DB Group @ unimo

3rd International Workshop on Data Integration in Life Sciences DILS 06,
European Bioinformatics Institute (EBI), Hinxton, UK

20 - 22 July 2006

# Virtual Integration of Existing Web Databases for the Genotypic Selection of Cereal Cultivars

Sonia Bergamaschi - Antonio Sala

www.dbgroup.unimo.it

Dipartimento di Ingegneria dell'Informazione

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

DB Group @ unimo

- To perform intelligent data integration of existing databases to create a Global Virtual View (GVV) for the genotypic selection of cereal cultivars.

- The GVV has been realized with the **MOMIS** system (Mediator envirOnment for Multiple Information Sources) developed by the Database Group of the University of Modena and Reggio Emilia as a part of the **CEREALAB** project conducted by the Agrarian faculty of the University of Modena and Reggio Emilia in collaboration and funded by the Regional Government of Emilia Romagna.

DB Group @ unimo
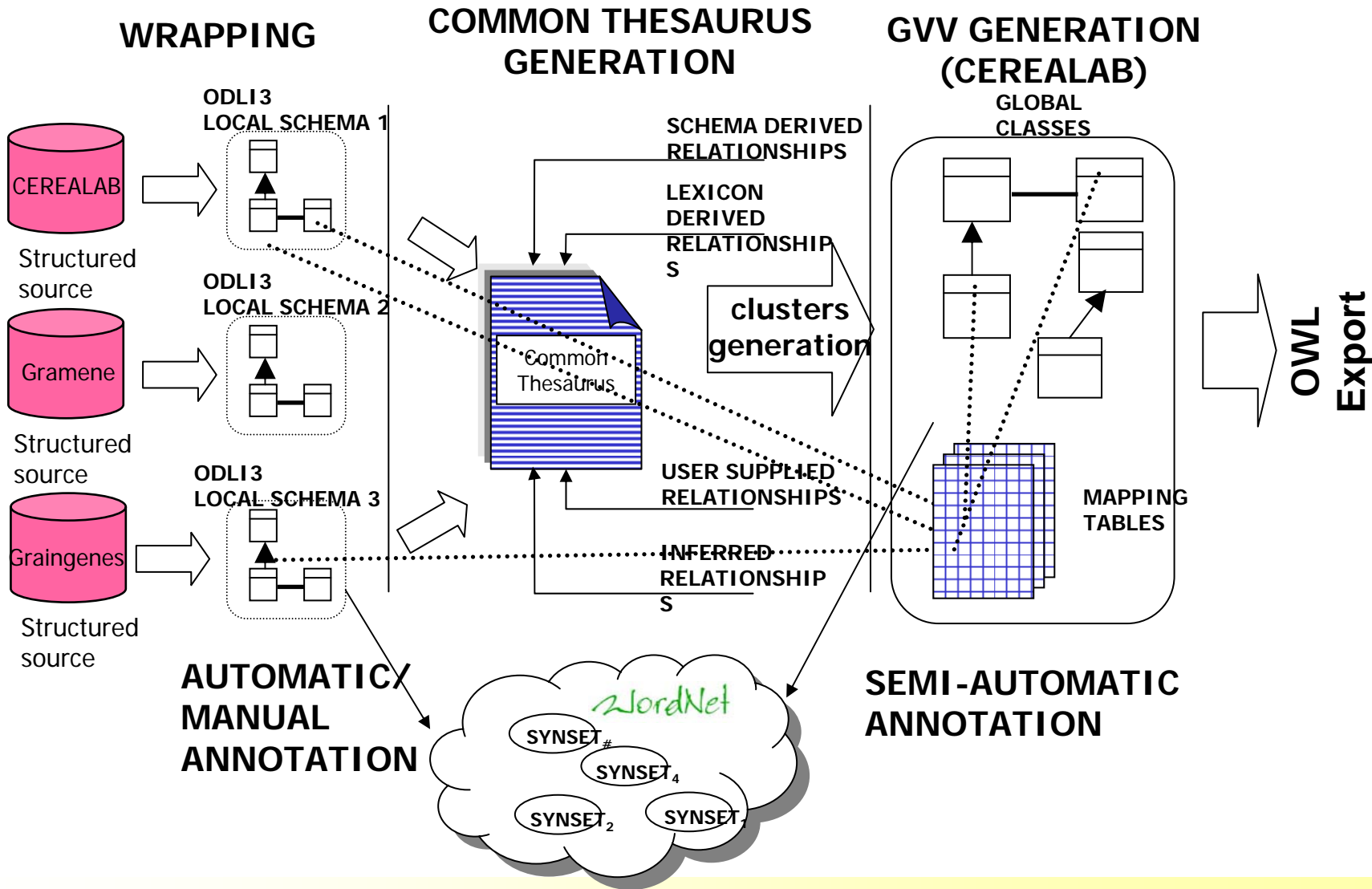
MOMIS uses an object-oriented language called ODL$_{i3}$ as a common data model for integrating a given set of local information sources.

ODL$_{i3}$ extends ODL with the following relationships expressing intra- and inter-schema knowledge for the source schemata:

- SYN (synonym of)
- BT (broader terms)
- NT (narrower terms)
- RT (related terms)

By means of ODL$_{i3}$, only one language is exploited to describe both the sources (the input of the synthesis process) and the GVV (the result of the process).

ODL$_{i3}$ is based on the OCDL description logics. Translators ODL$_{i3}$/OCDL and OCDL/ODL$_{i3}$ are available.

DB Group @ unimo

**MOMIS**

- Identifies and groups similar $ODL_{i3}$ classes (classes that describe the same or semantically related concept in different sources) into clusters (global classes)
- Generates mappings among global and local classes in the cluster

Cluster generation: affinity coefficients are evaluated for all possible pairs of $ODL_{i3}$ classes, based on the relationships in the Common Thesaurus properly strengthened

- Affinity coefficients determine the degree of matching of two classes based on:
  - their names (Name Affinity coefficient)
  - their attributes (Structural Affinity coefficient)
- Affinity coefficients are fused into Global Affinity coefficients calculated by means of the linear combination of the two coefficients
- Global affinity coefficients are used by a hierarchical clustering algorithm, to include $ODL_{i3}$ classes in clusters according to their degree of affinity

- The designer may interactively refine and complete the proposed integration results

DB Group @ unimo

- A Mapping Table (MT) is automatically generated for each global class of a GVV.

- The designer can extend the MT by adding:
  - Data Conversion Functions from local to global attributes
  - Join Conditions among pairs of local classes.
  - Resolution Functions for global attributes to solve data conflicts of local attribute values.
    - MOMIS provides some standard kinds of resolution functions for solving data conflicts for each global attribute mapping onto local attributes coming from more than one local source:
    - Random
    - Aggregation
    - Coalescence
    - Precedence function
    - All Values

DB Group @ unimo

The **MOMIS** Query Manager is the coordinated set of functions which allows the user to query the GVV

Query processing consists of the following steps:

- Query rewriting
    - to rewrite a global query as an equivalent set of queries expressed on the local sources (local queries)
- Local queries execution
    - the local queries are sent and executed at local sources
- Fusion and Reconciliation
    - The local answers are fused into the global answer