

Semantic Cockpit*

An Ontology-driven, Interactive BI Tool for Comparative Data Analysis

Bernd Neumayr

Michael Schrefl

Johannes-Kepler-Universität Linz
Department of Business Informatics
Data & Knowledge Engineering (DKE)

Konrad Linner

solvistas GmbH

* SemCockpit is a collaborative research project
funded by the *Austrian Ministry of Transport, Innovation, and Technology*
in program *FIT-IT Semantic Systems and Services*
Duration: March 2011 - August 2013

presented at:

The 1st International Workshop on Modeling and Reasoning for Business Intelligence (MoRe-BI 2011)
Brussels, Belgium, October 31, 2011

Semantic Cockpit:

An Ontology-driven,

- conceptual models with formal semantics
- automatic reasoning
- automatic transformation to logical/physical level

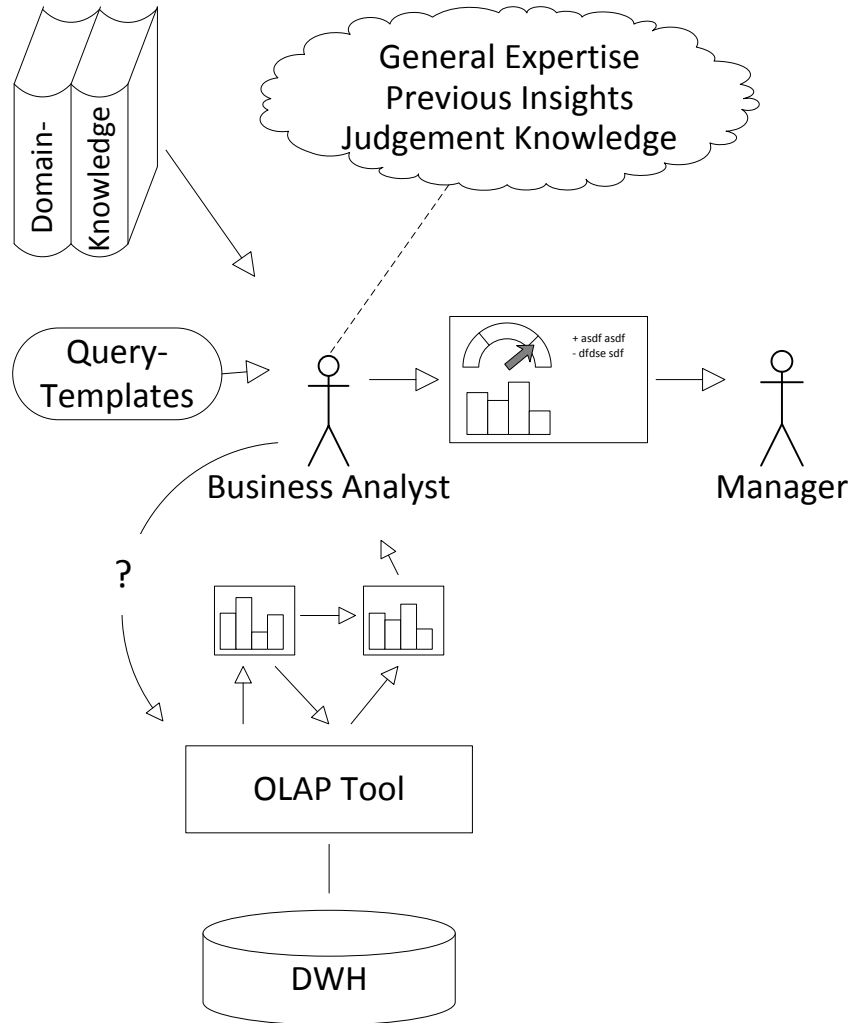
Interactive Business Intelligence Tool

- on top of existing data warehouse
- explorative data analysis
- assists BI designer
- assists and guides business analysts

