A Modularization Proposal for Goal-Oriented Analysis of Data Warehouses using i-star

Alejandro Maté *amate@dlsi.ua.es*
Juan Trujillo jtrujillo@dlsi.ua.es
Xavier Franch franch@essi.upc.edu

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

Universitat d'Alacant
Universidad de Alicante

Lucentia

Agenda

- Introduction
- Related work
- Definition of modules
- Example of application
- Experiments
- Conclusions & future work

Introduction

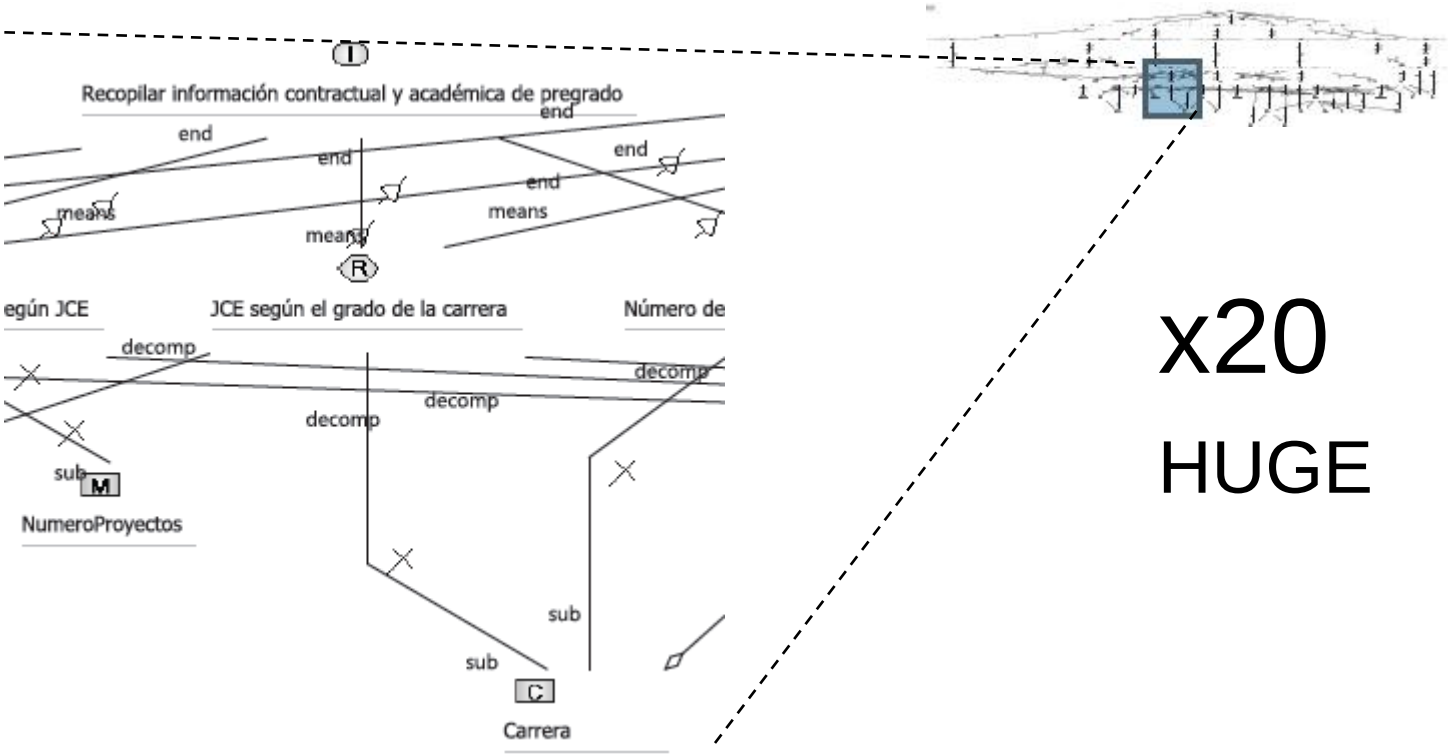
- Business Intelligence
  - Extracting **useful information** from the available **data** in order to take decisions
  - These data is usually stored at the Data Warehouse
    - Its structure must be designed according to the **users' needs**

