

# Incorporating Traceability in Conceptual Models for Data Warehouses by using MDA

*Alejandro Maté*  
amate@dlsi.ua.es

Juan Trujillo  
jtrujillo@dlsi.ua.es

*Lucentia*

 **Universitat d'Alacant**  
**Universidad de Alicante**

30th International Conference on Conceptual Modeling  
ER'11 2011. October 31th – November 3rd, Brussels, Belgium

## Content

- Introduction
- Related work
- Traceability of PIM to PIM Models
- Example of application
- Conclusions & future work

*Lucentia*

# Content

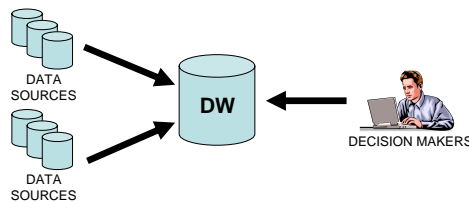
- Introduction
- Related work
- Traceability of PIM to PIM Models
- Example of application
- Conclusions & future work

*Lucentia*

3

# Introduction

- Data Warehouse
  - Integrates several heterogeneous data sources in support of management's decisions



*Lucentia*

4

# Introduction

- Building a Data Warehouse

*[Giorgini et al. 2008][Mazón et al. 2008]*

- Requires **matching** data source structures with both requirements and target DW structures
  - Defining the DW structures
  - Loading data from data sources
- The process produces a set of **reconciled elements** as a result
- What happens if **requirements change**?

*Lucentia*

5

# Introduction

## Requirements Model

**Teacher**  
oocm1a  
D Name  
DA YearsOfExperience

## Data sources Model

**TT TEA**  
oocod\_tea  
D NUMBER  
DA F\_NAME  
DA S\_NAME1  
DA S\_NAME2  
DA IS\_ACTIVE  
DA AOCCT  
DA BANK  
DA ADDR  
DA INCO  
DA UNIT\_COD  
DA COD\_CONT

Correct?

Result in the DW

**Teacher**  
oocod1a  
D Name  
DA IS\_ACTIVE  
DA INT\_COD

*Lucentia*

6

## Content

- Introduction
- **Related work**
- Traceability of PIM to PIM Models
- Example of application
- Conclusions & future work

*Lucentia*

7

## Related Work

### ▪ Related work

[Aizenbud-Reshef et al. 2006][Antoniol et al. 2002][Gotel et al. 2008][Ramesh et al. 2001]

- Traceability has been applied in other areas as a means of:
  - Identifying the scope of requirements
  - Impact change analysis
  - Rationale comprehension
- However, it has been **completely overlooked** in DWs
  - Recent rise in MDD has magnified the lack of traceability

*Lucentia*

8

## Content

- Introduction
- Related work
- **Traceability of PIM to PIM Models**
- Example of application
- Conclusions & future work

*Lucentia*

9

## Traceability of PIM to PIM Models

- Objectives:
  - Cutting costs and time required for the development process
  - Minimize the number of errors and impact of changes introduced
  - Enable support for automatic analysis
  - Easy adaptation of the proposal to other DW development approaches

*Lucentia*

10

## Traceability of PIM to PIM Models

- How?
  - **Preserving** the **existing relationships** between DW structures and data sources
    - Avoid repeating work when:
      - Requirements change
      - Each time a new data source is added
      - Designing ETL processes
  - Including **semantics** for automatic derivation & analysis by means of a **metamodel**
  - Relate elements without considering its type

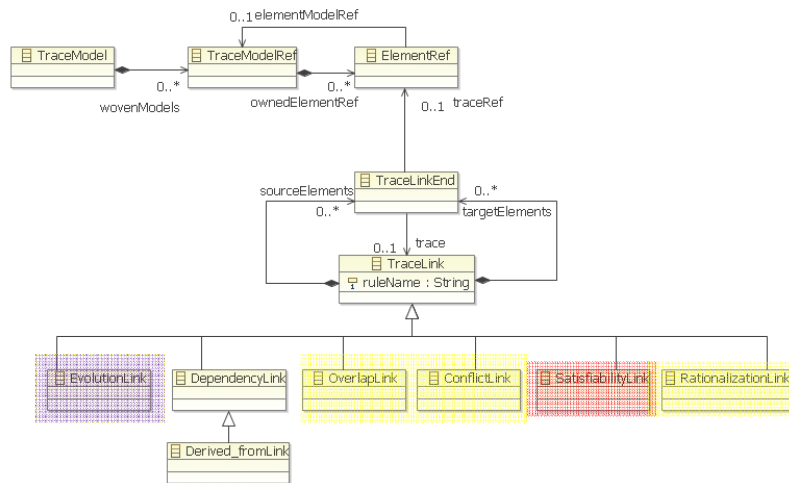
11

## Traceability of PIM to PIM Models

- Semantics
  - We can identify the following situations at conceptual level when developing a DW:
    - A conceptual element is refined into a new version of the same element → **Evolution**.
    - Two or more conceptual elements refer to the same concept but they differ in its representation → **Conflict**.
    - Two or more conceptual elements refer to the same concept AND their representation is compatible → **Overlap**.

