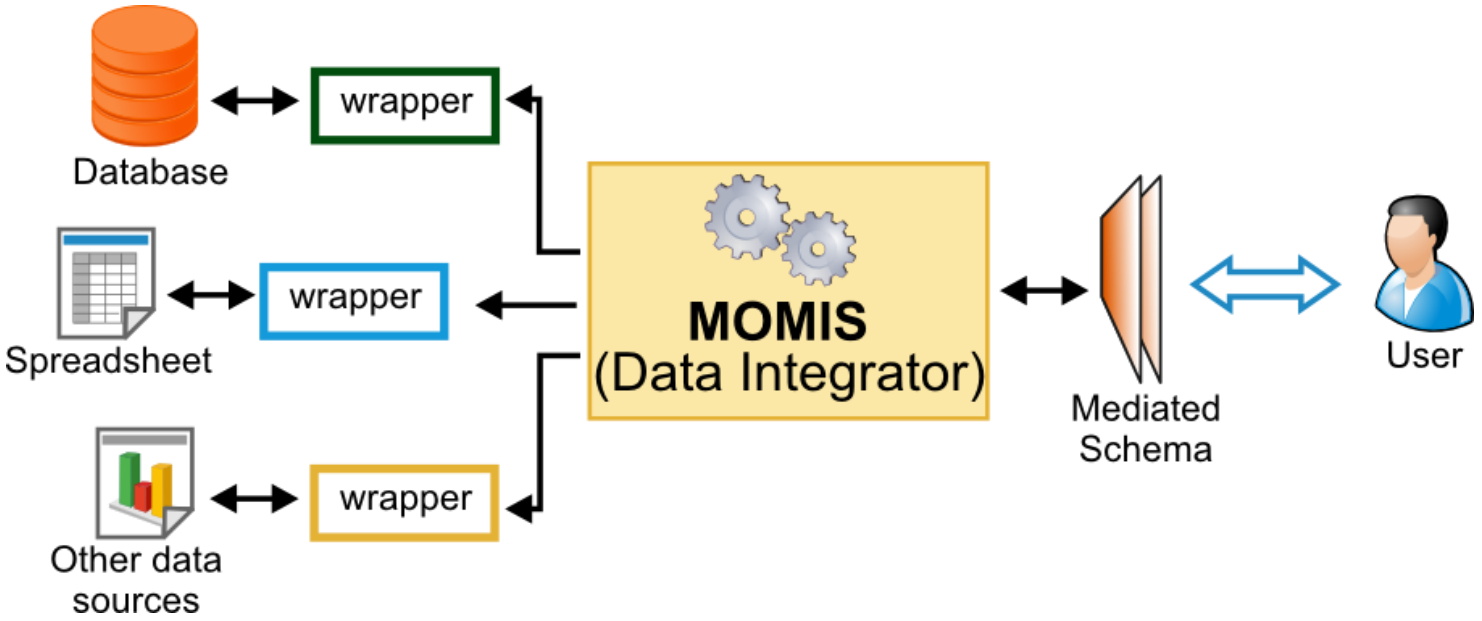
  Data Integration is a technological solution to build a shared and integrated knowledge base.

- Data Integration has to deal with the problems arising from the heterogeneity of data sources:

- **Structural Heterogeneity:**
  - Different data models
  - Same model but different conceptualization chosen

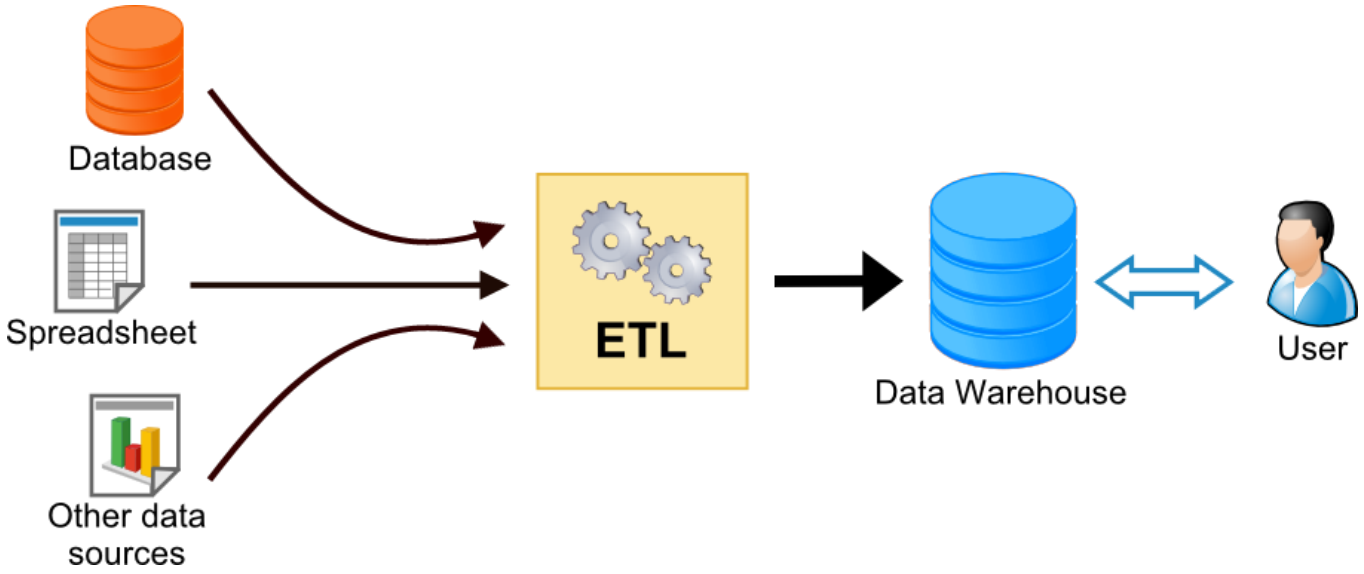
- **Semantic Heterogeneity**:
  different meaning and interpretation
  - Two schemata might use the same term to denote distinct concepts (homonyms), or different terms to denote the same concept (synonyms)

Virtual Data Integration



A Mediated Schema provides an integrated and virtual view of the data stored in several sources. No centralized copy of data is stored, a user query on the mediated schema is transformed into queries over the original sources.