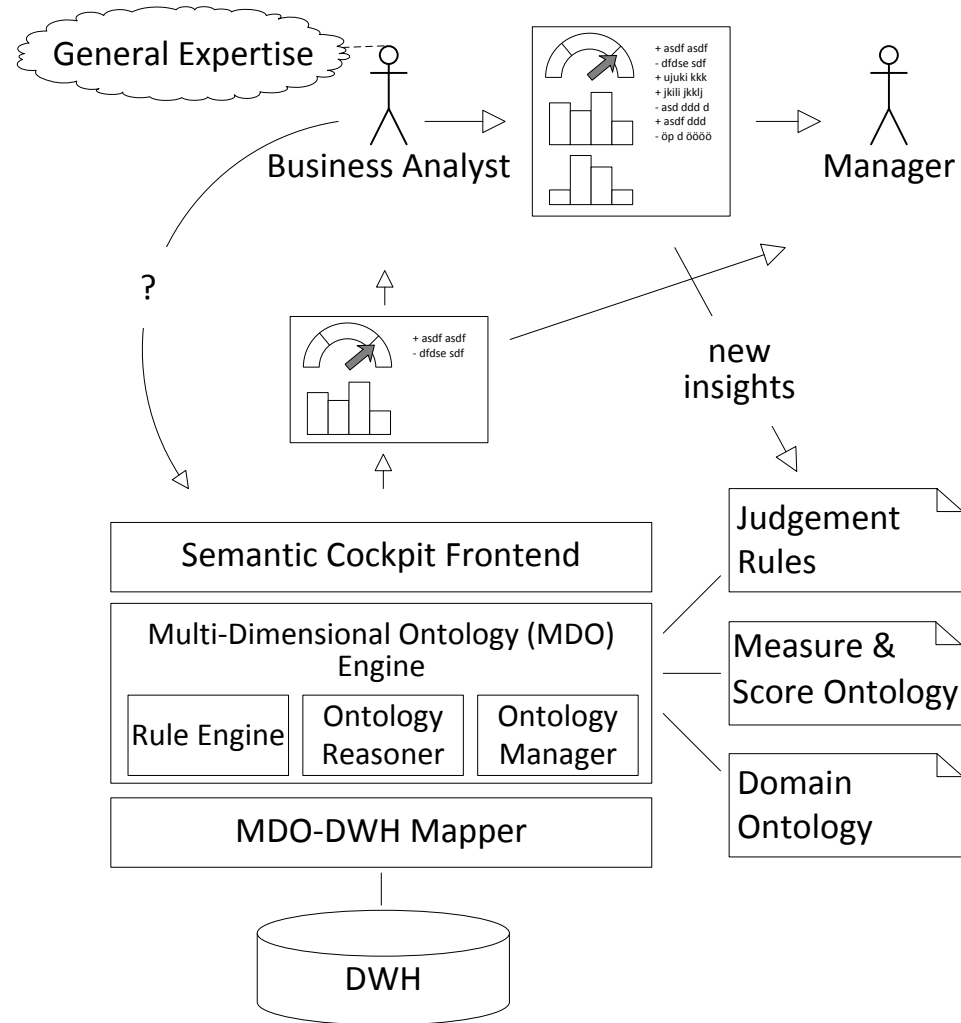
for Comparative Data Analysis

- inspect and compare data
- discover deviations or abnormalities

Conventional Data Analysis



Data Analysis using the Semantic Cockpit



Multidimensional Domain Ontologies

based on a conceptual, multi-dimensional data model

- dimensions, facts, levels, ...
- independent from DBMS and DWH modelling style

+ defined concepts (terminology)

