ESF
EIROPAS SOCIĀLAIS FONDS
EIROPAS SAVIENĪBA
IEGULDĪJUMS TAVĀ NĀKOTNĒ

This work has been supported by ESF projects No. 2009/0216/1DP/1.1.1.2.0/09/APIA/VIAA/044

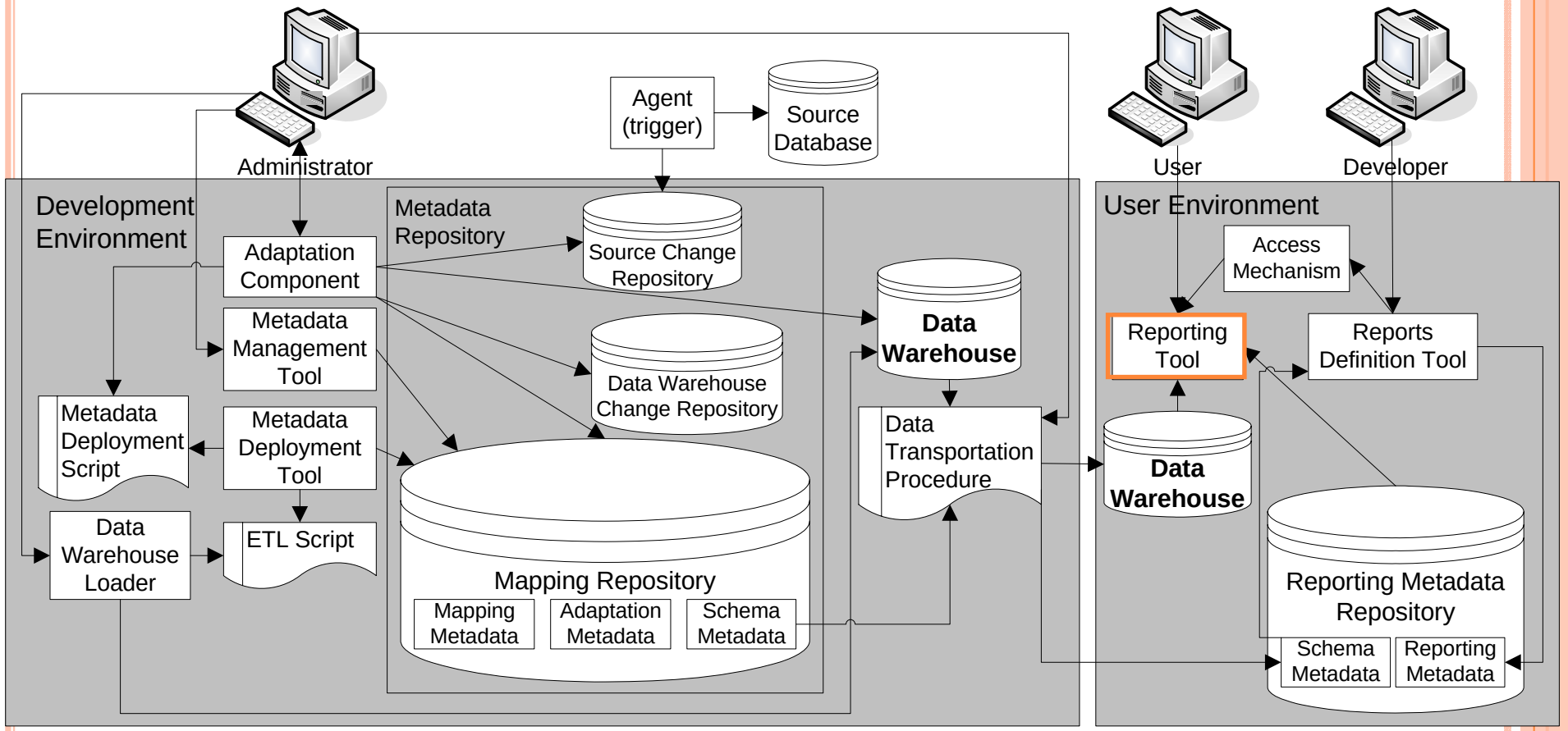
# Determining Preferences from Semantic Metadata in OLAP Reporting Tool