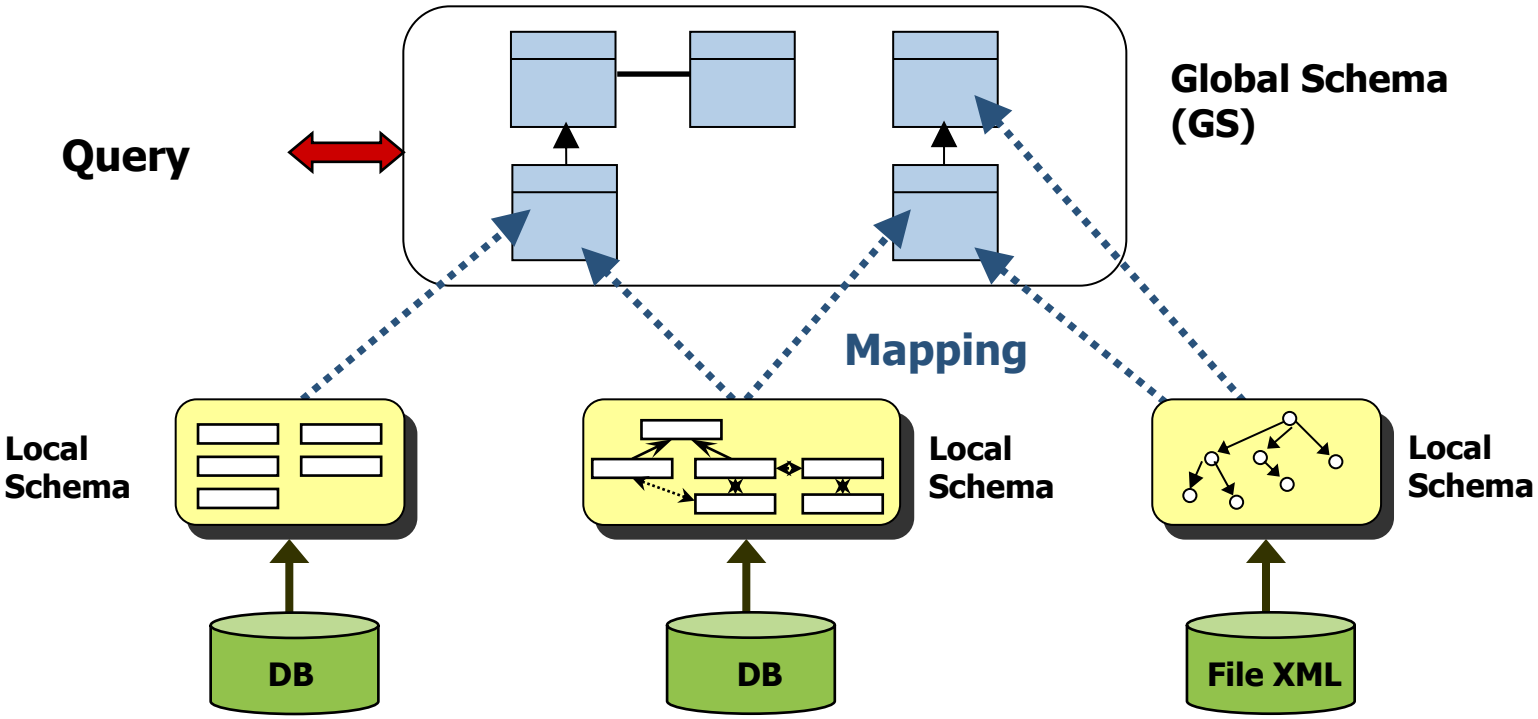
Data Warehouse & Data Integration



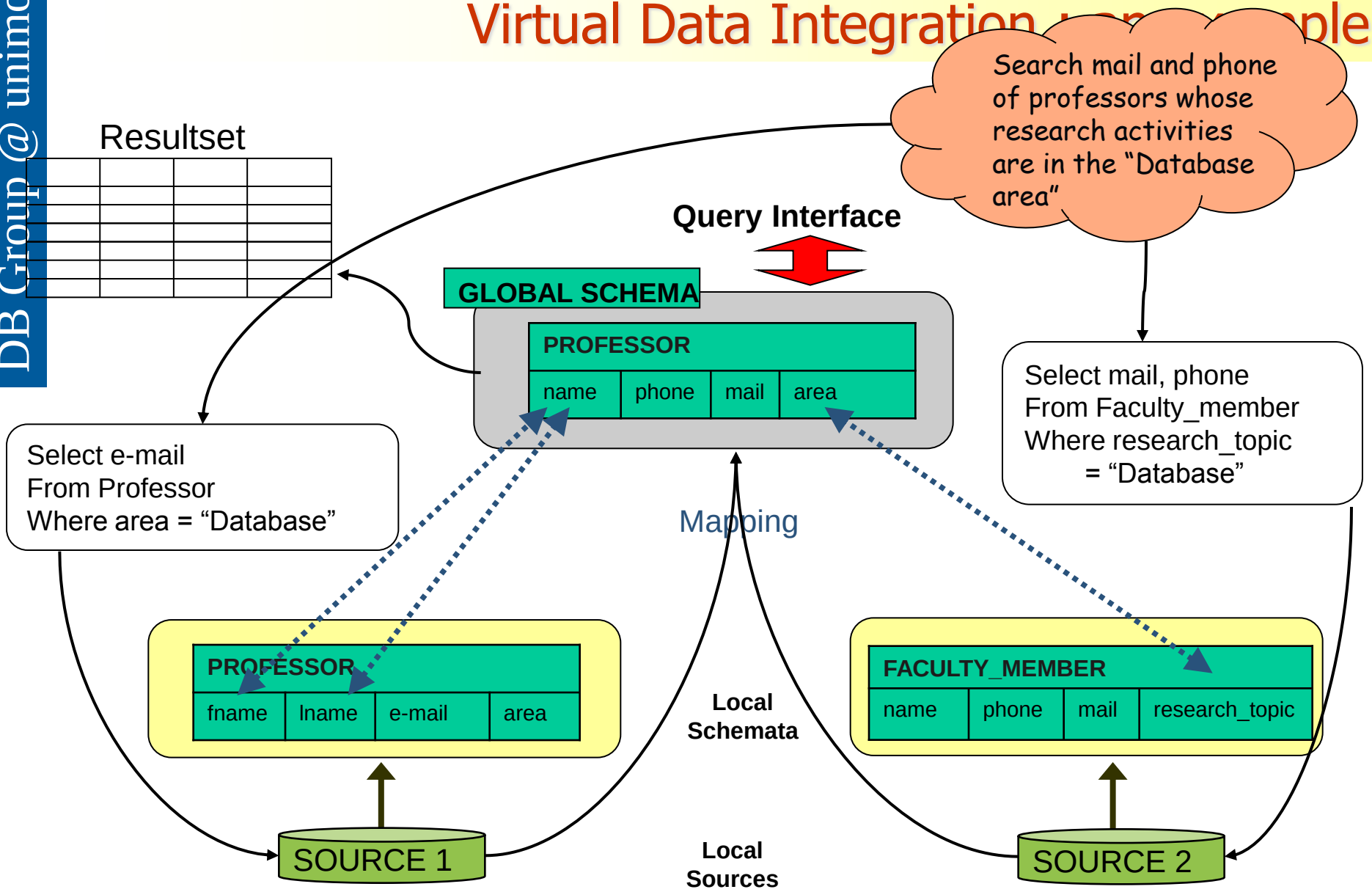
Data from several sources are extracted, transformed and loaded into a Data Warehouse, where users can run queries.

Semantic Integration of Heterogeneous Data

- Data integration provides a Global *virtual* Schema (GS) that
  - is a conceptualization (ontology) describing a set of distributed & heterogeneous data sources
  - allows a user to pose a query and receive a unique answer (transparently from the involved sources)



Virtual Data Integration example



MOMIS

- **MOMIS** (Mediator envirOnment for Multiple Information Sources) is a framework to perform information extraction and integration of heterogeneous, structured and semistructured, data sources developed by the DBGroup at the University of Modena and Reggio Emilia (www.dbgroup.unimo.it/Momis)

- Semantic Integration of Information
  - A common data model ODL_{I^3} (derived from ODL-ODMG and I^3)
  - The local schema of each source is available (source wrapping)

- Tool-supported techniques to construct the Global Schema
  - Local Schema Annotation w.r.t. a common lexical ontology (WordNet)
  - Semi-automatic generation of mappingss among local schemata
  - Clustering techniques
  - Semi-automatic generation of GAV (Global as View) mappings between the GS and local schemata (Mapping Table)
- **Global Query Management**

DB Group

- The **DataBase Group** (www.dbgroup.unimo.it) is the research database group at the Department of Computer Engineering of the University of Modena and Reggio Emilia, it is led by Professor Sonia Bergamaschi and is composed of the following researchers:

  - Sonia Bergamaschi (full professor)
  - Domenico Beneventano (professor)
  - Maurizio Vincini (professor)
  - Francesco Guerra (senior researcher)
  - Mirko Orsini (Phd – CEO of DATARIVER)
  - Laura Po (Phd - researcher)
  - Antonio Sala, Serena Sorrentino (Phd - research collaborator)
  - Fabio Benedetti, Giovanni Simonini(Phd student)
  - Silvia Rota, Dannaoui Abdul Rahman (Phd students)
  - Alberto Corni (Phd - research collaborator)

DB Group research activity

- **Big Data Management & Analysis**
  - NOSQL DBMS
  - Business Intelligence

- **Intelligent Information Integration**
  - to combine data residing at different autonomous sources, and providing the user with a unified view of these data

- **Semantic Search Engines**
  - to augment and improve traditional Web Search Engines by using not just words, but concepts and semantic relationships
  - Keyword Search on Relational Databases
  - Reccomendation Systems

National and International Research Projects

- Project Partecipation: "**D2I** (From Data to Information)" supported by MIUR: "Programma di ricerca scientifica di rilevante interesse nazionale (2000-2001)";
- Project Partecipation: "**Agenti software e commercio elettronico: profili giuridici, tecnologici e psico-sociali**", supported by MIUR "Programma di ricerca scientifica di rilevante interesse nazionale" (2001-2002)
- Project Partecipation: "**Tecnologie per arricchire e fornire accesso a contenuti**" supported by MIUR - Fondo Speciale Innovazione 2000 (2001-2002)
- Project Participation: "**CROSS** " supported by Regione Emilia-Romagna Iniziativa 1.1 PRRIITT(September 2005-2007)
- Project Partecipation: "**WINK** (Web-linked Integration of Network-based Knowledge)" supported by IST-UE RDT (cluster EUTIST-AMI) (2002-2003)

National and International Research Projects

- Project Participation: "**STIL**" supported by Regione Emilia-Romagna Iniziativa 1.1 del Piano Telematico Regionale (September 2005-2007)
- Project Coordination: "**SEWASIE** (SEmantic Web and AgentS in Integrated Economies)" supported by IST-UE RDT(2002-2005)
- Project Coordination: "**WISDOM** (Web Intelligent Search based on DOMain ontologies)" supported by MIUR "Programma di ricerca scientifica di rilevante interesse nazionale" (2005-2007)
- Project Coordination: "**NeP4B** (Networked Peers for Business) MIUR supported by MIUR "Programma Stategico"(2006-2009)- started on July 2006
- Project Participation: "**STASIS** (SofTware for Ambient Semantic Interoperable Services)" (2006-2008) supported by IST-EU RDT - started on september 2006

National and International Research Projects

- Project Participation: BIOGEST-SITEIA lab (2011-2013)

- Project Coordination: "**KEYMANTIC**" founded by the "Fondazione Cassa di Risparmio di Modena (http://www.fondazione-crmo.it/), whose aim is the development of a keyword-based query engine supporting users in querying data sources with complex and large schemas (2009-2011)

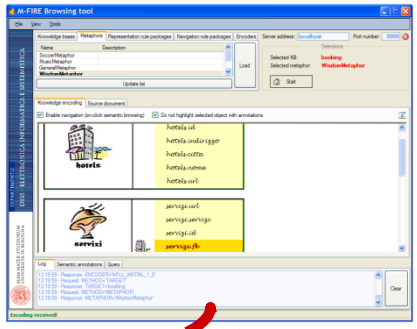
- Project Participation: "**FACILITATE** (SofTware for Ambient Semantic Interoperable Services)" (2010-2012) supported by IST-EU RDT