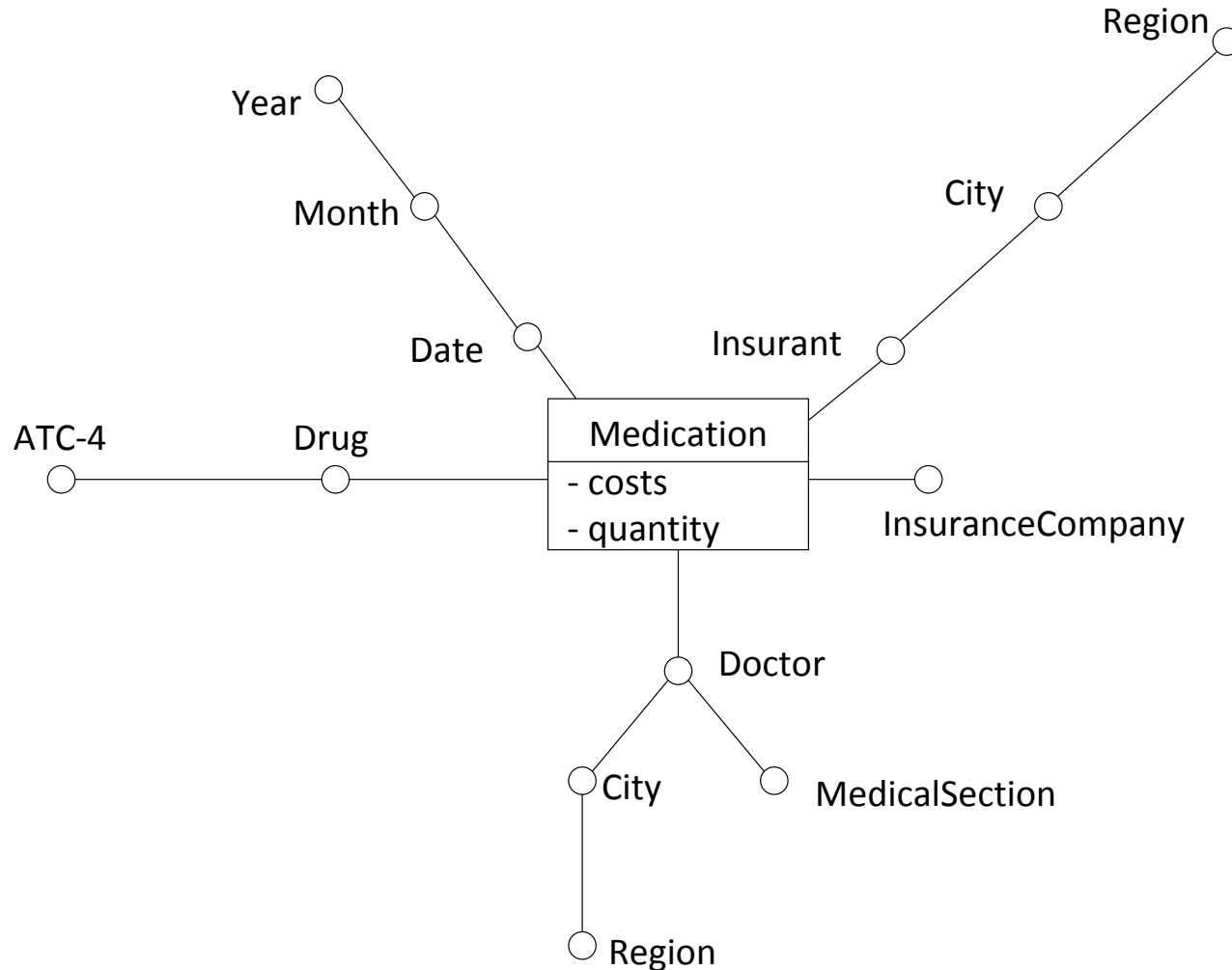
- single-dimensional (based on non-dimensional attributes)
- multi-dimensional (fact-based)

+ external ontologies

Reasoning tasks:

- Subsumption and consistency checking
(simplifies maintenance and facilitates query optimization)
- Query answering: transformation to SQL