*Darja Solodovnikova & Natalija Kozmina*
*(University of Latvia)*

# OUTLINE

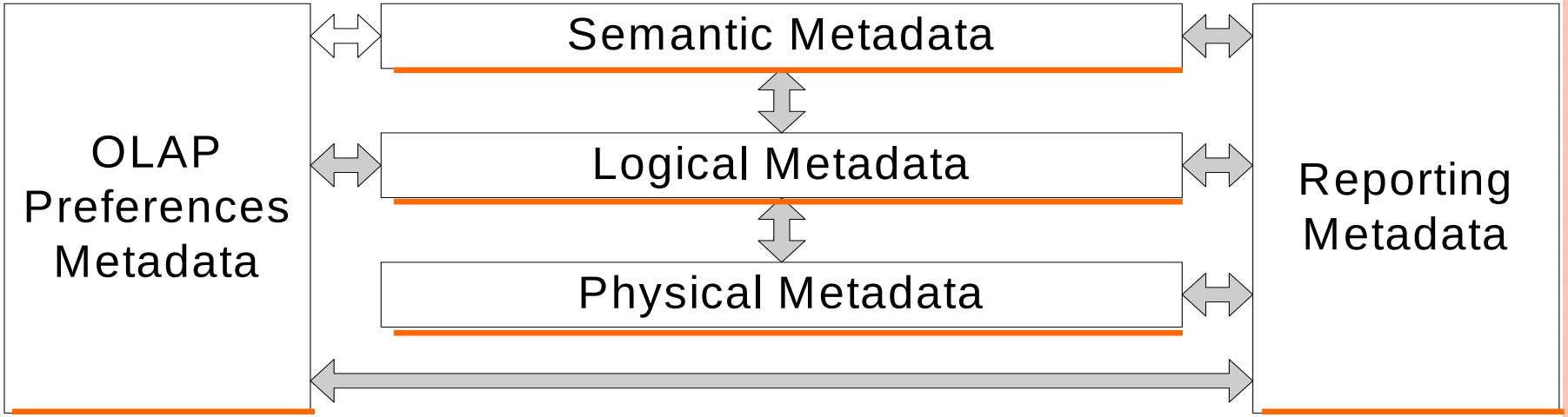
- Evolution framework architecture
  - OLAP reporting tool as a part of the evolution framework
  - Evolution support
- Five metadata layers
  - Logical, physical, semantic, reporting & OLAP preferences metadata
- Determining preferences from semantic description
  - Processing user preferences described with semantic metadata
  - Examples
- Conclusions and future work