- **MOMIS** has been already tested in the above mentioned research projects for the development of Vertical Web Portals and the integration of heterogeneus data sources in many domains:
  - Tourism (vertical web portal - WISDOM)
  - Textile (search engine - SEWASIE)
  - Mechanical (search engine - SEWASIE)
  - Logistics (logistic domain ontology - STIL)
  - Agro-Food (data integration for cereals breeding - CEREALAB)
  - Commercial (business intelligence - CROSS)

- **MOMIS** provides methods and tools for:
  - sharing legacy systems in an integrated information system
  - safeguarding the autonomy of systems and organizations
  - support the enterprise interoperability

MOMIS: Tourism domain

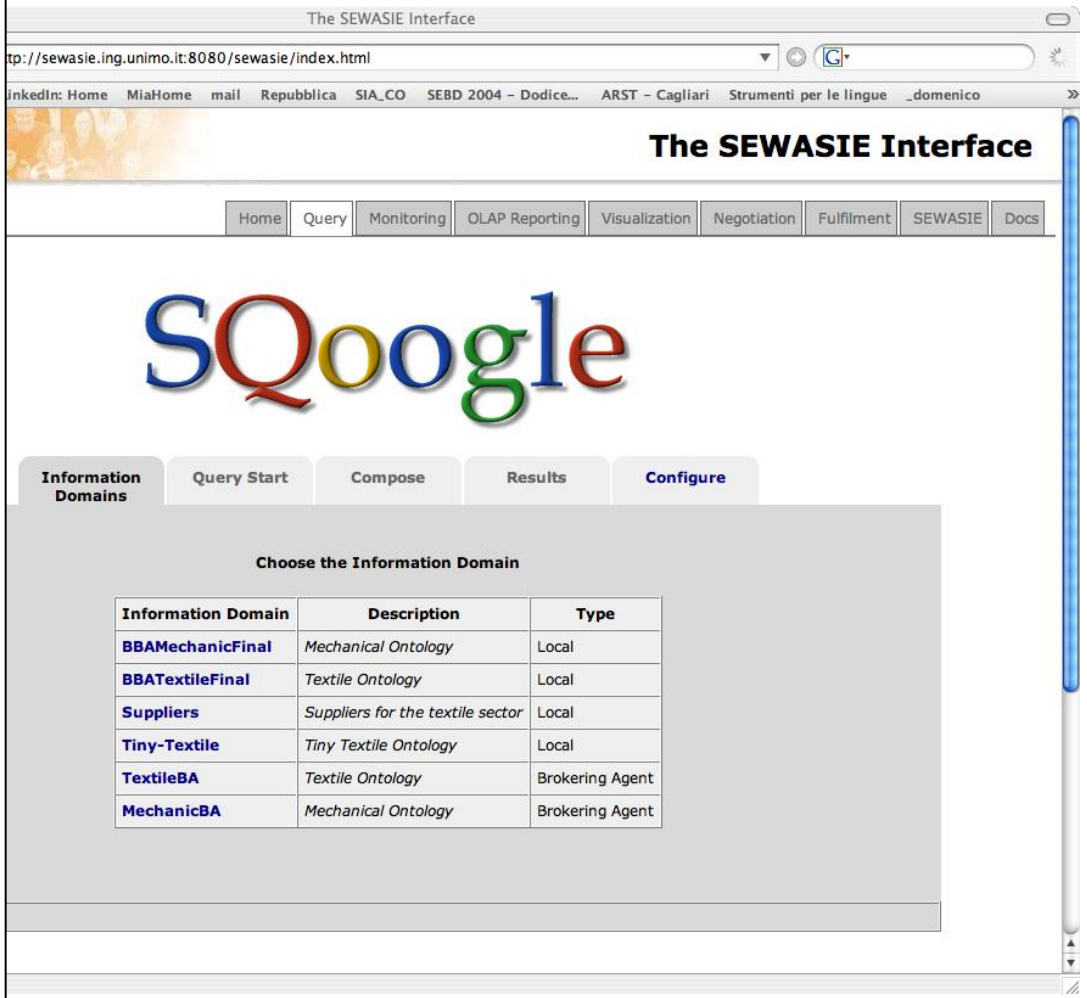
- In the **WISDOM** project (**W**eb **I**ntelligent **S**earch based on **DOM**ain ontologies) (www.dbgroup.unimo.it/wisdom) the **MOMIS** system was exploited for the integration of several tourism web sites and the development of a Tourism Vertical Web Portal



...

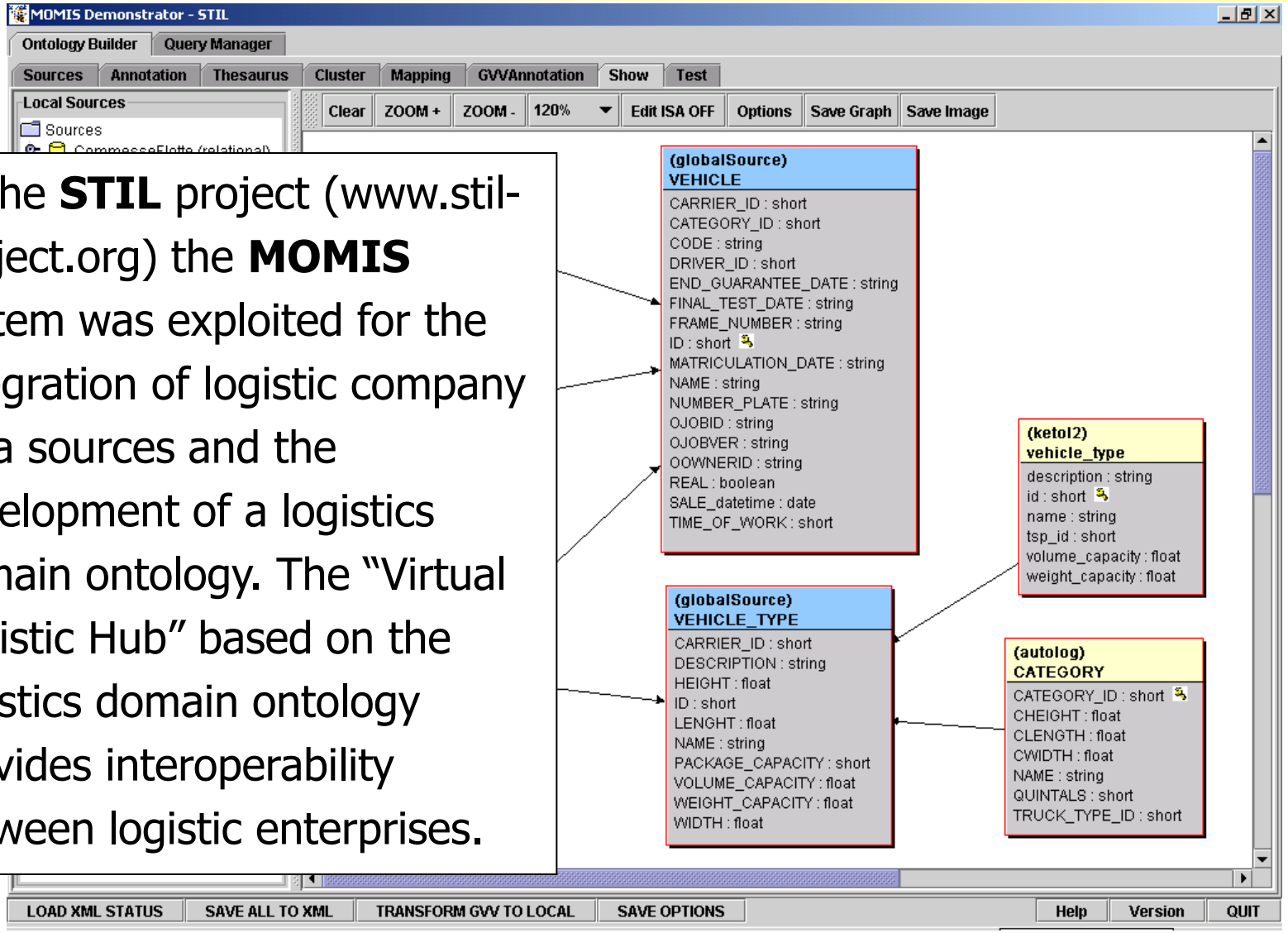
MOMIS: Textile and Mechanical domains

- In the **SEWASIE** project (**SE**mantic **W**eb and **A**gent**S** in **I**ntegrated **E**conomies) (www.sewasie.org) the **MOMIS** system was exploited for the integration of heterogeneus company data sources and the development of a Semantic Search Engine