Introduction

- Why i*?
  - i* helps us to **communicate** and **identify** relevant concepts for the DW
  - However, i* models **lack modularity**

Introduction

- Example



Related Work

- Related work

  [Franch 2010][Franch et al. 2011]

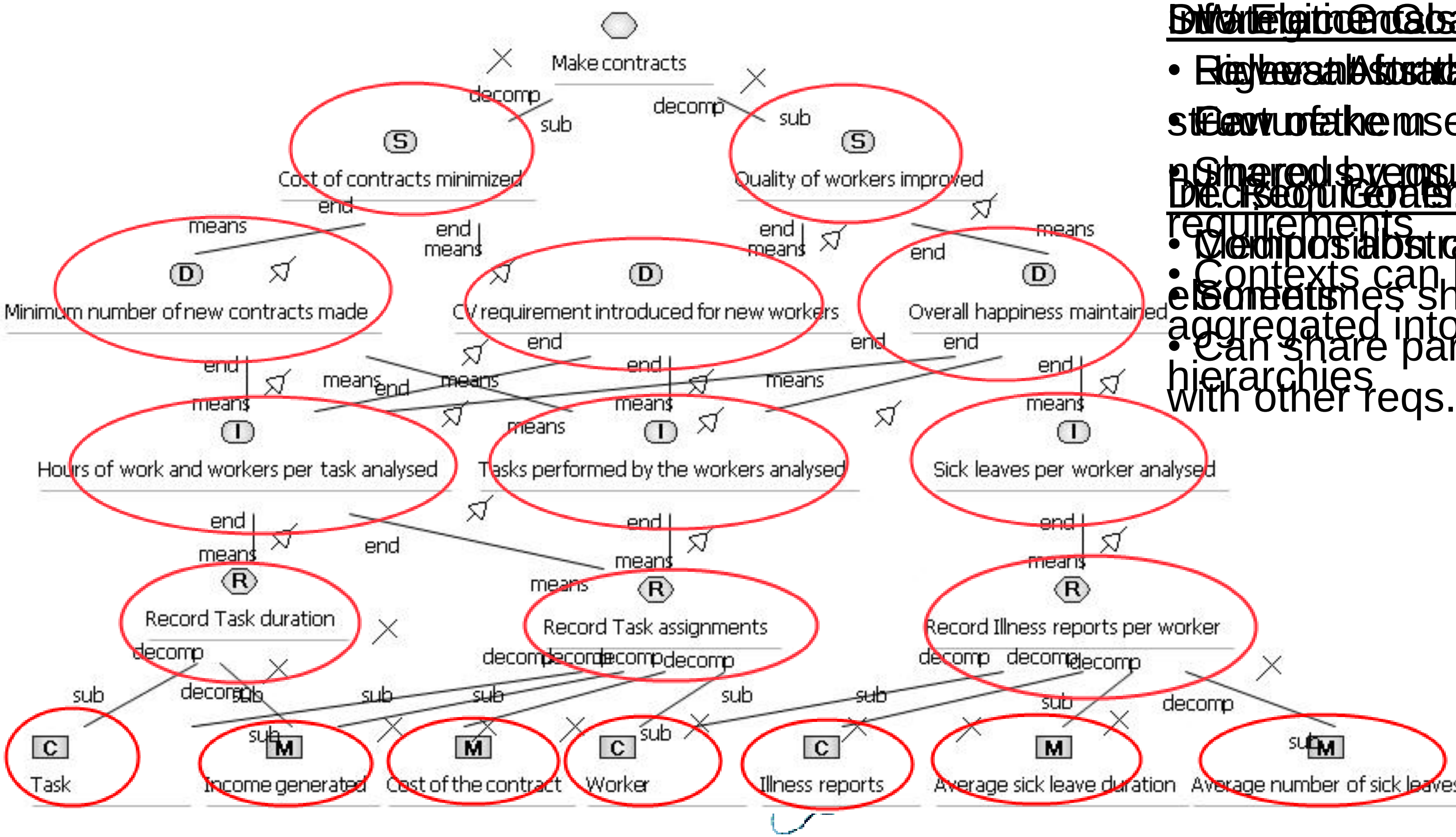
  - Other works have focused on incorporating modules on the **i* framework**
  - However, these modules **lack** any kind of **semantics**
  - Recently, it has been proposed to **tailor i* for the target domain**