12

## Traceability of PIM to PIM Models



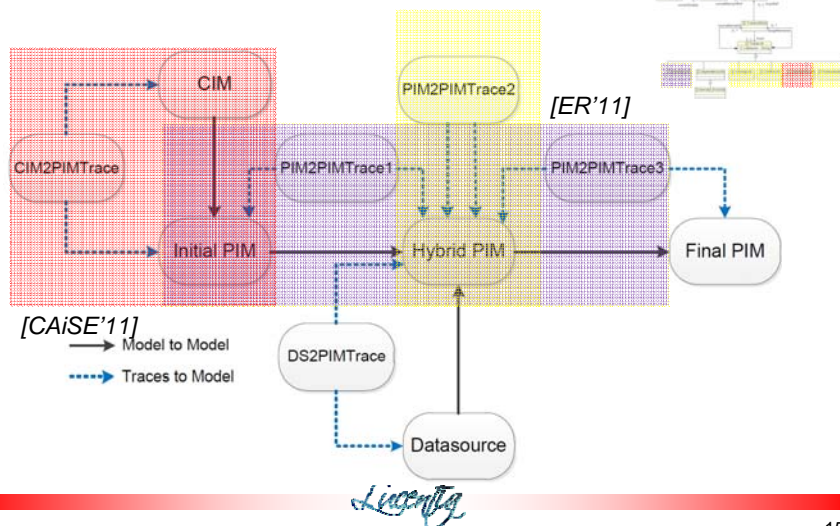
13

## Traceability of PIM to PIM Models

- The record of traceability links is performed in a semi-automatic way:
  - Elements are mapped from PIM to PIM by means of **QVT transformations** (*Evolution*)
  - The relationships between elements during the **reconciliation process** elicited by **the designer** (*Overlap, Conflict and Rationalization*)

14

# Traceability of PIM to PIM Models



15

# Content

- Introduction
- Related work
- Traceability of PIM to PIM Models
- **Example of application**
- Conclusions & future work

16



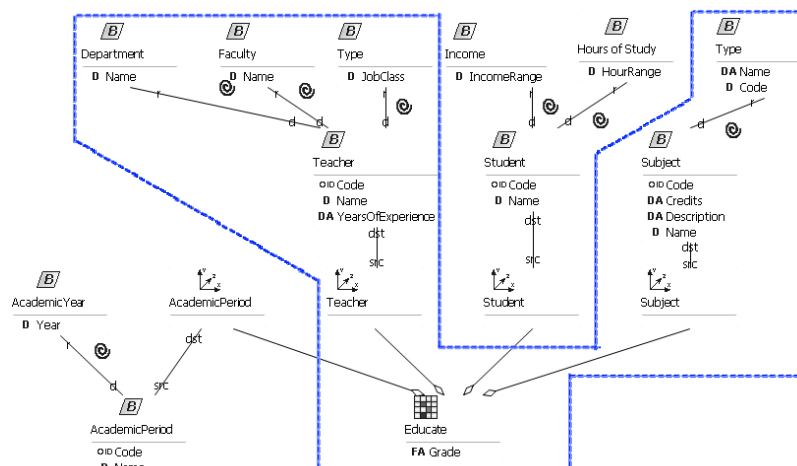
## Example of Application

- Let us show an example of application:
  - We have a requirements PIM, which satisfies the initial analysts' needs.
  - We also have a data source PIM, obtained by means of heuristics.
  - We perform the matching process by semantically relating elements.