# EVOLUTION FRAMEWORK ARCHITECTURE



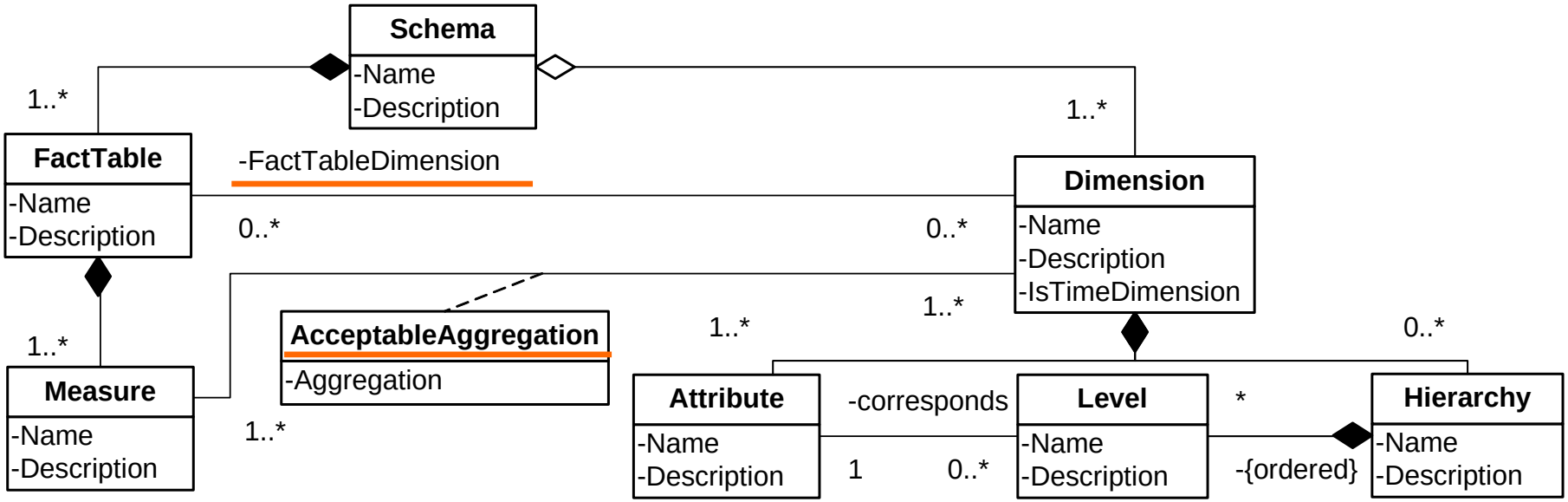

- Evolution support:
  - *Physical changes* operate with database objects
    - Examples: creation or deletion of an attribute, measure, dimension.
  - *Logical changes* modify schema metadata
    - Examples: creation or deletion of a hierarchy or level.
  - *Semantic changes* adapt meanings of data objects