Definition of Modules

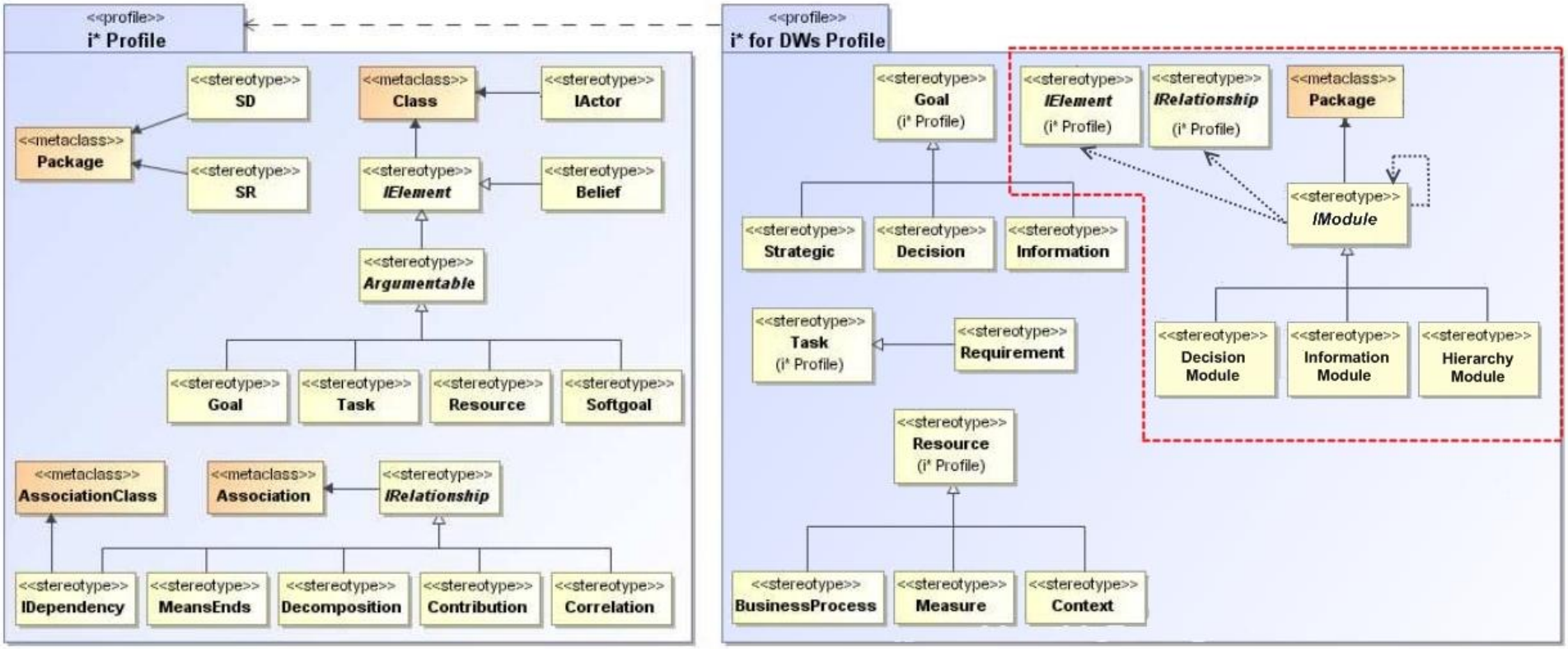
- The process consists of two steps:
  - First, perform an **ontological mapping** between i* and the target domain *(other work)*
  - Second, **analysis** and **definition of modules for the target domain**