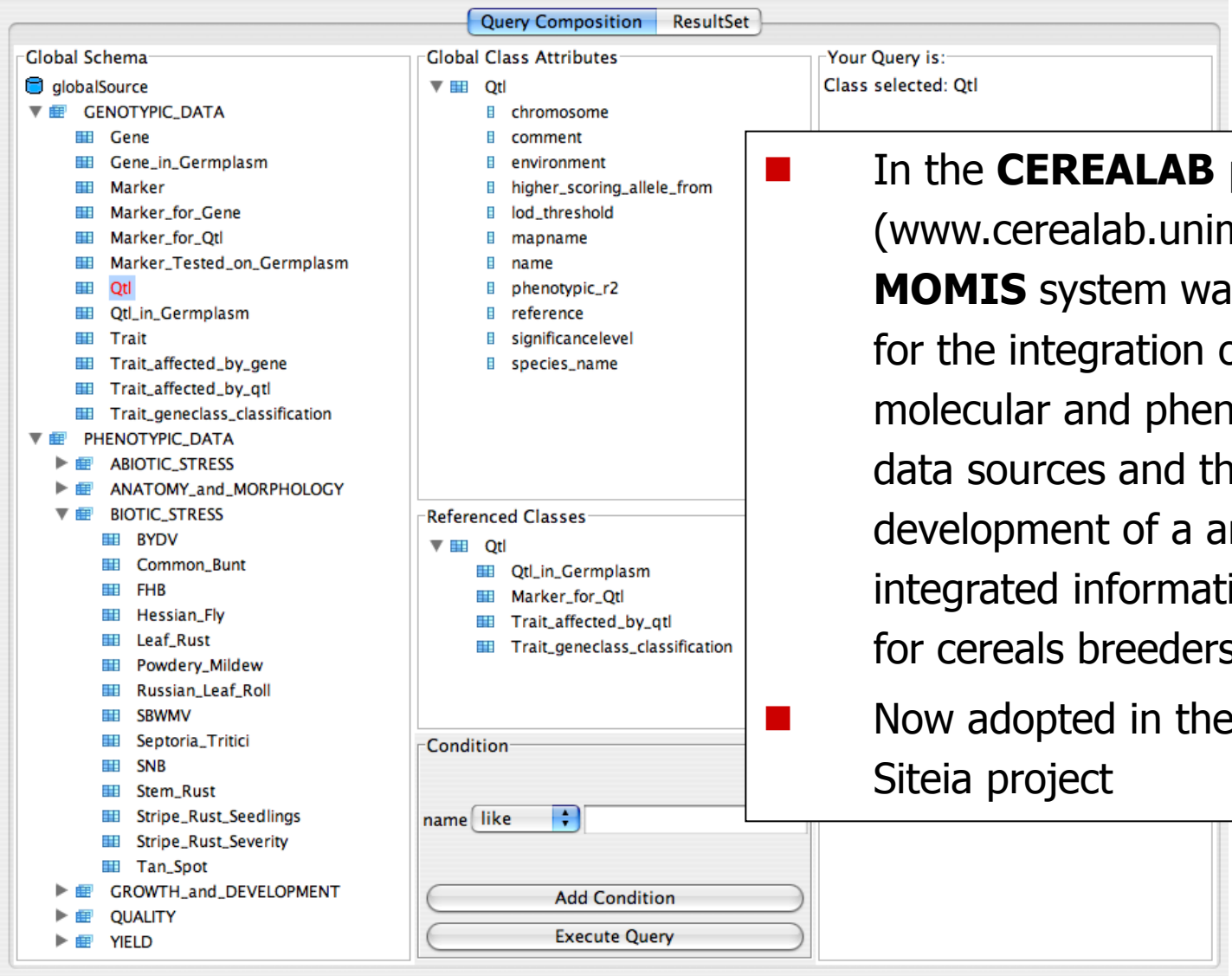


MOMIS: Logistics domain

- In the **STIL** project (www.stil-project.org) the **MOMIS** system was exploited for the integration of logistic company data sources and the development of a logistics domain ontology. The "Virtual Logistic Hub" based on the logistics domain ontology provides interoperability between logistic enterprises.