Starting Point: Dimensions & Facts



Defined Concepts

CONCEPT **UrbanRegion**

LEVEL: Region

EQUIVALENT-TO: inhabitants/sqMiles > 1000

CONCEPT **BigCity**

LEVEL: City

EQUIVALENT-TO: inhabitants > 1,000,000

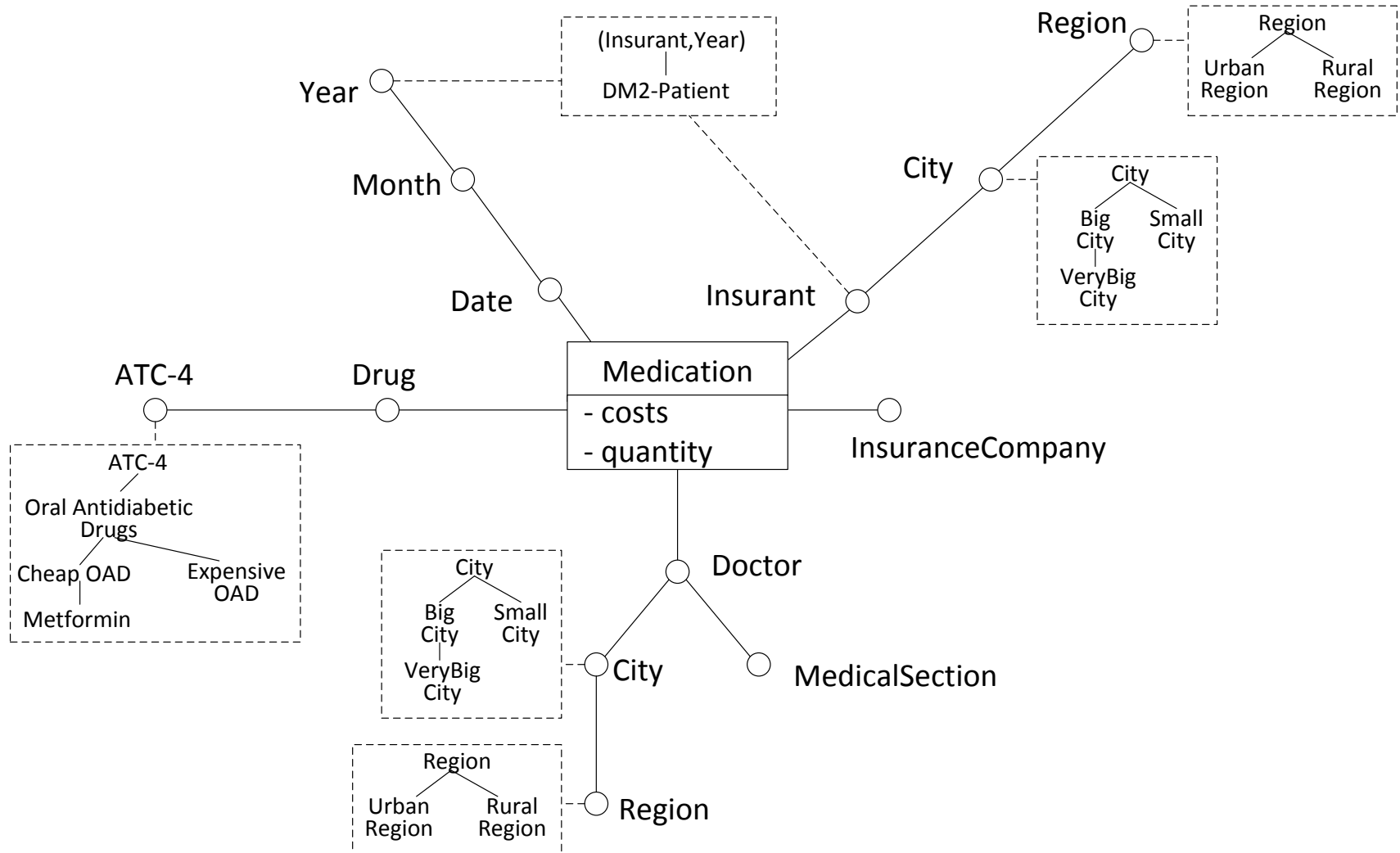
CONCEPT **VeryBigCity**

LEVEL: City

EQUIVALENT-TO: inhabitants > 5,000.000

...