Definition of Modules

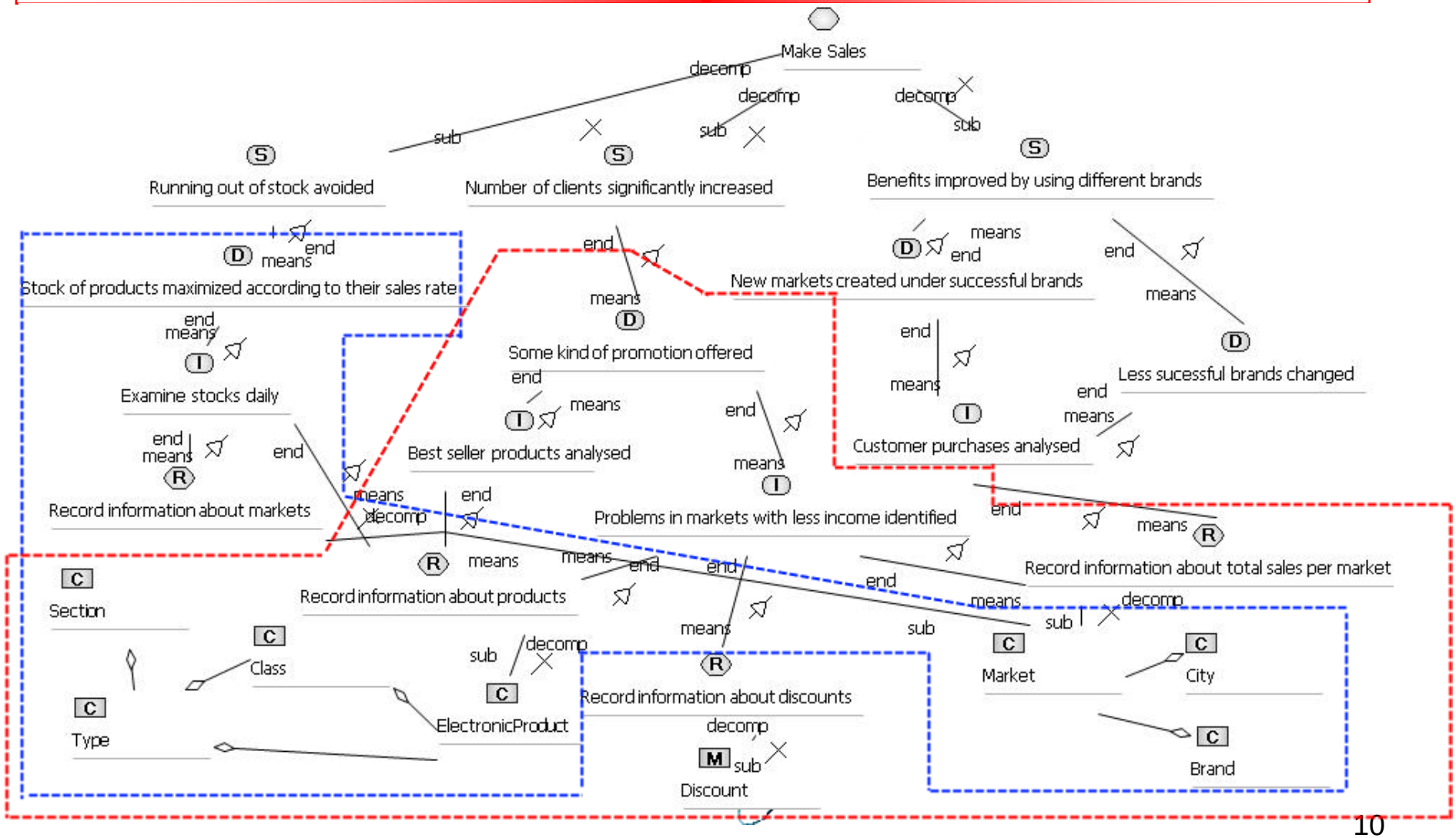
- There are 5 types of elements:



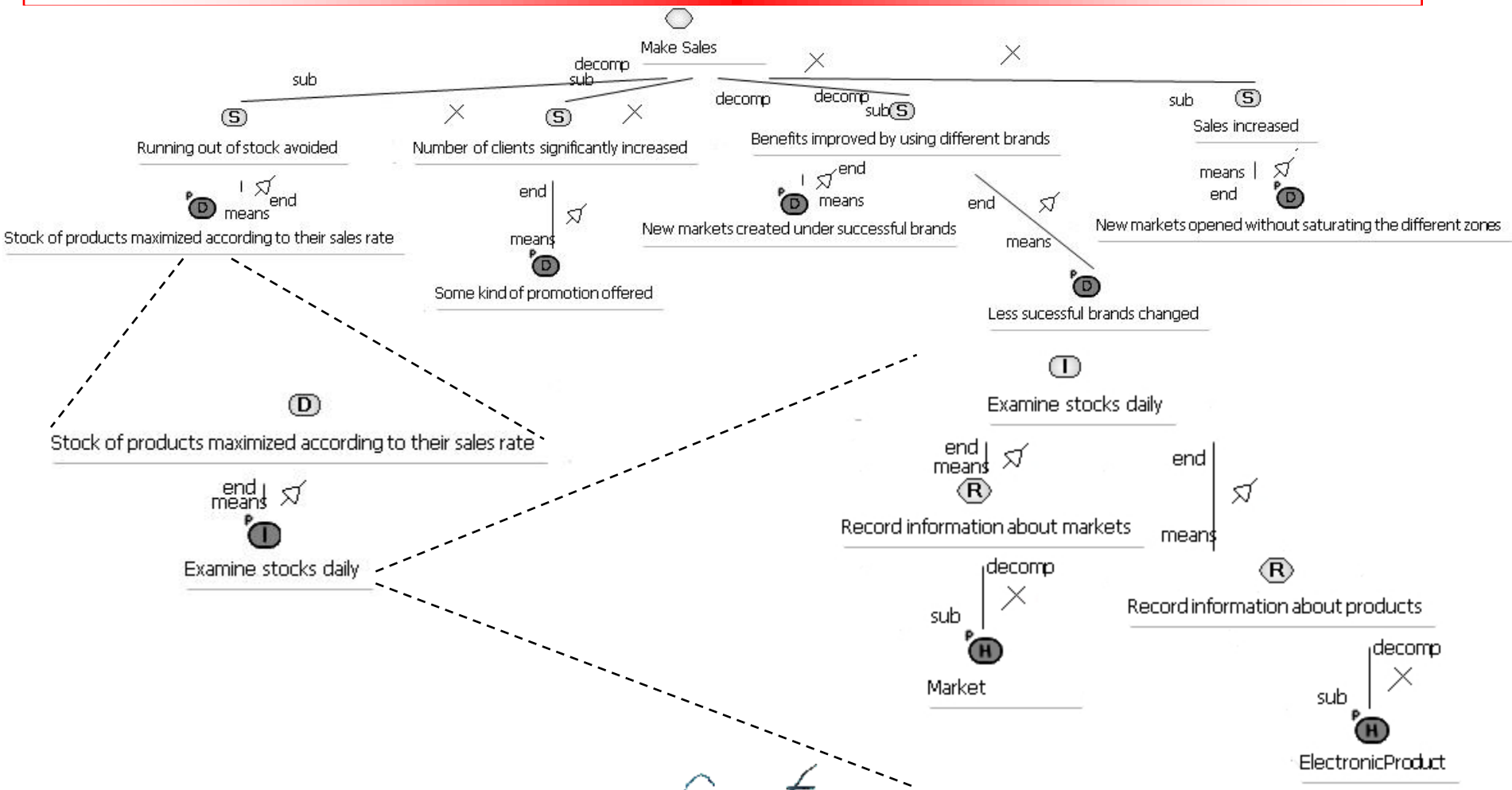
Definition of Modules



Example of application



Example of application



Experiments

- Which is the **impact** of the modularization **on designers**?
- **Correlation** between the **modularization**, and **time required** and **errors** in identification/modification tasks?
- We performed **two experiments** with people ranging from non-expert designers to experts on i* modeling

Experiment results

- First experiment tasks:
  - Identify **all** the elements related to a given decision goal
  - Identify **DW-only** elements related to another decision goal
  - Assign scores to the features perceived in the schema