# FIVE METADATA LAYERS



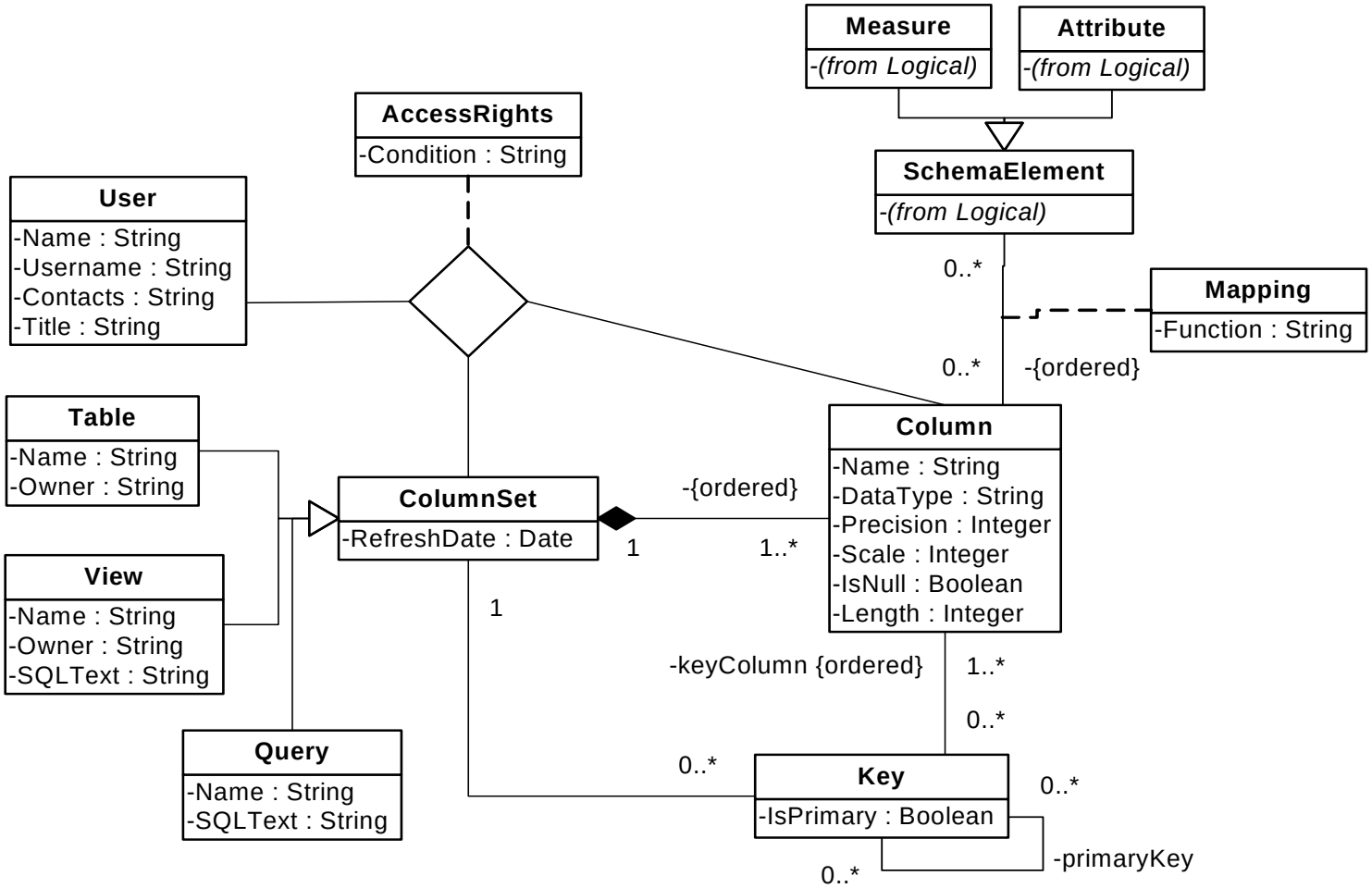

- *Logical*: data warehouse schemata
- *Physical*: storage of a data warehouse in relational database
- *Semantic*: data stored in a data warehouse and data warehouse elements in a way that is understandable to users
- *Reporting*: definitions of reports on data warehouse schemata
- *OLAP preferences*: definitions of user preferences on reports' structure and data

# LOGICAL METADATA



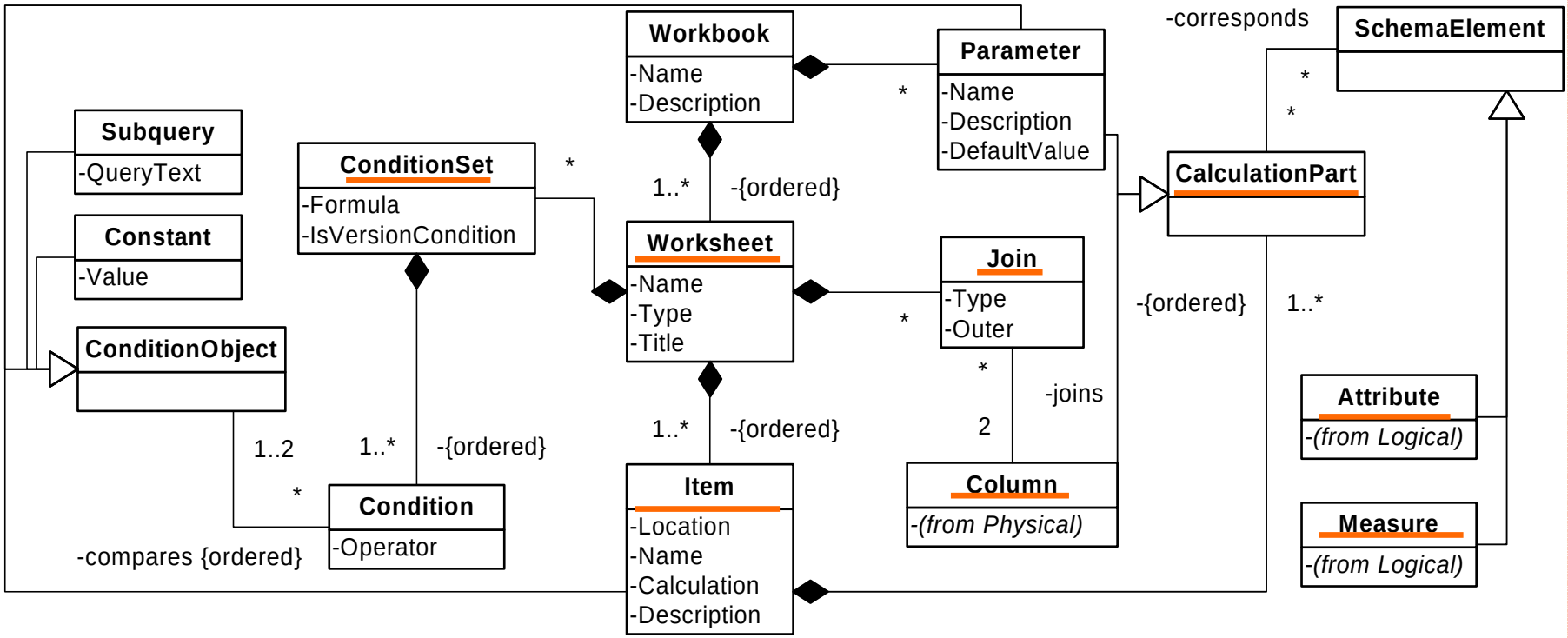

- Dimensions, attributes, hierarchies, fact tables and measures are adopted from the OLAP package of Common Warehouse Metamodel (*CWM*).
  - Fact tables & dimensions are connected by *FactTableDimension* association
- OLAP package of CWM was extended by the class *AcceptableAggregation.*
  - SUM, AVG, COUNT, MIN, MAX

# PHYSICAL METADATA




- Common Warehouse Metamodel (*CWM*) contains a package Relational, which was taken as a basis for physical metadata.
- Relational database schema of a data warehouse and mapping of a multidimensional schema to relational database objects is described.

