Requirements Engineering in Data Warehouses

Alejandro Maté
amate@dlsi.ua.es

Juan Trujillo
jtrujillo@dlsi.ua.es

PhD Colloquium at ER'11 2011
October 31th – November 3rd, Brussels, Belgium

Universitat d'Alacant
Universidad de Alicante

Lucentia

Content

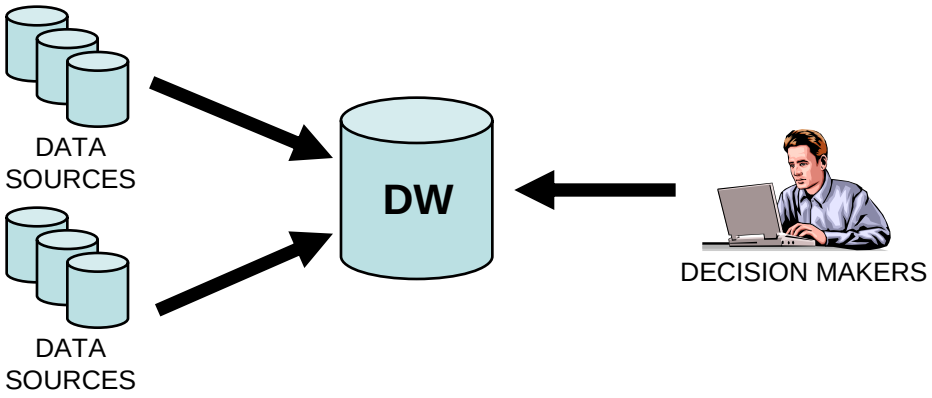
- Introduction
- Pitfalls in DW development
- Traceability as a solution
- Expected results & benefits
- Conclusions

Content

- Introduction
- Pitfalls in DW development
- Traceability as a solution
- Expected results & benefits
- Conclusions