Experiment results

	Monolitic	Modularized	ρ
Avg. reading time Sales	299.31	210.31	0.037
Identif. task 1 Sales	190.08	278.62	0.074
Identif. task 2 Sales	190.94	165.08	0.396
Avg. reading time Contracts	162.73	181.33	0.576
Identif. task 1 Contracts	150.07	211.5	0.112
Identif. task 2 Contracts	124.33	161.00	0.096
Avg. errors per questionnaire Sales	0.82	0.47	0.247
Avg. errors per questionnaire Contracts	0.33	0.36	0.906
Readability score Sales	2	1,93	0.826
Scalability score Sales	1,41	2,26	0.016
Comprehension score Sales	1,5	1,87	0.229
Modifiability score Sales	1,5	2,06	0.079
Readability score Contracts	2,27	2,33	0.803
Scalability score Contracts	1,67	2,41	0.011
Comprehension score Contracts	2,13	2,05	0.857
Modifiability score Contracts	1,73	2,17	0.128

Measured in seconds

Designers perceive “big improvements in scalability and modifiability

Experiment results

- Second experiment tasks:
  - Modification tasks over the models:
    - Two modification tasks over the sales model
    - One modification task over the contracts model
  - Design a goal model from scratch

Experiment results

	Monolitic	Modularized	ρ
Modif. task 1 Sales	202	154,27	0.327
Modif. task 2 Sales	223,6	290	0.217
Modif. task Contracts	128,73	197,6	0.002
Avg. Time drawing	1306,67	1891,44	0.019
Avg. Time/element	50,10	44,34	0.809
Avg. number of elements	25,67	42,89	0.000
Avg. unique non package elements	25,67	27,67	0.021

Measured in seconds

More elements discovered with modules!

Redundancy on monolitic design!

Conclusions & Future Work

- Conclusions:
  - We have presented a modularization proposal for DW including **semantics**
  - The modularization **improves the perceived scalability** of models
  - Performing tasks over modularized models:
    - Are **less error prone**
    - Consume **more time**

Lucentia

Conclusions & Future Work

- Future work:
  - Perform further experiments to validate other aspects of the proposal
  - Carefully evaluate the impact of the proposal for experts in DW design
  - Consider the simplification or addition of new modules

A Modularization Proposal for Goal-Oriented Analysis of Data Warehouses using i-star

QUESTIONS?

Alejandro Maté — *amate@dlsi.ua.es*
Juan Trujillo — jtrujillo@dlsi.ua.es
Xavier Franch — franch@essi.upc.edu