# REPORTING METADATA



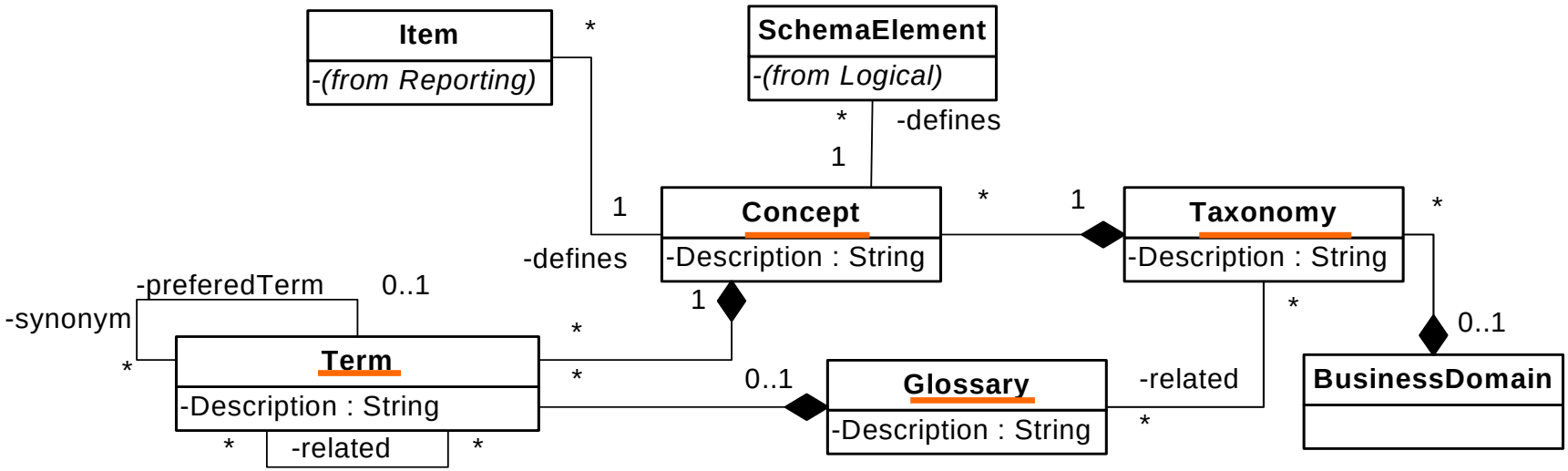

- Common Warehouse Metamodel (*CWM*) contains a package Information Visualization, but this metadata isn't sufficient, therefore, a new metamodel was developed.
- Reports in the tool can be defined by developers or experienced users themselves by choosing desired elements of a data warehouse schema and defining conditions, parameters, etc.
- When a user runs a report in the OLAP reporting tool, an SQL query is built based on the report definition in reporting metadata, and its result is displayed to a user.

# SEMANTIC METADATA (1/2)

- ***Why it is necessary to describe each element of the data warehouse model in business language?***

- ...To analyze this data using all necessary features, including OLAP operations “drill-down” and “roll-up” and using hierarchies.
- ...To modify or construct reports by users from elements, which are familiar to them, so that reports’ creation becomes transparent.
- ...To state OLAP preferences, operating with business language terms.
  - ...To provide users of different skill levels (e.g., expert, novice) with recommendations on potentially interesting reports.
- ...To express user requirements for information and also changes in requirements, thus, making the understanding between users and developers of data warehouse clearer.

# SEMANTIC METADATA (2/2)




- Common Warehouse Metamodel (*CWM*) Business Nomenclature package represents business metadata, and it was taken as a basis for semantic metadata.
- A *Concept* is the semantic meaning or notion of some data warehouse element or data stored in some element.
  - *Concepts* are united in *Taxonomies*.
- A *Term* is particular word or phrase employed by users to refer to a concept.
  - *Terms* are united in *Glossaries*.