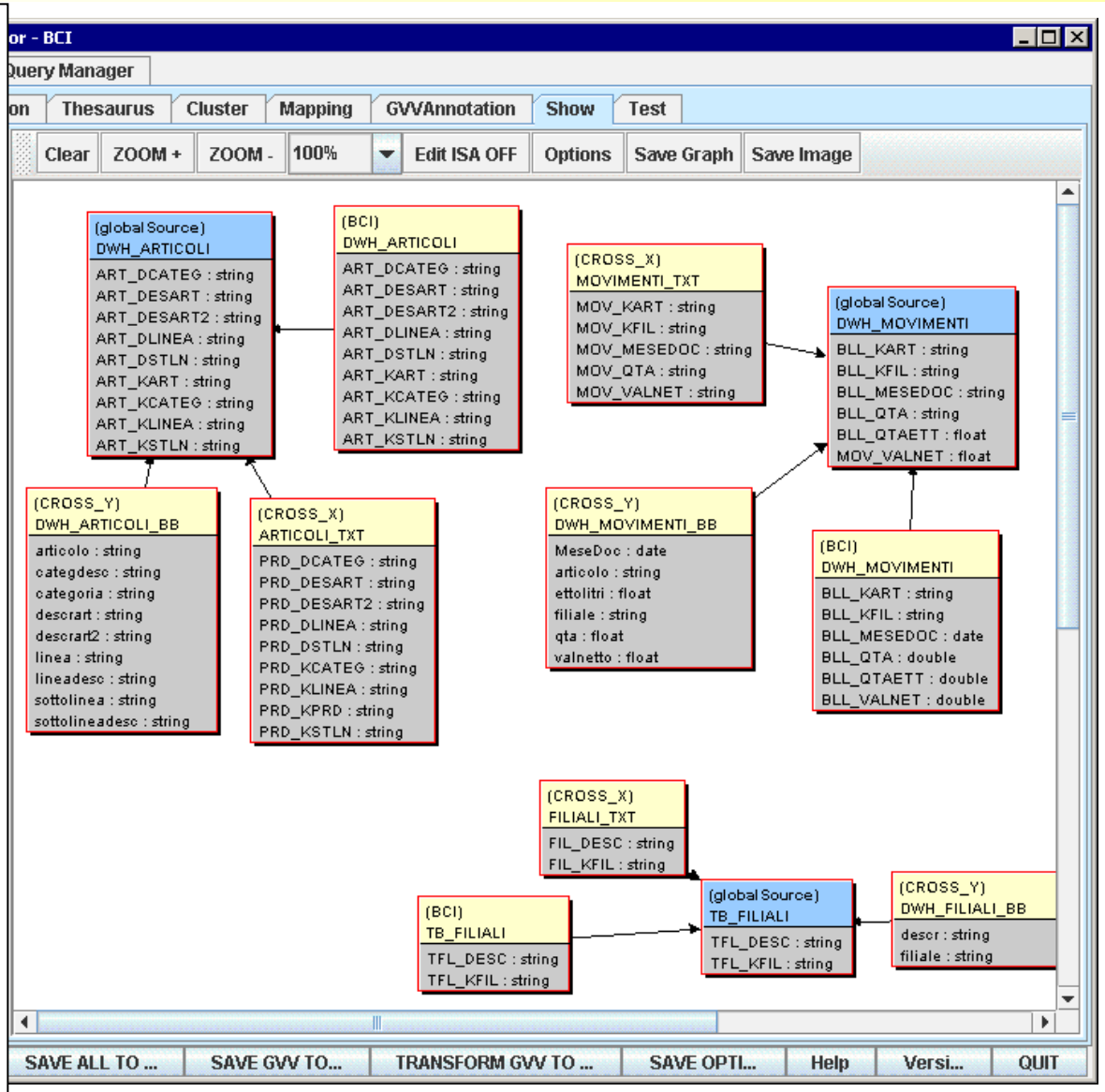
MOMIS: Agro-Food domain



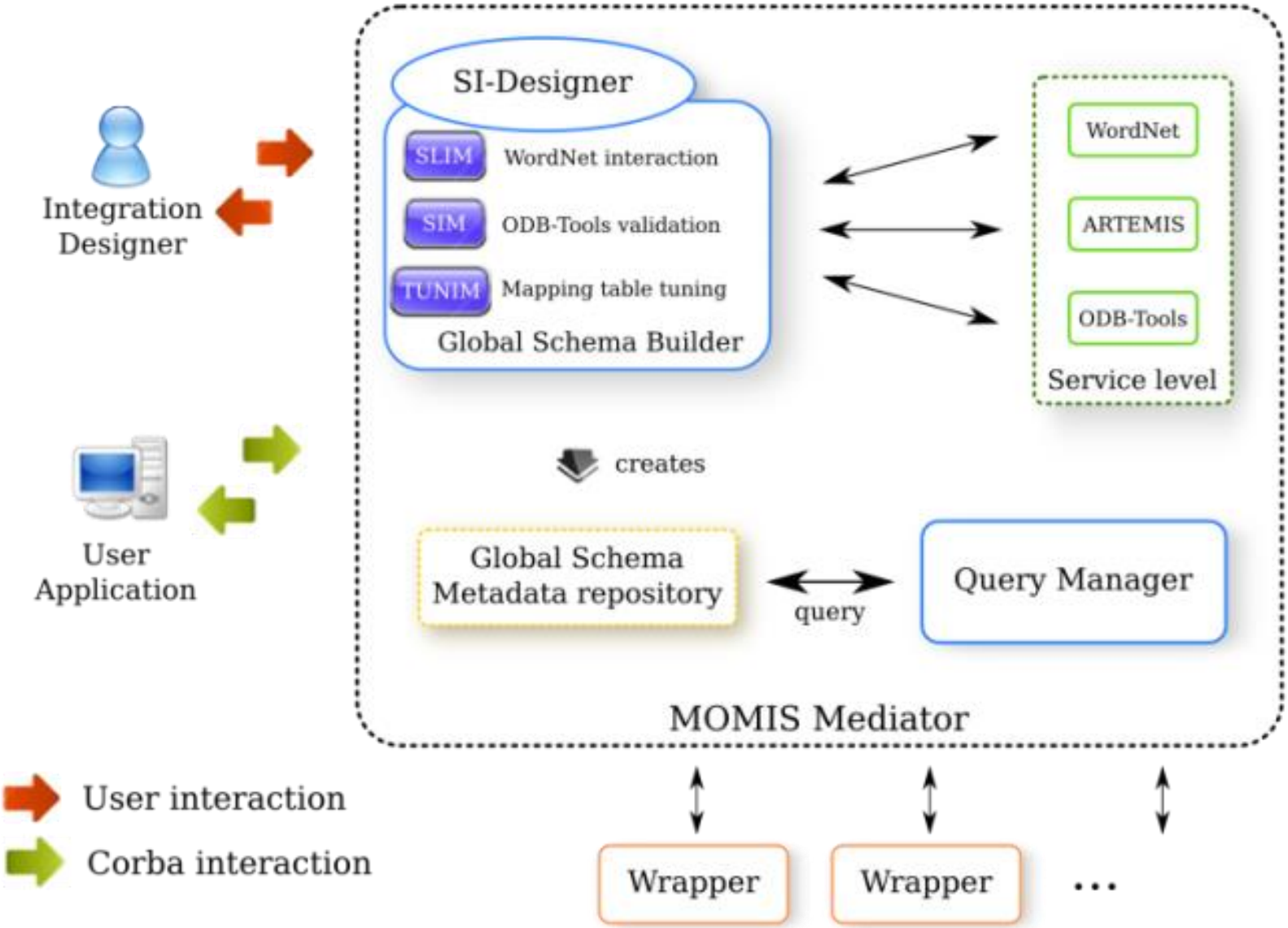
- In the **CEREALAB** project (www.cerealab.unimore.it) the **MOMIS** system was exploited for the integration of molecular and phenotypic data sources and the development of a an integrated information system for cereals breeders.
- Now adopted in the Biogest-Siteia project

MOMIS: Commercial domain

- In the **CROSS** project (www.cross-lab.it) the **MOMIS** system was exploited for the integration of heterogeneus sales order data sources and the population of a Data Warehouse in a business intelligence environment.



MOMIS Architecture



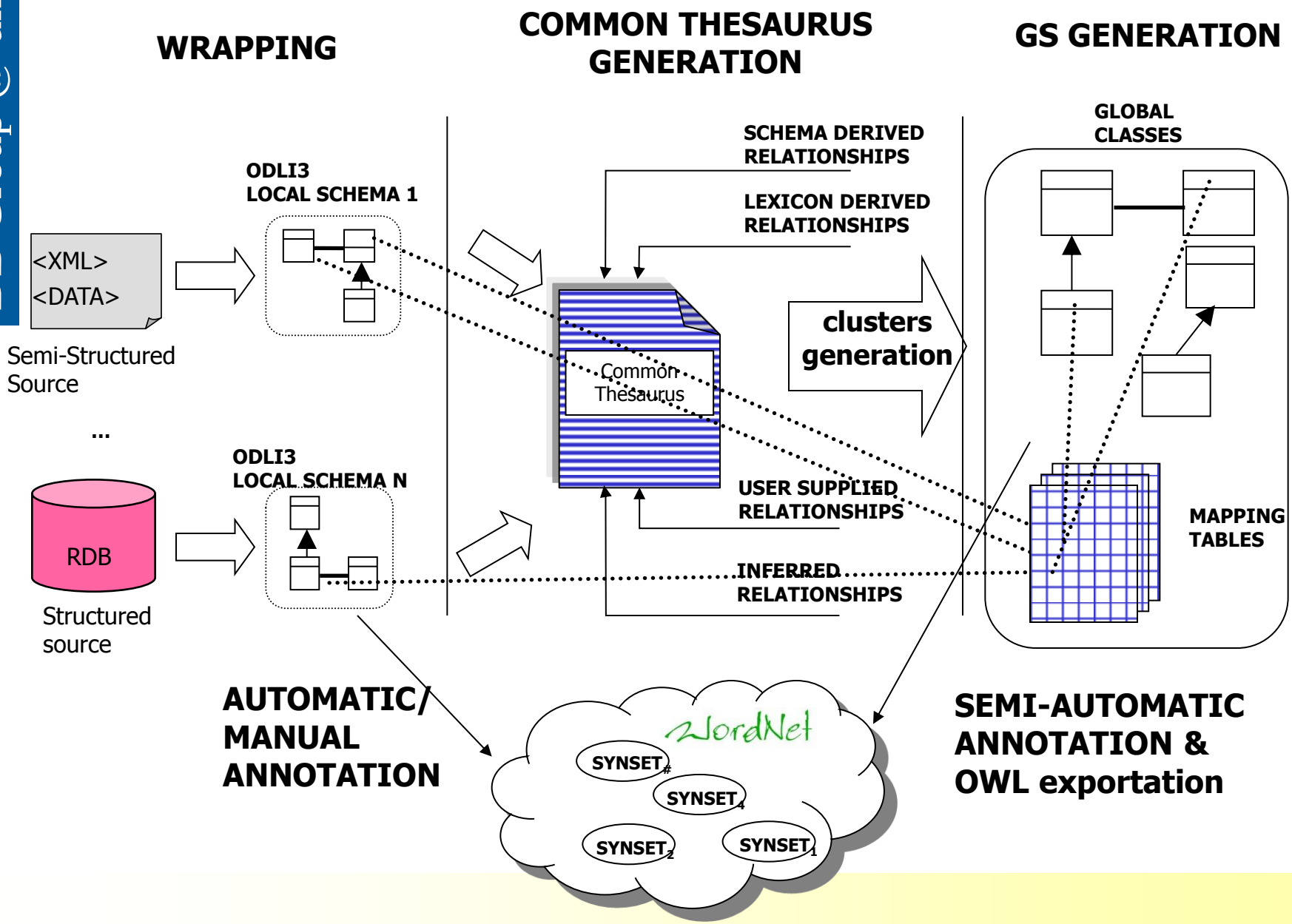
1. Global Schema and Mapping generation

→ the Global Schema is generated and its mappings with the data source are defined

2. Querying the Integrated Data

→ the integrated data can be queried by users and software applications

Overview of the GS generation process



Common Thesaurus

- Set of intensional and extensional relationships expressing intra-schema and inter-schema knowledge
  - **Intensional Relationships**
    between class and attribute names (T)
    < Ti SYN Tj > *Synonymy*
    < Ti NT Tj > *(Narrower Term - NT)*
    < Ti RT Tj > *(Related Term - RT)*
  - **Extensional Relationships** - between classes (*C*)
    the instances of C1 are ...
    <C1 SYN_{Ext} C2> : ... the same instances of C2
    <C1 NT_{Ext} C2> : ... a subset of the instances of C2
    <C1 DIS_{Ext} C2> : ... disjoint from the instances of C2
- **Common Thesaurus generation**:
  (1) *schema derived relationships*
  (2) *lexicon derived relationships*
  (3) *designer supplied relationships*
  (4) *inferred relationships (exploiting ODB-Tools capabilities)*

Lexicon-derived Relationships

- Extracted from the WordNet thesaurus (lexical ontology)

- In WordNet:
  - Word forms are organized in synonym sets (*synsets*)
  - Semantic relationships between *synsets* (meanings)
    - Hyponymy (Hypernymy)
    - Meronymy
    - Correlation (between synsets having the same Hypernym)

- Relationships between class and attribute names are obtained using the WordNet semantic relationships as follows:
  - Synonymy ⇒ SYN
  - Hyponymy ⇒ NT
  - Meronymy and Correlation ⇒ RT

Annotation and Lexicon-derived Relationships

(Narrower Term)
NT

	Word form		
Meaning (synset)	*Book*	*Volume*	*Publication*
a written work or composition that has been published (printed on pages bound together)	✕		
physical objects consisting of a number of pages bound together; "he used a large book as a doorstop"	✕	✕	
the amount of 3-dimensional space occupied by an object		●	
a copy of a printed work offered for distribution			✕

Hyponymy (is a kind of)

Lexicon derived relationships

Book	SYN	Volume
Book	NT	Publication

Extending WordNet

- **WordNet Editor**

- If a class or attribute name has no correspondent in WordNet, the designer may add a new meaning and proper relationships to the existing meanings.

- The designer may add a new meaning (for an existing word-form or for a new one) by:
  - writing the gloss explicitly, or
  - using an existing synset chosen among a list of candidates obtained by an explicit search (using one or more keywords) or by exploiting similarity search techniques.

- The designer may add relationships for the new synset
  - Related synsets are obtained by an explicit search (using one or more keywords) or by exploiting similarity search techniques.