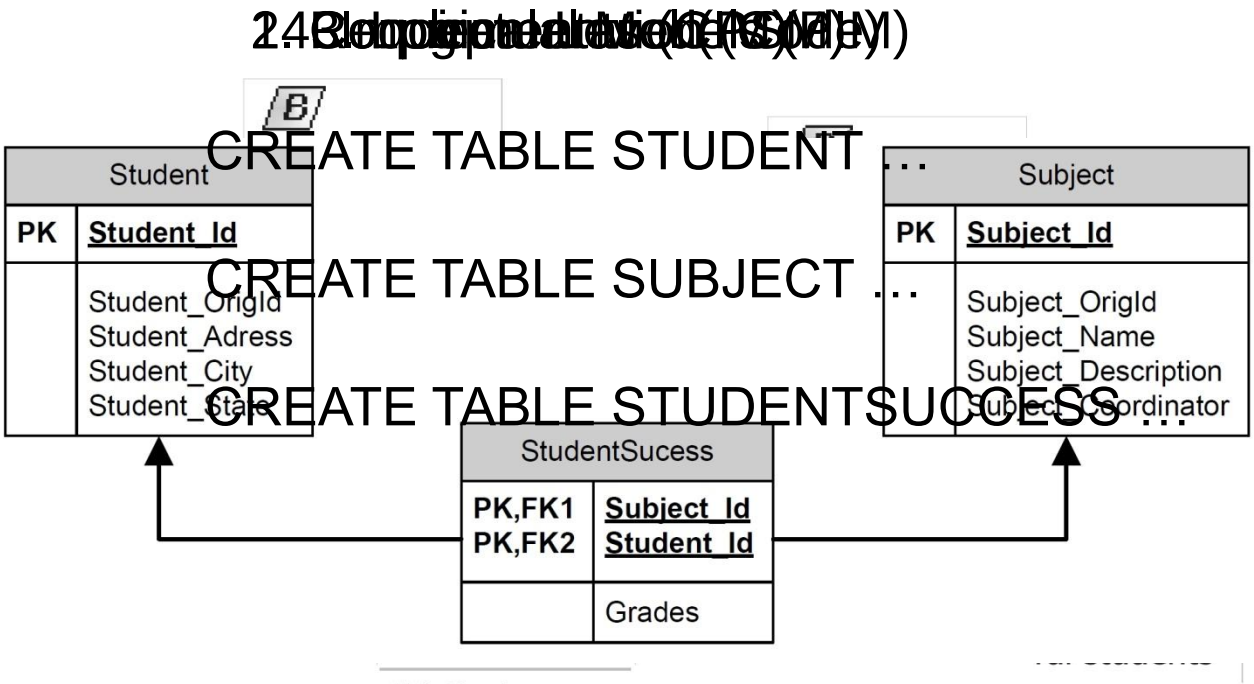
Introduction

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



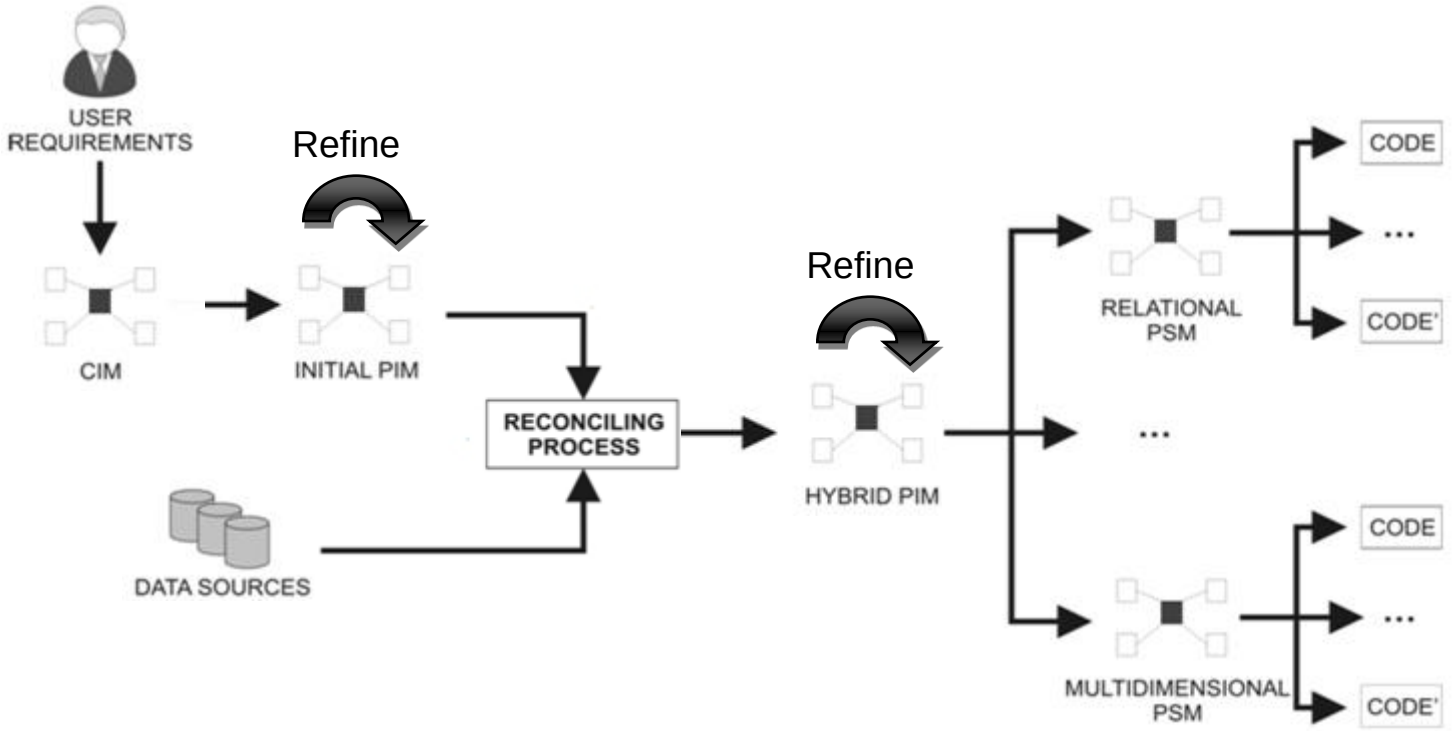
Introduction

- Development
  - Current development approaches make use of up to 4 layers:



Introduction

- Overview



Content

- Introduction
- Pitfalls in DW development
- Traceability as a solution
- Expected results & benefits
- Conclusions

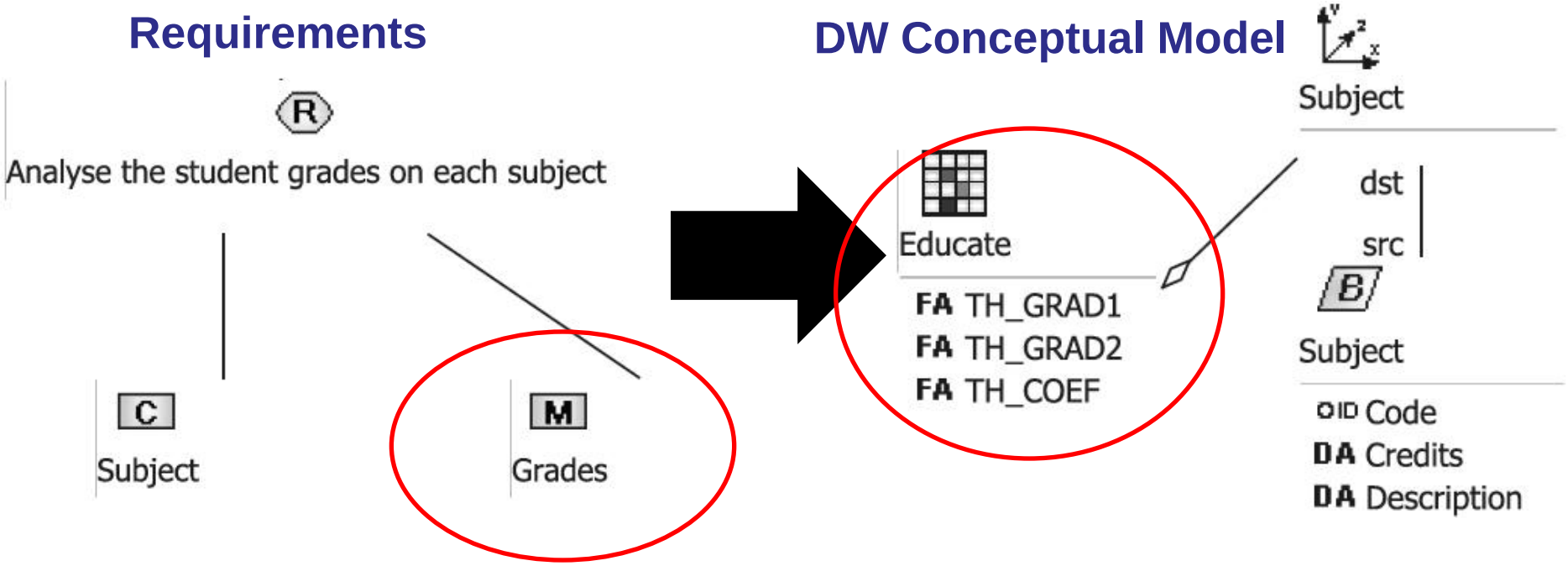
Pitfalls in DW development

- The lack of traceability makes us unable to perform operations over multiple models:
  - How do we calculate the "**quality**" of the DW?
  - Which requirements **cannot** be fulfilled?
  - How do we introduce **changes** without losing all the **previous work**?
  - Why perform the **matching** between requirements and sources **multiple times**?