# OLAP PREFERENCES METADATA (1/3)

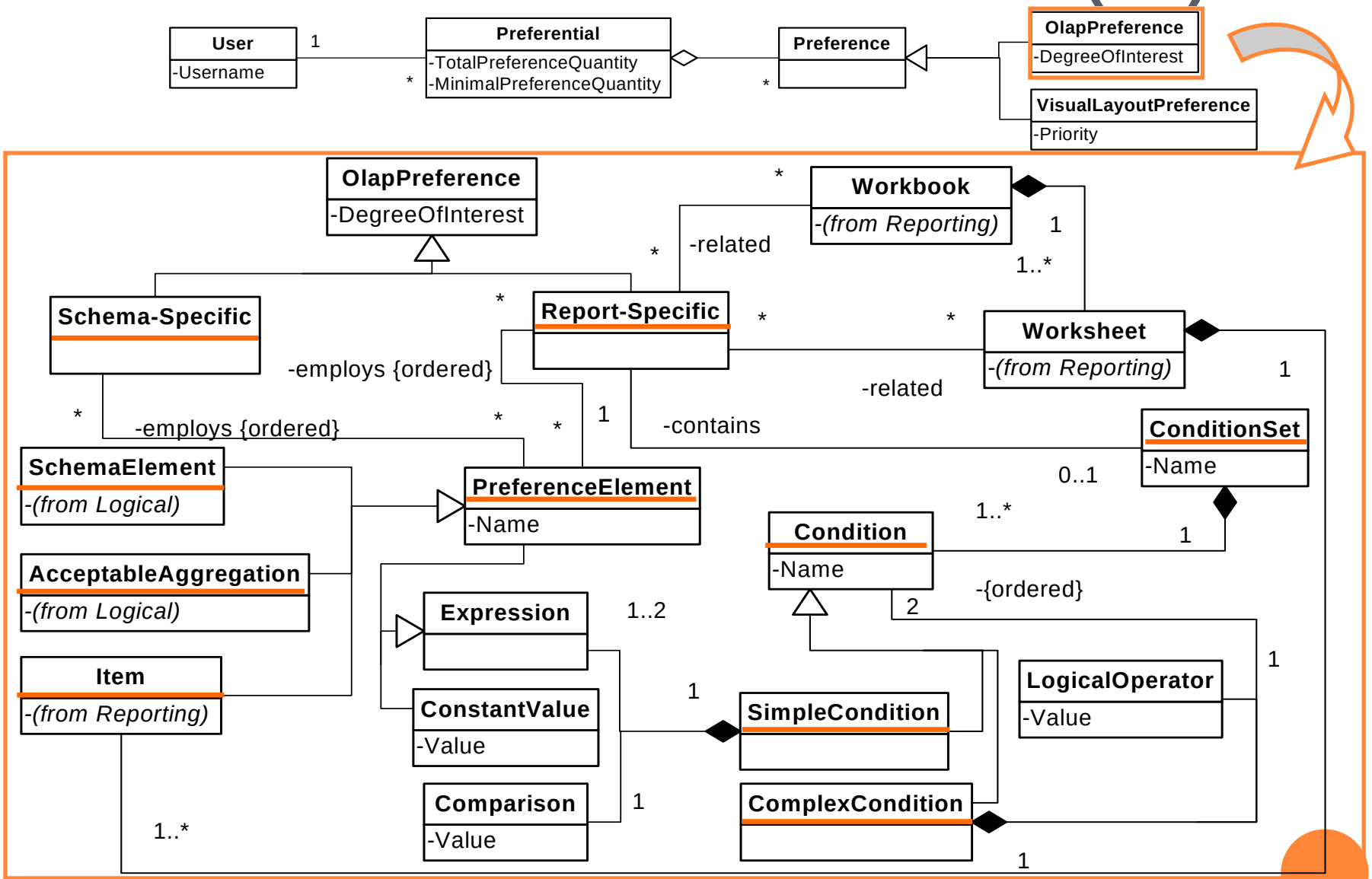
- ***Why is OLAP personalization important?***

- Typical problems in the field of data warehousing:
  - Large volumes of data
  - Burdening data exploration
  - While exploring previously unknown data, the OLAP query result may highly differ from expectations (i.e., empty result sets or data floods)
- Possible solution – introducing *personalization* in the field of data warehousing.
- *Personalization* – is a process of providing users with selected information on their specific needs*.