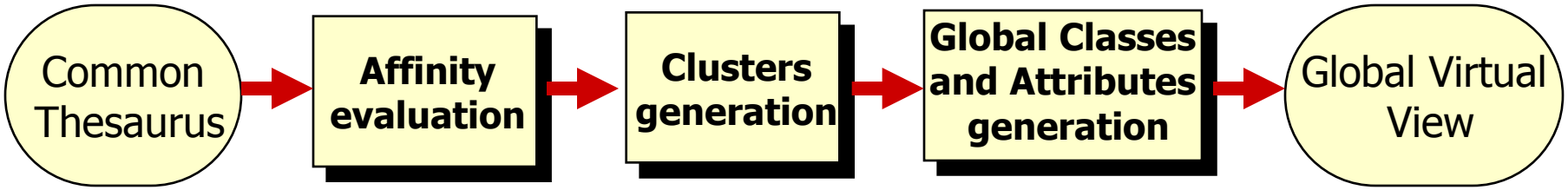
Common Thesaurus Generation: Other rules

- **Schema-derived relationships**
  - RT relationships derived from foreign keys in a relational schema
  - NT relationships from inheritance in a object-oriented schema
  - NT relationships from couples IDs and IDREFs in XML data files
  - ...

- **Inferred relationships**
  - Exploiting Description Logics techniques (by using ODB-Tools) a new set of relationships are inferred

- **Designer supplied relationships**
  - The designer can add/delete relationships to the Common Thesaurus

Global Virtual View and Mapping Table Generation



- **GS generation :**
  A global class C=(**L**,**GA**) is generated for each cluster :
  - **L** are the local classes of the cluster
  - **GA** are the global attributes of C
    - Union of the local attributes
    - Fusion of "similar attributes" (by using the Common Thesaurus)

- **MT generation :**
  For each global class C=(**L**,**GA**), a *Mapping Table* (MT) is generated, to represent the mappings between global and local attributes
  - MT is a table **GA**x**L** : An element *MT[GA][L]* represents the attributes of the local class *L* mapped into the global attribute *GA*.

GS and MT generation : example

- Cluster **Company={prontocomune.Azienda,fibre2fashion.Company ,usawear.Company}**

- Mapping Table of C

	prontocomune. Azienda	fibre2fashion. Company	usawear. Company
Name	Nome	Name	CompanyName
Address	Indirizzo	Address	Address
Description		AboutUs	Description
Category	Categoria	Category	
Phone	Telefono	Tel	Phone

- **MT generation :**
  Since " AboutUs SYN Description" is in CT, these local attributes are "fused" into to the same global attribute "Description"

- **GS annotation :**
  - the name and the meaning of the class Company correspond to the name and the meaning of fibre2fashion.Company (the most general class)

Mapping Table Refinement

- Global-As-View (**GAV**) approach:
  the GS is expressed in terms of the local schemata

- **Global-as-View (GAV) mappings**:
  for each global class C we define a **view** $\mathbf{V_C}$ over the local classes of C.

- The integration designer, supported by the Ontology Builder graphical interface, can implicitely **define** $\mathbf{V_C}$ by the Mapping Table refinement:

1. ***Data Transformation :*** *converting data from local source data formats into a global schema format (Conversion Functions)*
2. ***Data Fusion :*** *fusing records representing the same real-world object into a single, consistent, and clean record:*
   1. ***Object Identification***
   2. ***Data Reconciliation***

Data Transformation: THALIA Benchmark

- THALIA: **T**est **H**arness for the **A**ssessment of **L**egacy information **I**ntegration **A**pproaches

  public available testbed and benchmark for information integration systems

  provides over 40 downloadable sources representing University course catalog from computer science around the world

  systematic classification of the different types of syntactic and semantic heterogeneities described by the twelve queries provided

- MOMIS Data Transformation can deal with all the twelve queries of the THALIA benchmark by using a simple combination of declarative translation functions and without the overhead of new code.

THALIA's Query 2 example

Q2: 'Find all database courses that meet at 1:30pm on any given day'

***Complex Mappings**:* Mapping between the Time attribute of Carnegie Mellon University and the Times attribute of University of Massachusetts.

Course Mapping Table

Course	Course (cmu)	Course (umb)
CourseTitle	CourseTitle	TitleCredits
Time	Time	MDTF[Time] [umb.Times]

```
MDTF[Time][umb.Times] =

CASE WHEN ISNUMERIC(SUBSTRING(Times, 1, 2)) = 1
     THEN CASE WHEN CAST(SUBSTRING(Times, 1, 2) AS int) > 12
           THEN CAST(CAST(SUBSTRING(Times, 1, 2) AS integer)- 12 AS nvarchar(2))
               ELSE SUBSTRING(Times, 1, 2)
               END
     + SUBSTRING(Times, 3, 4) +
          CASE WHEN CAST(SUBSTRING(Times, 7, 2) AS int) > 12
THEN CAST(CAST(SUBSTRING(Times, 7, 2) AS integer)- 12 AS nvarchar(3))
               ELSE  SUBSTRING(Times, 7, 2)
               END
     + SUBSTRING(Times, 9, 3)
END AS Time
```

Mapping Refinement: Data Conflicts Resolution

- **Data Conflicts :** the same attribute from one or more sources do not agree on its value

1) **Uncertainty :** it is a conflict between a not-null value and one or more null values that describe the same attribute of the same object
2) **Contradictions :** it is a conflict between two or more different not-null values that describe the same attribute of the same object.

- **Example**: data contradictions on the `Phone` attribute

L1

Name	Address	Phone
RAMOTEX	…Mirpur-1216Dh	+390828015393
CASTORAMA	…Casalecchio (BO)	+390516113011

L2

Name	Address	Phone
RAMOTEX	…Mirpur-1216Dh	880-5-801466
Koramsa Corp	…Guatemala City	+502 439 6868

Data fusion: Full Join Merge

- What operator for Data Fusion ?
- **Full Join Merge** Operator
  - **Full Join** : to include into the result *all tuples of all local sources*
    - Computed on the basis of the Object Identification/Join Conditions
  - **Merge** : to perform data reconciliations
    - Application of Resolution functions (including all the results)
- In MOMIS the **Full Join Merge** is the *default* operator, i.e., is *implicitly defined* by using the Ontology Builder graphical interface (see next slide)
- The designer can change this default operator to other join operators (inner join, left/right join)

From the Mapping Table to the Full Join Merge

Mapping Table of the Global Class Hotel = {resort, hotel}

Resolution Functions

SUM

Join Conditions

Object identifier

	resort	hotel
Name	name	name
Room	rooms	hotelrooms
Price	amount	price
Star	star	
Wifi		wifi

Resolution Functions

AVG

```
Select Name,
 AVG(L1.amount, L2.price) as Price,
 SUM(L1.rooms,L2.hotelrooms) as Room
…
    from resort L1 full join hotel L2
       using (name)
```

Full Join

Full Join Merge

Data Integration and Data Fusion: an example

Global Class **G** = {**L1**, **L2** }

G	L1	L2
ID	ID	ID
A	A	
B		B
C	C	C

Data Fusion

G as Full Join Merge of L1 and L2

```
SELECT   ID,
         L1.A  AS A,
         L2.B  AS B,
         AVG (L1.C,L2.C) AS C
FROM     L1 FULL JOIN L2
         USING (ID)
```

result

ID	A	B	C=AVG(L1.C,L2.C)
1	3	4	3
2	3		1
3			3
4	8		3
5			

integration

L1

ID	A	C
1	3	4
2	3	1
3		3
4	8	3

L2

ID	B	C
1	4	2
2		
3		3
5		

Global Query Management

- **The querying problem**:
  How to answer queries expressed on the GS (**global queries**)?

- In a Virtual Data Integration system, data reside at the data sources then the query processing is based on **Query rewriting** :
  to rewrite a global query as an equivalent set of queries expressed on the local schemata data sources (**local queries**).

- **GAV** approach: query rewriting is performed by **unfolding,** i.e. by expanding a global query on C according to the **view** associated to C

  - When the view is defined with an *outer-join merge* operator, the query rewriting performs the fusion (object identification and conflict resolution) of the local answers into the global answer.

- **Query Manager**

  - Distributed Query Processing

  - Query Optimization

Query unfolding: Predicate push down

global constraint on *one-to-one attributes* can be push down on local queries

Global Query

Query 1:

```
Select Name,  Room
from  Hotel
where Price = 100 and Stars >  3
```

Hotel	resort	hotel
Name	name	name
Room	rooms	hotelrooms
Price	amount	price
Star	stars	-
Wifi	-	wifi

Local queries

Q1 to local source "resort".

Q1_resort:

```
Select name, amount,  rooms
from  resort
where stars >  3
```

Q1 to local source "hotel".

Q1_hotel:

```
Select name, price,
hotelrooms
from  hotel
```

Query unfolding: Full Outer Join simplification

The local answers (Q_Li) are fused into the global answer on the basis of the Full Outer Join-merge operation:

Q_L1 full join Q_L2 on JC(L1,L2)

✓**Full Join simplification**:

(1) FOJ = Q_L1 left join Q_L2 on JC(L1, L2)
if there exists predicate pushed down only on L1

(2) FOJ = Q_L1 inner join Q_L2 on JC(L1, L2)
if there exists a predicate pushed down only on L1 and a predicate pushed down only on L2.