Lucentia

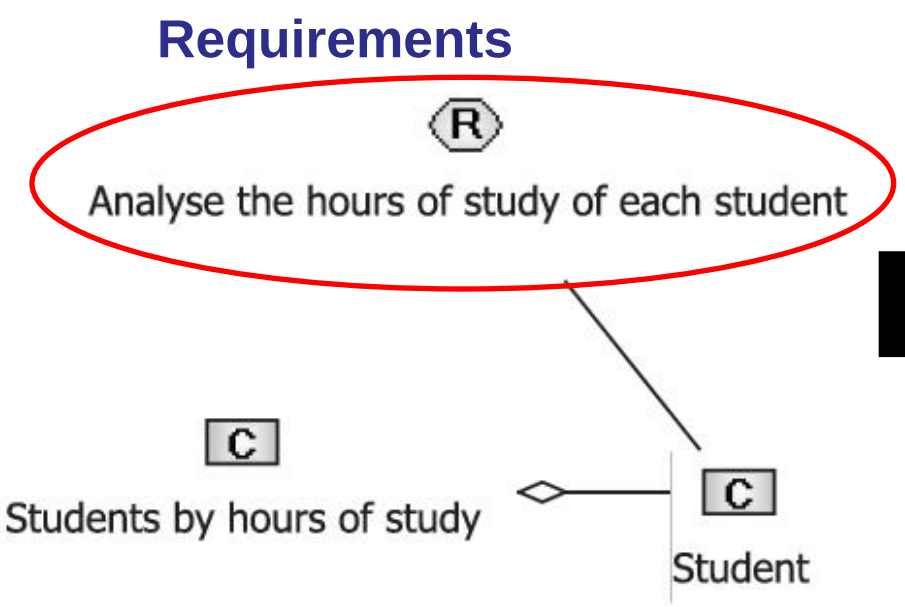
Pitfalls in DW development

- Quality of the DW
  - How ***complete*** is the **current design**?



Pitfalls in DW development

- Traceability of user requirements
  - Do we have the **necessary data**?