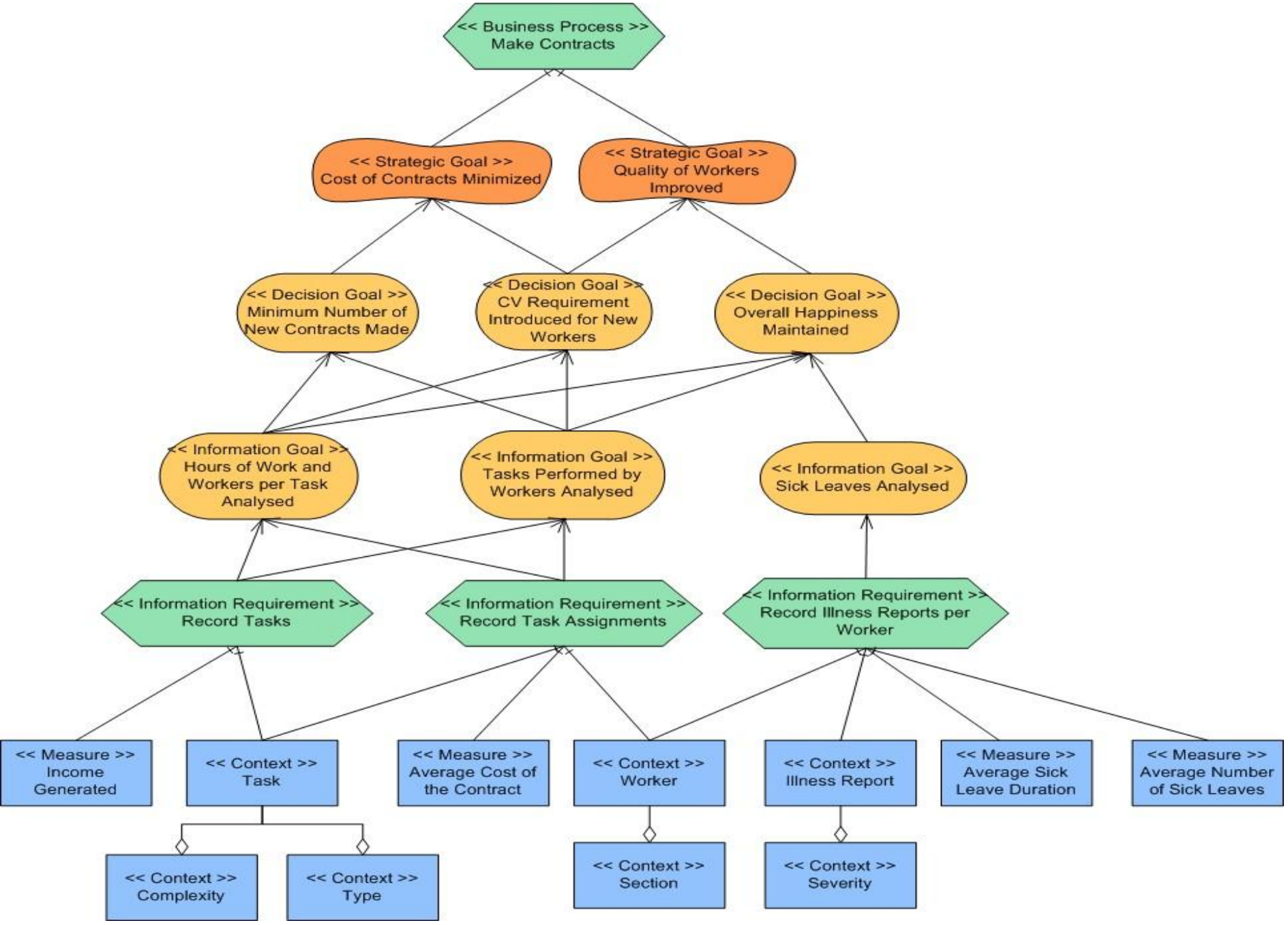
30th International Conference on Conceptual Modeling
ER'11 2011
October 31th – November 3rd, Brussels, Belgium

Universitat d'Alacant
Universidad de Alicante

Lucentia

Experiment results

- Experiment preliminars:
  - 2 models: contracts and sales
  - 4 versions of the questionnaires
  - Each person fills 1 version only
  - 1 sheet for monolitical models
  - 4 sheets for modularized models
  - Statistical analysis to filter outliers
  - The rest of the data is used to perform an ANOVA analysis ($\rho < 0.05$)


<< Business Process >>
Make Contracts
<< Strategic Goal >>
Cost of Contracts Minimized
<< Strategic Goal >>
Quality of Workers Improved
<< Decision Goal >>
Minimum Number of New Contracts Made
<< Decision Goal >>
CV Requirement Introduced for New Workers
<< Decision Goal >>
Overall Happiness Maintained
<< Information Goal >>
Hours of Work and Workers per Task Analysed
<< Information Goal >>
Tasks Performed by Workers Analysed
<< Information Goal >>
Sick Leaves Analysed
<< Information Requirement >>
Record Tasks
<< Information Requirement >>
Record Task Assignments
<< Information Requirement >>
Record Illness Reports per Worker
<< Measure >>
Income Generated
<< Context >>
Task
<< Measure >>
Average Cost of the Contract
<< Context >>
Worker
<< Context >>
Illness Report
<< Measure >>
Average Sick Leave Duration
<< Measure >>
Average Number of Sick Leaves
<< Context >>
Complexity
<< Context >>
Type
<< Context >>
Section
<< Context >>
Severity