For Query Q1:

```
FOJ_Q1 =    select *
            from Q1_resort   left  join Q1_hotel
                    on Q1_hotel .name = Q1_resort.name
```

Query unfolding: Resolution Function

RES_FOJ: Application of the resolution funtions to FOJ

```
RES_FOJ_Q1 =   select
               COALESCE(Q1_hotel.name, Q1_resort.name) AS Name,
               SUM(Q1_hotel.rooms, Q1_resort.hoteltooms) AS Room,
               AVG(Q1_hotel.amount, Q1_resort.price) AS Price
               from FOJ_Q1
```

Query Resul: Application of the residual conditions to RES_FOJ

```
Query Result  =  select
                        Name,
                        Room
                 from RES_FOJ_Q1
                 where Price = 100
```

Data Provenance for Data Integration

- **Data Provenance** or **Lineage** describes where data came from, how it was derived and how it was updated over time
- Provenance provides valuable information that can be exploited for many purposes: managing data uncertainty, identifying and correcting data errors.
- Provenance is one of the open problems and desiderata for data fusion systems.
  - to know the origin of the visualized data
  - to explain merging decisions by tracking which original values were involved and how they have been fused.

➢ Design and Development of a Provenance Management component for the MOMIS System

Data Provenance for the MOMIS system

- **Question**: Which data provenance model for the MOMIS system?
- Classical *Data Provenance models*
  - *Lineage* encodes the tuples that were used in some derivation of the query result (set of tuples)
  - *Why-provenance* encodes all the different derivations of a tuple in the query result (set of sets of tuples)
- But they are limited to UCQs (Union of Conjunctive Queries): The full join merge of MOMIS is more expressive than UCQs
  - **full join** + **merge** (using the resolution functions).
- **PI-provenance** (Perm Influence)
  - A why-provenance model including the full join.
  - An open source system (PERM - Provenance Extension of the Relational Model) supports provenance for SQL queries.

PI-provenance

- **PI-Provenance as a set of witness lists**:

  a **witness list** contains a local tuple from each local class or the special value ⊥, indicating that no tuple from a local class was used to derive the output tuple.

i-esima tupla of L1

L1

	ID	A
$L1^1$	1	3
	2	3
	3	12
	4	12

L2

	ID	B
$L2^1$	1	4
	2	
	3	3

```
SELECT   PROVENANCE DISTINCT A
FROM     L1 FULL JOIN L2
         USING (ID)
WHERE    L1.A > 3 OR L2.B  > 3
```

A	PI-Provenance
3	{<$L1^1$, $L2^1$>}
12	{<$L1^3$, $L2^3$>, <$L1^4$, ⊥>}

PI-provenance for the **full join merge** operator

- Problem: To encode all different derivations of a tuple in the query result in presence of **full join + resolution functions**
- Solution: The PI-provenance was extended to resolution functions in order to obtain all possible derivations

L1

ID	A
1	3
2	3
3	12
4	12

L2

ID	A
1	4
2	
3	3

```
SELECT  DISTINCT COALESCE(L1.A,L2.A)
FROM    L1 FULL JOIN L2
        USING (ID)
WHERE L1.A > 3 OR L2.A  > 3
```

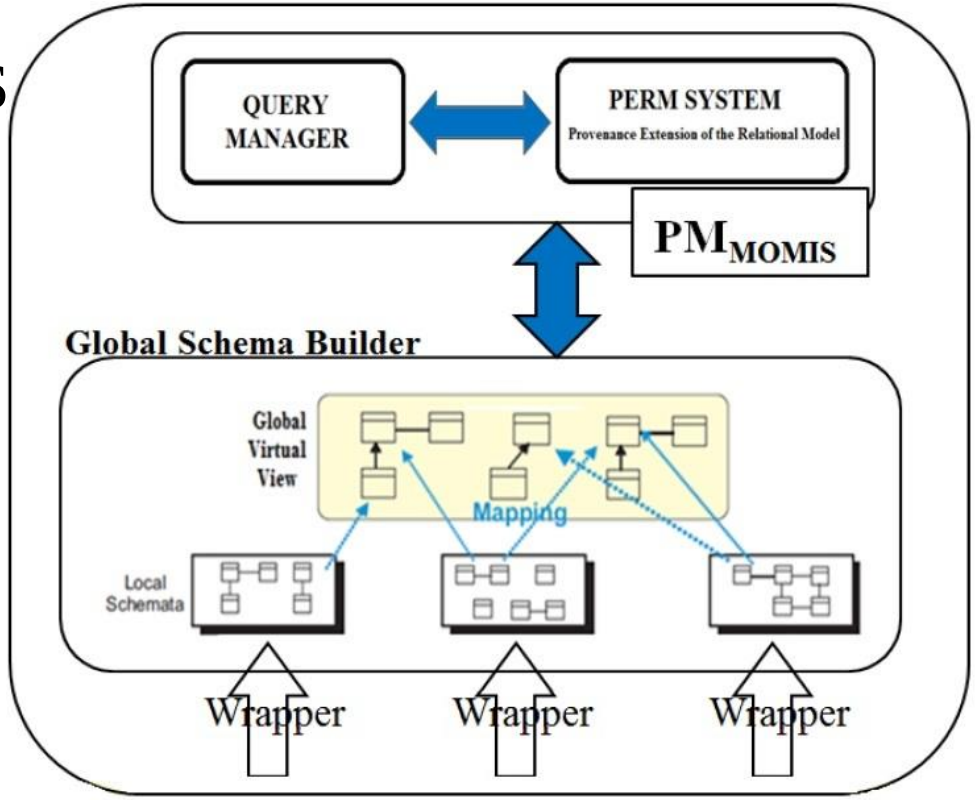
A	PI- Provenance
3	{<$L1^1$, $L2^1$>}
12	{<$L1^3$, $L2^3$>, <$L1^4$, ⊥>}

+

all possible derivations
{<$L1^3$, ⊥>}

Data Provenance in the MOMIS system: Architecture

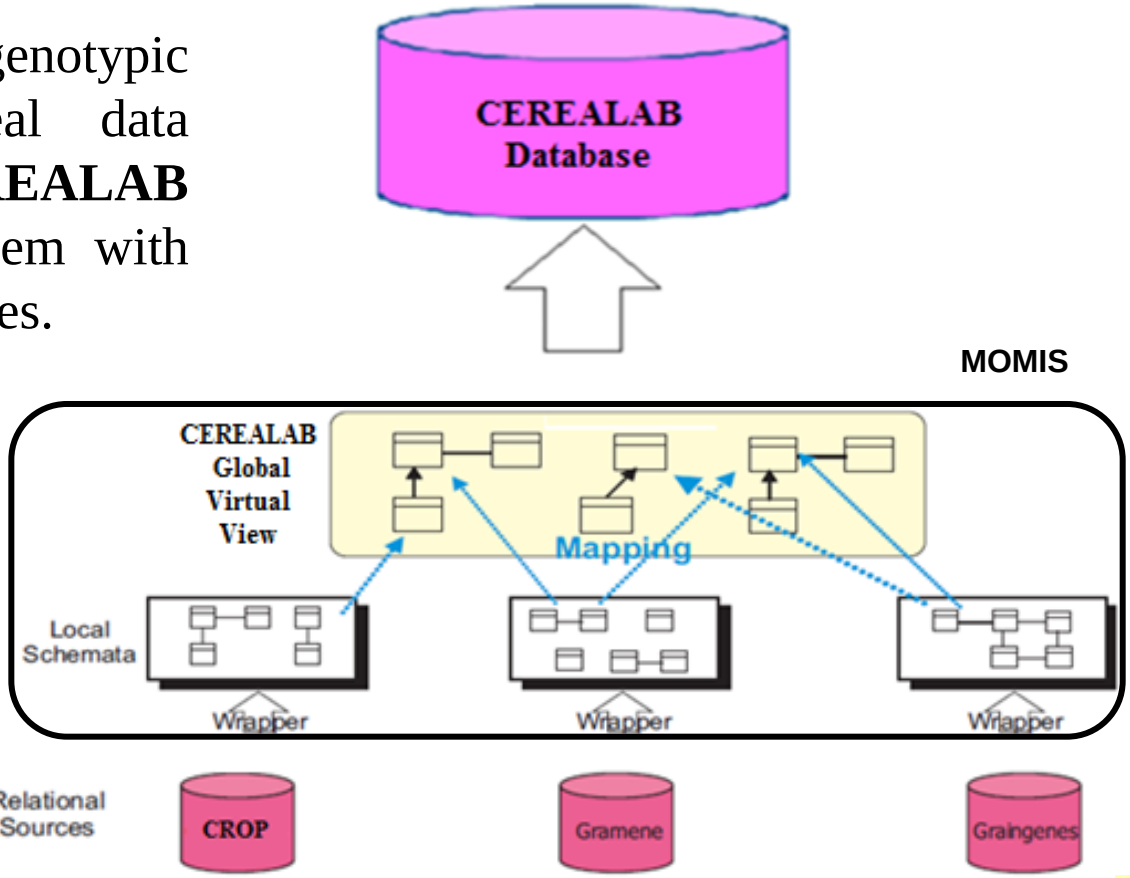
- The "*PI*-provenance" is fully implemented in the "Perm" system, an open-source provenance management system.
- The "Perm" system used as the SQL engine of MOMIS (Provenance computation for the **full join**)
- **Extensions** implementation for Resolution Functions (Provenance computation for the **full join-merge**)



Example : Integration of biological data

- Use of the MOMIS for an integrated and unified access to many available sources of genomic and phenotypic data

CEREALAB DB stores genotypic and phenotypic cereal data collected within the **CEREALAB project** and integrates them with already existing data sources.



Example : Integration of the local schemata

- Two local classes (**GermplasmA** and **GermplasmB**) with the same attributes

GPN : GermPlasm Name.
FHB : Fusarium Head Blight.
Yield: Production in t/ha.
Type: Germplasm type.