Multidimensional Domain Ontology



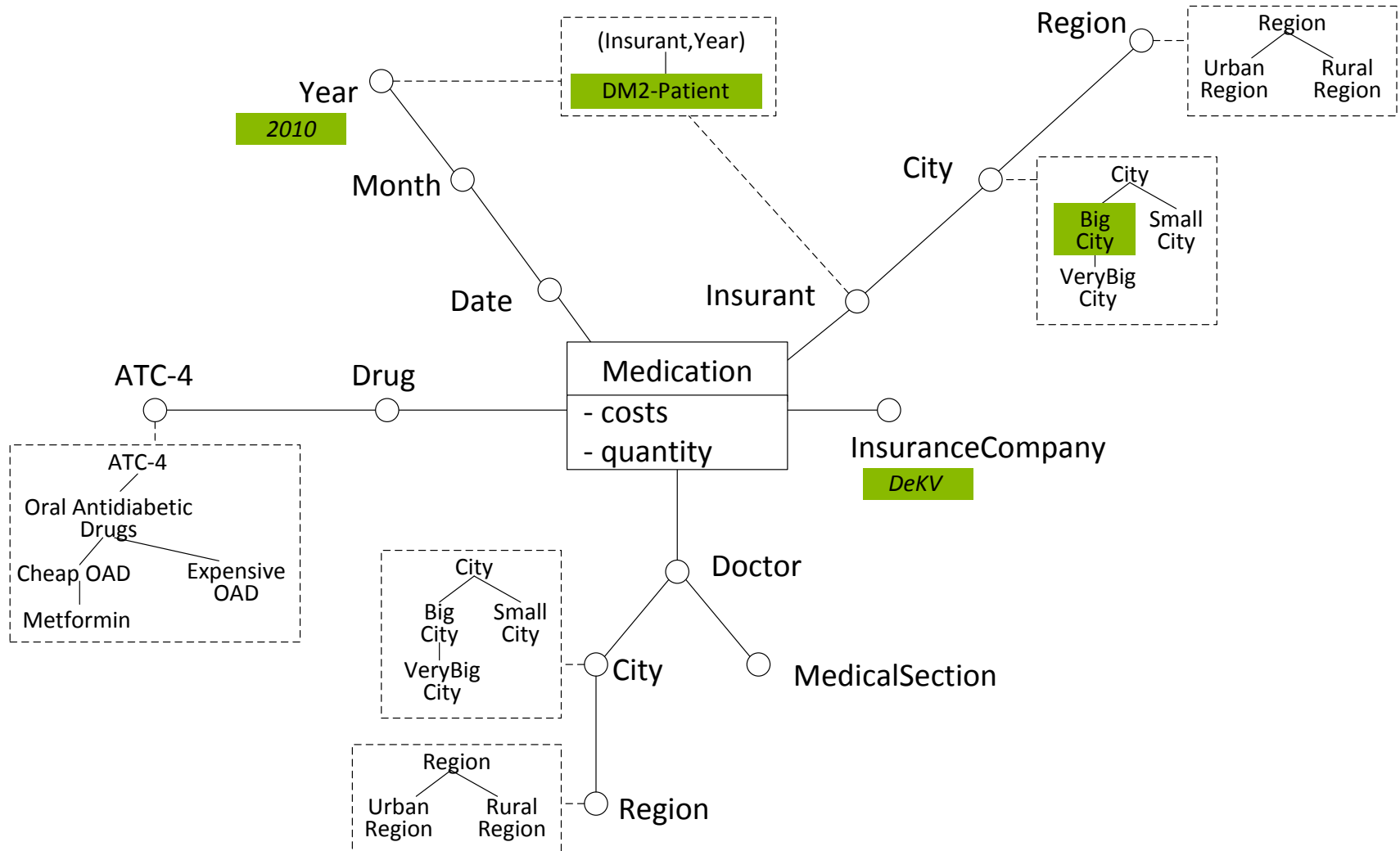
Querying with Scope Restrictions

- **Sample scope restriction:**

Medications paid for by insurance company DeKV in the year 2010 for insurants living in a *big city* and suffering from *diabetes mellitus 2* in this year:

```
Medication (  
    Time/Year:2010,  
    Insurant/City:[BigCity],  
    InsuranceCompany:DeKV  
) [DM2-Patient]
```


Querying with Scope Restrictions



Measure Ontologies

- **Measure definitions**

- at conceptual level
- based on other measures

- **Different types of measures**

- flexible-granularity vs. fixed-granularity
- simple vs. complex measures
- ratios, ...