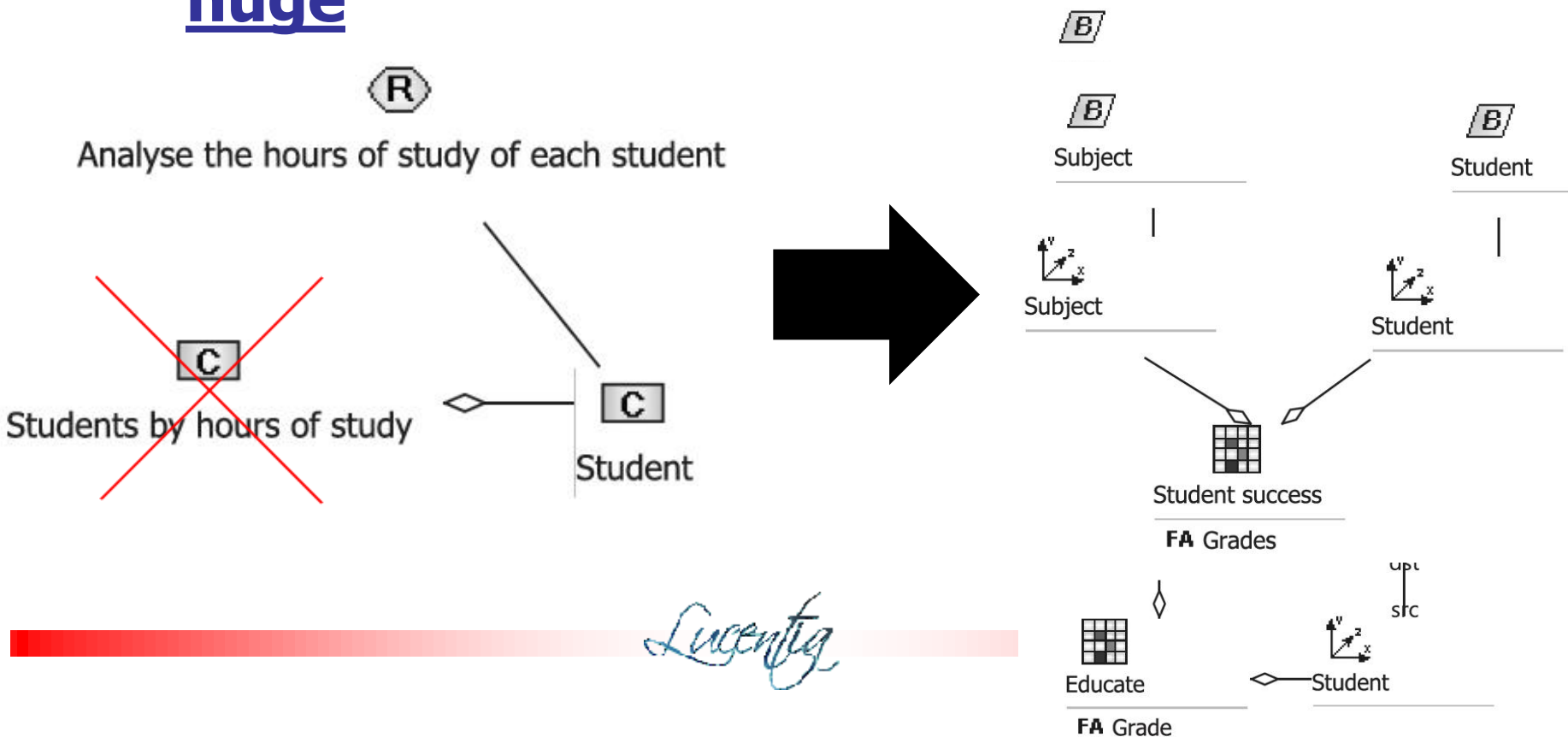
Out of 183 tables

…And I am not sure if this is the right table…

FI_PER	
PK	**file_code**
	hs_code id_code nam_name nam_app addres city birth_city birth_date

Pitfalls in DW development

- Propagation of changes
  - The **impact** of **little modifications** can be **huge**



Pitfalls in DW development

- Reconciliation
  - Whenever we **introduce or modify** an element, we have to **match it** against the data sources
    - **No data, no use**
  - Once we **finish** building **the DW**, we **still** have to **load the data**

Where did you say I had to put this….?

FI_PER	
PK	**file_code**
	hs_code id_code nam_name nam_app addres city birth_city birth_date