*Lucy*

17

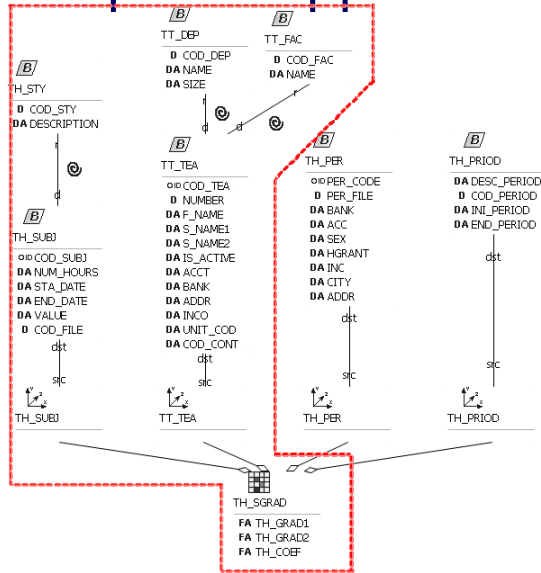
## Example of Application



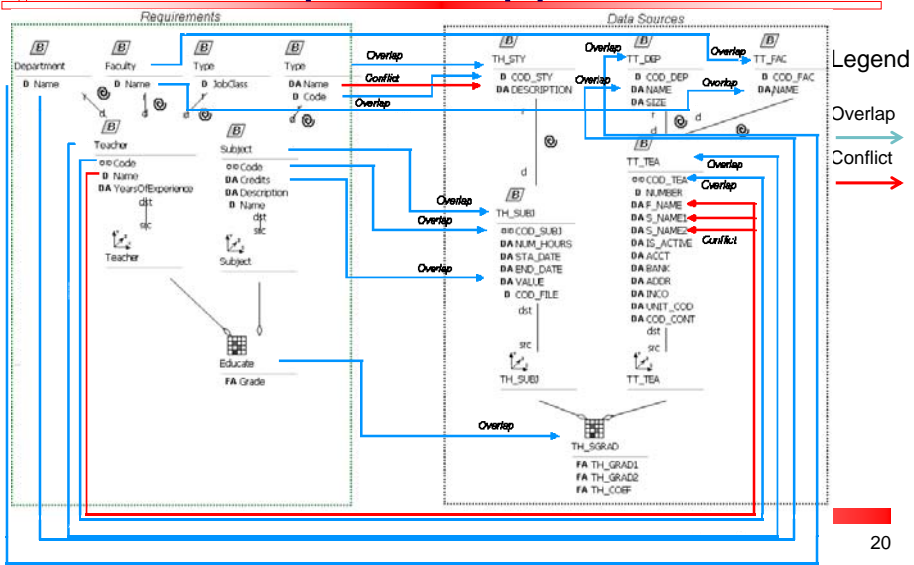
*Lucy*

18

# Example of Application



# Example of Application



## Example of Application

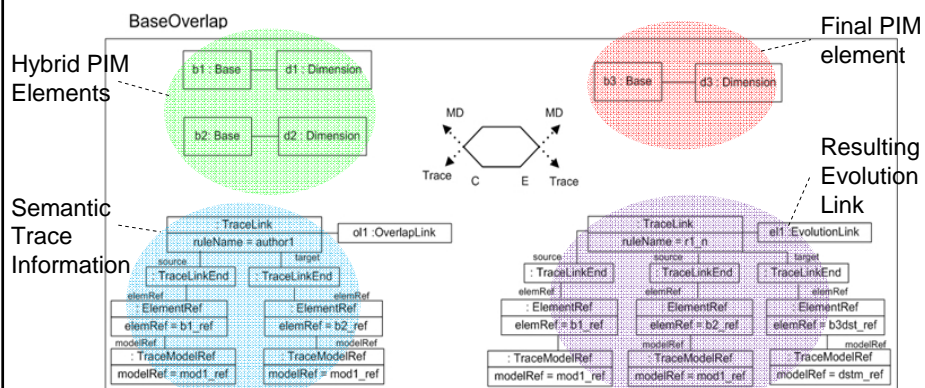
- Once the relationships have been elicited, trace models allow us to:
  - Perform impact change analysis
  - Cut ETL process design time
  - Derive different versions of the DW according to the semantics specified

*Lucentia*

21

## Example of Application

- Deriving the final DW using QVT:



*Lucentia*

22

## Content

- Introduction
- Related work
- Traceability of PIM to PIM Models
- Example of application
- **Conclusions & future work**

*Lucentia*

23

## Conclusions & Future Work

- Conclusions:
  - We have presented an approach to **preserve the traceability** of elements in DW development
  - Benefits:
    - **Minimizes effort** and is **less error prone**, since the matching is only performed once
    - **Allows** us to perform **impact change analysis**, both from **requirements AND data sources**
    - Enables **support for automatic analysis** tasks providing useful information about the current status of the DW

*Lucentia*

24

## Conclusions & Future Work

- Future work:
  - Define a complete set of QVT relations which use the defined semantics to derive DW models
  - Develop automatic analysis algorithms exploiting the information of traces
  - Consider the definition of an algebra for the relationships between requirements and data sources.

*Lucentia*

25

## Incorporating Traceability in Conceptual Models for Data Warehouses by using MDA

# Thank you

*Alejandro Maté*  
amate@dlsi.ua.es

Juan Trujillo  
jtrujillo@dlsi.ua.es

*Lucentia*