* - BNET Business Dictionary:

http://dictionary.bnet.com/definition/personalization.html

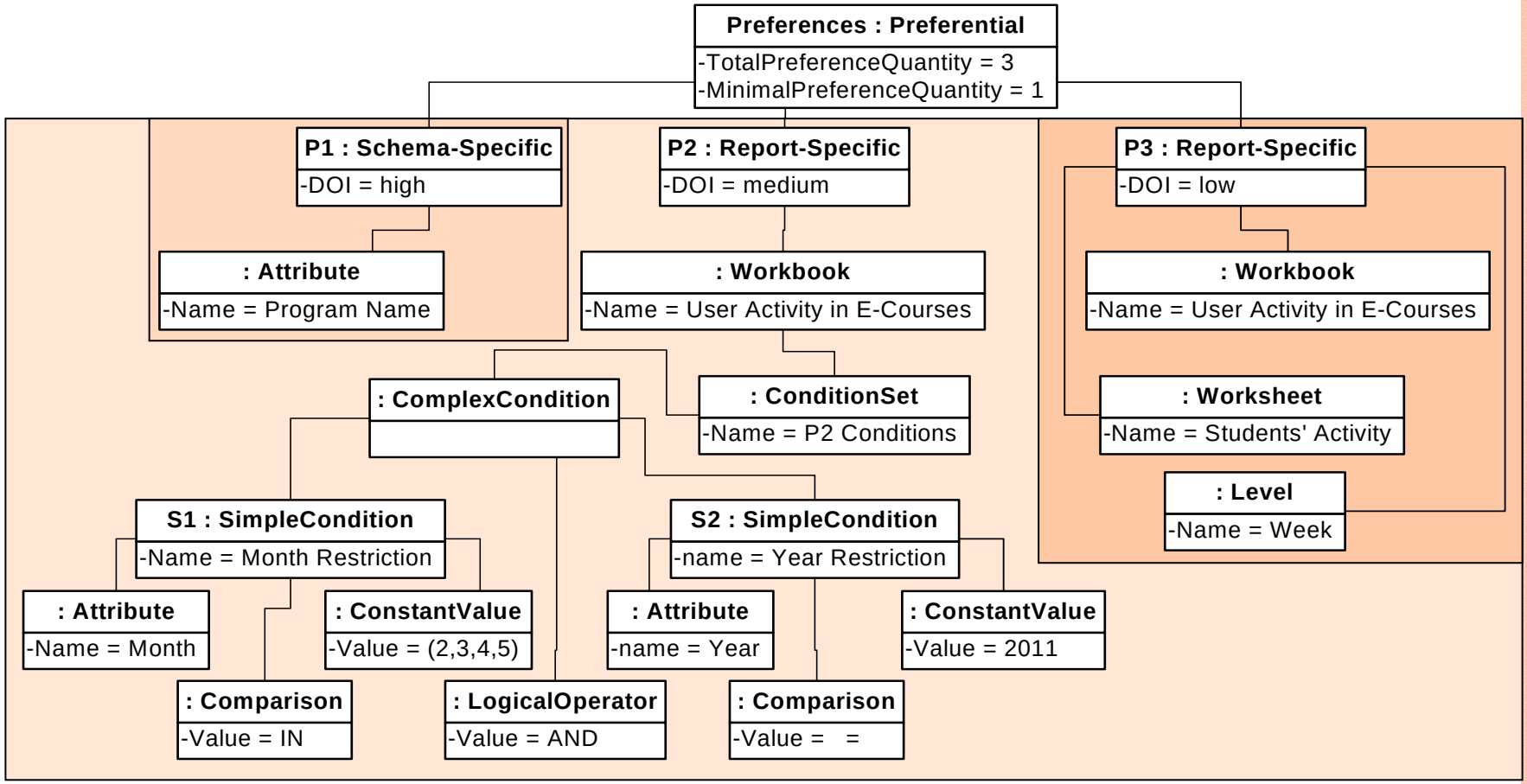
# OLAP PREFERENCES METADATA (2/3)




- *Schema-Specific preferences* are set on OLAP schema, its elements and acceptable aggregate functions.
- *Report-Specific preferences* include restrictions on data in reports.

# OLAP PREFERENCES METADATA (3/3)

- *Instances of Schema-Specific preferences*
  - P1: I have a high interest in any report that includes an attribute “ProgramName”.
- *Instances of Report-Specific preferences*
  - P2: I have a medium interest in reports from the workbook “User Activity in E-Courses” from Feb-2011 to May-2011.
  - P3: I have a low interest in a report “Students’ Activity” from the workbook “User Activity in E-Courses” by weeks.