Content

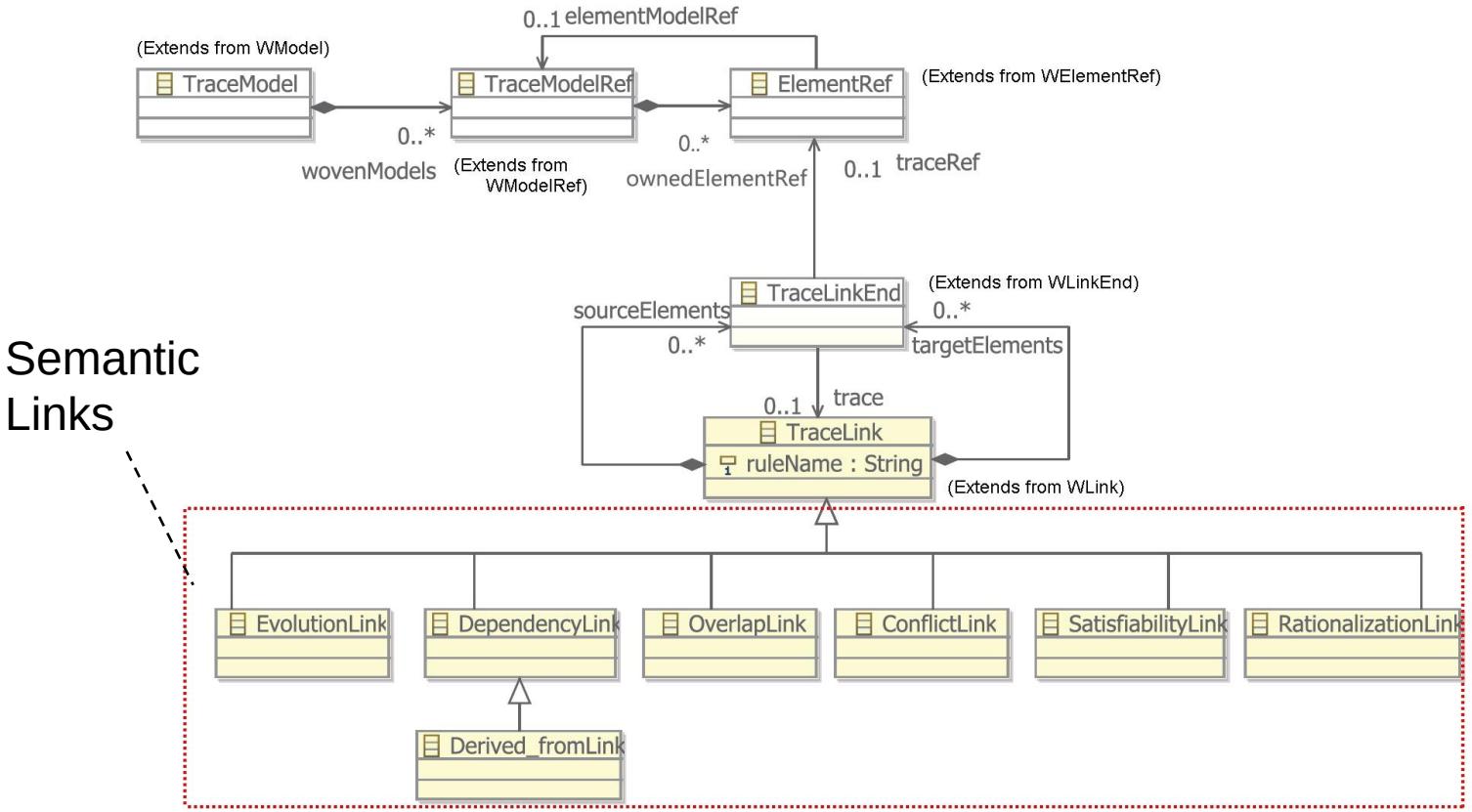
- Introduction
- Pitfalls in DW development
- Traceability as a solution
- Expected results & benefits
- Conclusions

Traceability as a solution

- Including traceability in the process:
  - First, define a **trace metamodel** with the necessary **semantics**
  - Second, **automate trace generation**
  - Third, **define** the necessary **trace models** and **restructure** the **current process**