- **Reasoning tasks**

- consistency checking
- discover relationships between measures
- summarizability checking
- query answering: transformation to SQL Query

Sample Measure Definitions

Flexible-Granularity Measures

```
MedicationCosts  
= SUM(Medication.Costs)
```

Fixed-Granularity Measures

```
MedicationCostsPerInsurant  
= MedicationCosts PER Insurant
```

Complex Measures

```
AvgMedicationCostsPerInsurant  
= AVG(MedicationCostsPerInsurant)
```

Ratios

```
RatioOfCheapOADs =  
MedicationCosts (Drug/ATC4 : [CheapOAD])  
/  
MedicationCosts (Drug/ATC4 : [OralAntidiabeticDrugs])
```

Measure Applications ...

Ratio of cheap oral antidiabetic drugs for insurance company DeKV in the year 2010:

```
RatioOfCheapOADs (  
    InsuranceCompany:DeKV,  
    Time/Year:2010  
)
```

... SQL query generation

Ratio of cheap oral antidiabetic drugs for insurance company DeKV in the year 2010:

```
RatioOfCheapOADs (
  InsuranceCompany:DeKV,
  Time/Year:2010
)
```



```
SELECT (
  SELECT SUM(m_costs)
  FROM f_medication
  WHERE d_drug IN (
    SELECT l_drug FROM h_drug
    WHERE l_atc4 IN (
      SELECT l_atc4
      FROM c_cheapoad)
    AND d_insurer = 'DeKV'
    AND d_time IN (
      SELECT l_day FROM h_time
      WHERE l_year = '2010'))
) / (
  SELECT SUM(m_costs)
  FROM f_medication
  WHERE d_drug IN (
    SELECT l_drug FROM h_drug
    WHERE l_atc4 IN (
      SELECT l_atc4
      FROM c_oad)
    AND d_insurer = 'DeKV'
    AND d_time IN (
      SELECT l_day FROM h_time
      WHERE l_year = '2010'))))
```

Comparative Scores

■ Comparisons

- Scope of interest
- Scope of comparison
- Scoring instruction

- Result: comparative score

■ Reasoning tasks

- Comparability of scope of interest and scope of comparison
- Query answering: transformation to SQL

Comparative Scores – Example

Evaluate prescribing habits of general practitioners with regard to oral antidiabetic drugs in the year 2010 in comparison with internists:

COMPARATIVE-QUERY:

SCOPE-OF-INTEREST: (

Doctor/medicalSection:GeneralPractitioner,

Time/Year:2010,

InsuranceCompany:DeVK)

SCOPE-OF-COMPARISON: (

Doctor/medicalSection:Internist)

SCORING-INSTRUCTION:

PercentageDifference (RatioOfCheapOADs)

Judgement Rules

Analysis-Situation, Condition → Recommended-Action
→ Explanation

Reasoning tasks

- Consistency and subsumption checking for analysis situations
- Detection of analysis situations

Judgement Rules – Example

- Whenever general practitioners score badly in comparisons with internists with regard to their prescribing habits for oral antidiabetic drugs, the system recommends an extra training for the general practitioners.

JUDGEMENT-RULE :

ANALYSIS-SITUATION:

SCOPE-OF-INTEREST:

(Doctor/medicalSection:GeneralPractitioner)

SCOPE-OF-COMPARISON:

(Doctor/medicalSection:Internist)

SCORING-INSTRUCTION:

PercentageDifference (RatioOfCheapOADs)

CONDITION: SCORE < -5

RECOMMENDATION: 'Extra Training'

Summary

- **Semantic cockpit**
 - is an interactive tool on top of existing data warehouses
 - is developed in the ongoing semCockpit project
 - assists and guides BI designers and business analysts
- **Knowledge represented as**
 - defined concepts
 - measures and scores
 - judgement and guidance rules
- **Reasoning tasks**
 - consistency and subsumption checking
 - summarizability checking
 - comparability checking
 - detection of analysis situations
 - query answering: transformation to SQL

State-of-the-Project

■ Finished work

- Overall architecture
- Multi-dimensional domain ontologies
- First proof-of-concept prototype (simple measures, simple domain ontologies, SQL transformation)

■ Current and future work

- Language constructs, reasoning techniques, implementation
 - measures, scores, judgement rules
- Mapping to database level
- Visualization and frontend

Thanks for your attention