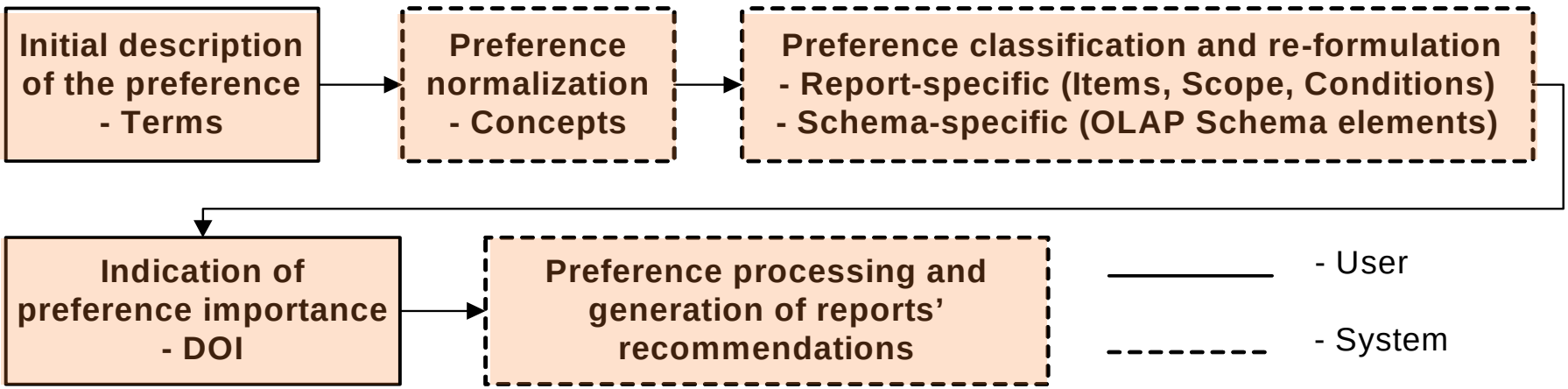
# DETERMINING PREFERENCES FROM SEMANTIC DESCRIPTION

- Processing user preferences described with semantic metadata
- Examples

**Terms:**
"study program",
"acad. specialization",
"branch", "field of study"

**Concept:**
"study program"

**Report-specific:**
-Scope: worksheet
-Condition: "study program name = "Information Systems" "



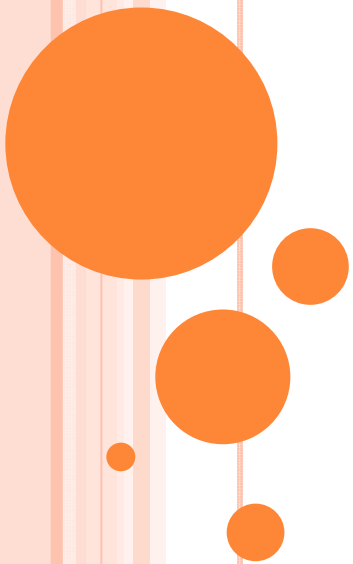

**DOI:**
medium = 0.5.

New methods to be developed for
- Explicitly defined report-specific preferences
- Implicitly discovered schema-specific preferences → Explicitly defined schema-specific preferences

# CONCLUSIONS AND FUTURE WORK

- Five different layers of metadata (logical, physical, semantic, reporting & OLAP preferences metadata) that intersect each other were exposed.
- A possibility for a user to create OLAP preferences, using description in business language, operating with synonym terms and choosing the most appropriate among them was considered.
- A concept of the algorithm of OLAP preference creation, transformation and processing was set forth.
- TODO:
  - To extend and supplement the algorithm of OLAP preference creation, transformation and processing;
  - To review the existing approach* for generation of reports' recommendations, which is based on implicitly discovered schema-specific user preferences;
  - To adapt the approach* to explicitly set schema- and report-specific user preferences and to evaluate it;
  - To handle semantic changes, i.e. different versions of meanings of the same data objects.

- * Described in the paper: "On Implicitly Discovered OLAP Schema-Specific Preferences in Reporting Tool" (N. Kozmina, D. Solodovnikova), to be presented at BIR2011.



Thank

You