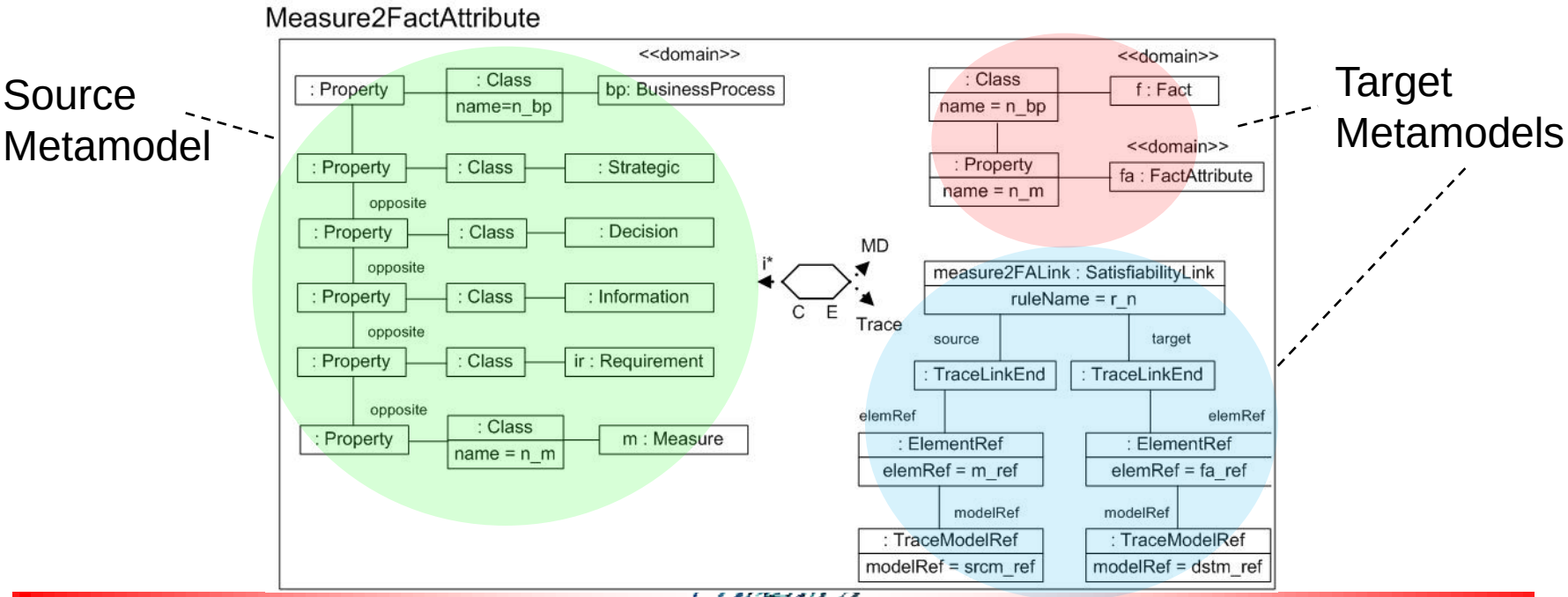
Traceability as a solution

- Trace metamodel:

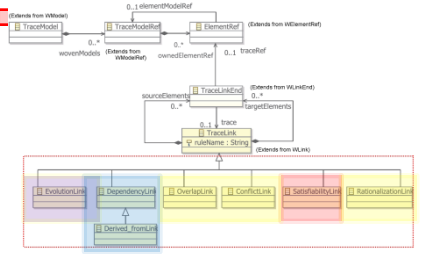


Traceability as a solution

- Automatic trace generation:
  - **Extend** the current **MDD approach considering** the **trace metamodel**



Traceability as a solution



1. Requirements

4.Reconciliation

2. Evolution

CIM

PIM2PIMTrace2

CIM2PIMTrace

PIM2PIMTrace1

PIM2PIMTrace3

Initial PIM

Hybrid PIM

Final PIM

Model to Model

Traces to Model

DS2PIMTrace

3.Datasource information

Datasource

Content

- Introduction
- Pitfalls in DW development
- Traceability for DW development
- Expected results & benefits
- Conclusions

Expected results & benefits

- Once we have support for traceability, we expect the following benefits:
  - **Quality metrics**: Traceability allows us to calculate quality metrics for the DW
    - E.g. number of implemented and non-implemented requirements, number of data sources, quality of the data, etc.
  - **Traceability of user requirements**: We are able to asses the status of each requirement at any point in development

Expected results & benefits

- **Propagation of changes**: We no longer need to be concerned about re-starting the process all over again due to changes

- **Reconciliation process**: The reconciliation process is no longer behaves as a black box
  - We have **explicit** record of the relationships between requirements and data sources
  - ETL processes now have an **initial plan** instead of starting from scratch

Lucentia

Content

- Introduction
- Pitfalls in DW development
- Traceability for DW development
- Expected results & benefits
- Conclusions

Conclusions

- Conclusions:
  - **Tasks** in DW development **require information** from **multiple levels** of the architecture
  - Development approaches lack **traceability** and are **unable** to **match elements** in different levels
  - Introducing traceability in the process **reduces time and development costs** at the same time as it helps to **increase the quality** of the DW

Requirements Engineering in Data Warehouses

QUESTIONS?

(and suggestions!)

Alejandro Maté
amate@dlsi.ua.es

Juan Trujillo
jtrujillo@dlsi.ua.es

PhD Colloquium at ER'11 2011
October 31th – November 3rd, Brussels, Belgium

Universitat d'Alacant
Universidad de Alicante

Lucentia