- Global class **GERMPLAS** (GPN,YIELD,FHB,TYPE) where
  - GPN is the shared identifier
  - YIELD,FHB and TYPE are conflicting attributes

Example : Data Fusion

GPA (GermplasmA)

GPN	yield	FHB	Type
Eureka	18	MR	
Fortuna	7	MR	
Mentana		S	Line
Kenora	20	MR	Landrace
Oasis	21	MR	Cultivar

GPB (GermplasmB)

GPN	yield	FHB	Type
Eureka	6	S	Cultivar
Fortuna	15	S	Landrace
Mentana	20	MR	Line
Kenora			Cultivar

```
SELECT  GPN,
        Yield =AVG(GPA.Yield,GPB.Yield),
        FHB = COALESCE(GPA.FHB, GPB.FHB),
        Type = ALLVALUES(GPA.type,GPB.type)
FROM    GPA  FULL JOIN GPB
        USING (GPN)
```

GPN	yield	FHB	Type
Eureka	12	MR	Cultivar
Fortuna	11	MR	Landrace
Mentana	20	S	Line
Kenora	20	MR	Landrace, Cultivar
Oasis	21	MR	Cultivar

Example : Provenance

Query: types of varieties that are resistant to FHB?

GERMPLASM

GPN	yield	FHB	Type
Eureka	12	MR	Cultivar
Fortuna	11	MR	Landrace
Mentana	20	S	Line
Kenora	20	MR	Landrace, Cultivar
Oasis	21	MR	Cultivar

```
TYPE_MR =
    SELECT  DISTINCT Type
    FROM    GERMPLASM
    WHERE   FHB='MR'
```

TYPE_MR

Type
Landrance
Cultivar
Landrace,Cultivar

*P*rovenance for the *TYPE_MR* query

Type	Provenance as a set of witness lists
Landrance	{<GPA ^{Fortuna}, GPB ^{Fortuna}>}
Cultivar	{<GPA ^{Eureka}, GPB ^{Eureka}>, <GPA ^{Oasis}, $\perp$>}
Landrace,Cultivar	{<GPA ^{Kenora}, GPB ^{Kenora}>}

Example : Provenance

- In the MOMIS+PERM system witness lists are represented in a relational form:

Each witness list of an output tuple is represented by a single tuple

Type	GPA.GPN	GPA.yield	GPA.FHB	GPA.type	GPB.GPN	GPB.yield	GPB.FHB	GPB.type
landrace	Fortuna	7	MR		Fortuna	15	S	landrace
cultivar	Eureka	18	MR		Eureka	6	S	cultivar
cultivar	Oasis	21	MR	cultivar				
landrace,cultivar	Kenora	20	MR	landrace	Kenora			cultivar

Selected Publications

S. Bergamaschi, S. Castano e M. Vincini: **Semantic Integration of Semistructured and Structured Data Sources**, SIGMOD Record Special Issue on Semantic Interoperability in Global Information, Vol. 28, No. 1, March 1999.

D. Beneventano, S. Bergamaschi, S. Castano, A. Corni, R. Guidetti, G. Malvezzi, M. Melchiori e M. Vincini: **Information Integration: the MOMIS Project Demonstration**, International Conference on Very Large Data Bases (VLDB'2000), Cairo, Egypt, Settembre 2000.

S. Bergamaschi, S. Castano, D. Beneventano e M. Vincini: **Semantic Integration of Heterogeneous Information Sources**, Special Issue on Intelligent Information Integration, Data & Knowledge Engineering, Vol. 36, Num. 1, Pages 215-249, Elsevier Science B.V. 2001.

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini: **Synthesizing an Integrated Ontology**, IEEE Internet Computing, Vol.7,N.5, September/October 2003.

S. Bergamaschi, G. Cabri, F. Guerra, L. Leonardi, M. Vincini, F. Zambonelli, **Exploiting Agents to Support Information Integration,** Special Issue of the International Journal on Cooperative Information Systems vol. 11(3-4): 293-314, 2002, ISSN 0218-8430

I. Benetti, D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini, **An Information Integration Framework for E-Commerce,** IEEE Intelligent Systems Magazine, Jan/Feb 2002, pp. 18-25,

I. Benetti, S. Bergamaschi, F. Guerra, M. Vincini, **Soap-enabled web services for knowledge management** to appear in Int. J. Web Engineering and Technology,Vol. 1. N.2., 2004.

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini: **Building a Tourism Information Provider with the MOMIS System**, Information Technology & Tourism Journal(ISSN 1098-3058), 7:3_4, 2005.

Selected Publications

S. Bergamaschi, F. Guerra, **Peer to Peer Paradigm for a Semantic Search Engine**, in proceedings of the International Workshop on Agents and Peer-to-Peer Computing, Bologna, 15 July 2002, LNCS 2530, Springer ISBN 3-540-40538-0

S. Bergamaschi, F. Guerra, M. Vincini, **Product Classification Integration for E-Commerce**, Second International Workshop on Electronic Business Hubs - WEBH 2002 in conjunction with DEXA 2002, September 2-6 2002, Aix En Provence, France, published by IEEE Computer Society, Los Alanitos (CA), ISBN 0-7695-1668-8, pp. 861-867

D. Beneventano, S. Bergamaschi, S. Castano, V. De Antonellis, A. Ferrara, F. Guerra, F. Mandreoli, G. Ornetti, M. Vincini, **Semantic Integration and Query Optimization of Heterogeneous Data Sources,** 1st Int.l Workshop on Efficient Web-based Information Systems (EWIS), 2002, Montpellier, France, pp.154-165.

S. Bergamaschi, F. Guerra, M. Vincini, **A peer-to-peer information system for the semantic web,** in proceedings of the International Workshop on Agents and Peer-to-Peer Computing, in AAMAS 2003 Melbourne, Australia, July 14, 2003

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini: **Building an Ontology with MOMIS**, in proceedings of the Semantic Integration Workshop within the Second International Semantic Web Conference, October 20, 2003 Sundial Resort, Sanibel Island, Florida, USA.

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini, **Building an integrated Ontology within SEWASIE system,** in proceedings of the First International Workshop on Semantic Web and Databases, Co-located with VLDB 2003 Berlin, Germany, (2003)

S. Bergamaschi, G.Gelati, F. Guerra, M. Vincini, **WINK: a Web-based Enterprise System for Collaborative Project Management in Virtual Enterprises,** 4th International Conference on Web Information Systems Engineering, Roma Italy, 10-12 December 2003

Selected Publications

S. Bergamaschi, F. Guerra, M. Orsini, C. Sartori, **Extracting Relevant Attribute Values for Improved Search**, IEEE Internet Computing, vol. 11, no. 5, pp. 26-35, Sept/Oct, 2007 (special issue on Semantic-Web-Based Knowledge Management)

S. Bergamaschi, F. Guerra, M. Orsini, C. Sartori, **Relevant values: new metadata to provide insight on attribute values at schema level**, International Conference on Enterprise Information Systems (ICEIS 2007), 12-16, June 2007, Funchal, Madeira – Portugal

D. Beneventano, S. Bergamaschi, M. Vincini, M. Orsini, R. Carlos Nana, **Query Translation on Heterogeneous Sources in MOMIS Data Transformation Systems**, VLDB 2007 - Third International Workshop on Database Interoperability (InterDB 2007), September 24, 2007 – Vienna, Austria.

D. Beneventano, S.Bergamaschi: **Semantic search engines based on data integration systems**. In Semantic Web Services: Theory, Tools and Applications. IGI Global, Information Science Reference, Hershey, New York, 2006, pages 317-341

S.Bergamaschi, P. Bouquet, D. Giacomuzzi, F. Guerra, L. Po, and M. Vincini: **MELIS: an incremental method for the lexical annotation of domain ontologies**, in International Journal on Semantic Web and Information Systems (IJSWIS) 3(3), p.p.57-80 2007

S.Bergamaschi, A.Sala: **Virtual Integration of existing web databases for the genotypic selection of cereal cultivars**, The 5th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2006), Montpellier, France, Oct 31 - Nov 2, 2006

Selected Publications

S. Bergamaschi, A. Sala, **Creating and Querying an Integrated Ontology for Molecular and Phenotypic Cereals Data**, Special Session on Agricultural Metadata & Semantics of the 2nd International Conference on Metadata and Semantics Research (MTSR'07), Corfù, Greece, October 11-12 ,2007

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini:**The SEWASIE MAS for semantic search**, In proceedings of the 1st International workshop on Agent supported Cooperative Work 2007, Lyon, France, 29 October 2007

D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini: **Querying a super-peer in a schema-based super-peer network**, In proceedings of the 3rd VLDB International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P 2005) held at VLDB 2005 31st International Conference on Very Large Data Bases, Trondheim, Norway, August 28-29, 2005

Sonia Bergamaschi, Laura Po, Serena Sorrentino: **Automatic annotation for mapping discovery in data integration systems** 16th Italian Symposium on Advanced Database Systems (SEBD 2008) Mondello (PA), Italy, June 22-25, 2008

I. F. Cruz, R. Gjomemo, M. Orsini: **A Secure Mediator for Integrating Multiple Level Access Control Policies**, 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2008), Zagreb, Croatia, September 3-5, 2008