

CNRS - INPT - UPS - UT1 - UTM



Institut de Recherche en Informatique de Toulouse

# Analysis Framework for Reduced Data Warehouse

Authors : Franck Ravat, **Jiefu Song**, Olivier Teste  
Team SIG, IRIT, Toulouse.

EDA-2015

Brussels, Belgium

# Plan

---

## 1. Context

## 2. Multi-states analyses

2.1 Data management

2.2 Analysis processing

## 3. Implementation

## 4. Conclusion

# Context

- General Context

- Data is stored permanently in a **M**ultidimensional **D**ata **W**arehouse (MDW)
- New data are periodically added

- Issues

- Low analysis performance due to important and increasing data volume.
- Decreasing relevance of detailed information with age.

- Objective

OLAP analyses in MDW keeping only useful data over time:

- Propose a complete analysis process.
- Facilitate decision-makers' tasks.

# Context

- Related Work

1. Modeling solutions for data reduction in MDW

- Partial (incomplete) solutions : either the fact or dimensional data updates [Chen et al. 2002], [Skyt et al. (2008)], [Kimball and Ross (2011)], [Golfarelli and Rizzi (2009)], [Iftikhar and Pedersen (2011)]
- ✓ Generalized data reduction process to the whole MDW schema [Atigui et al. (2014)]

2. Analysis framework compatible with reduced MDW

- No discussion about analysis-relative component [Chen et al. 2002], [Golfarelli and Rizzi (2009)], [Kimball and Ross (2011)], [Iftikhar and Pedersen (2011)]
- Limited discussion about querying solution in reduced MDW [Morzy and Wrembel (2004)], [Skyt et al. (2008)]
- ✓ Complete analysis support framework including
  - ▣ data management,
  - ▣ analysis processing
  - ▣ interactive data restitution.

# Plan

---

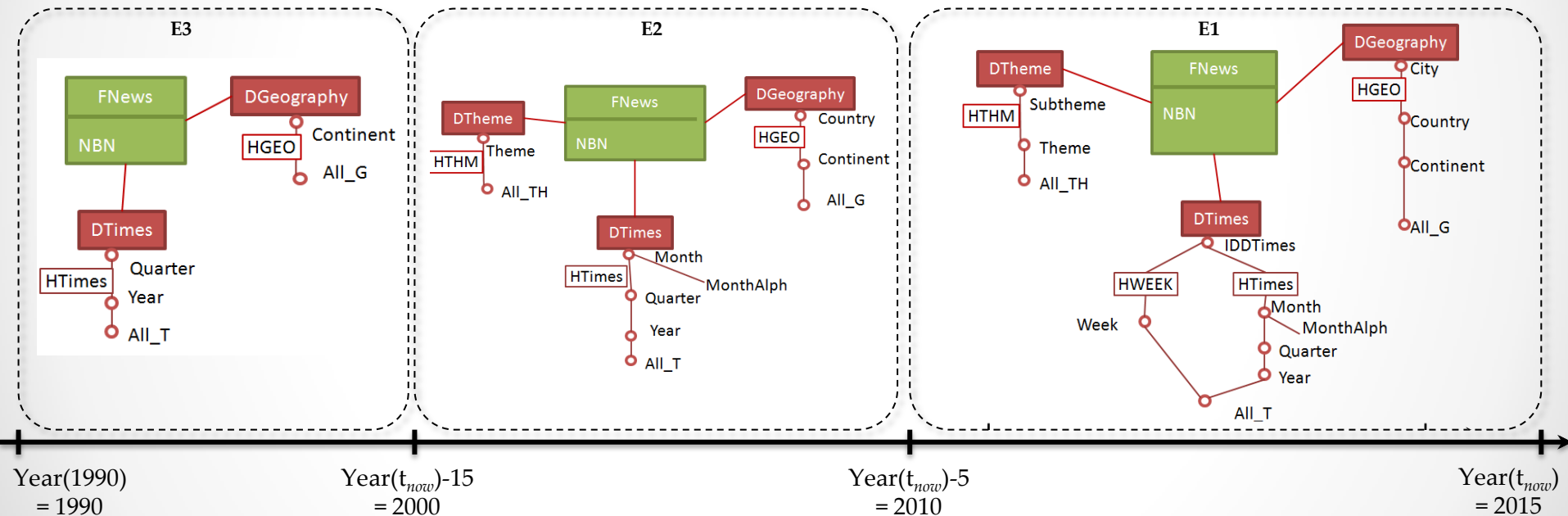
1. Context
2. Multi-states analyses
  - 2.1 **Data management**
  - 2.2 Analysis processing
3. Implementation
4. Conclusion

# Multi-states analysis

- Data management

Extended classical MDW model [Atigui et al. (2014)] :

- A MDW is composed of a set of star schemas;
- Each star schema, called state, is valid for a certain period of time.



# Multi-states analysis

- Data management  
Metamodel

Meta\_Dimension

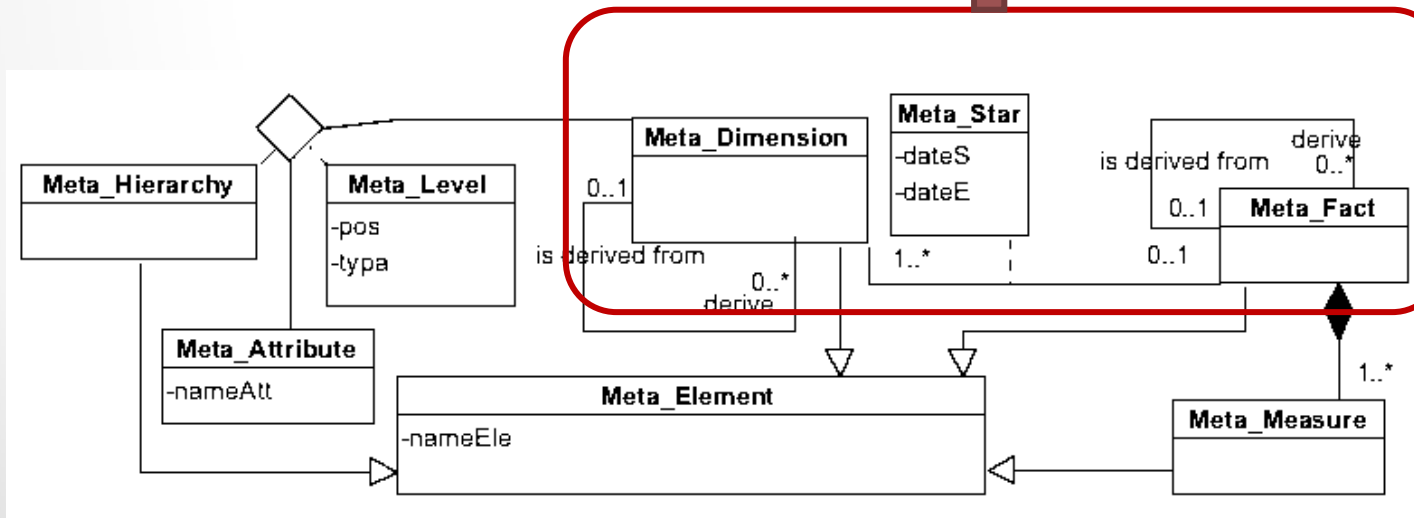
ID	NAME	IDP
d1	DGEOGRAPHY	
d2	DTIMES	
d3	DTHEME	
d4	DGEOGRAPHY_E1	d1
d5	DTIMES_E1	d2
d6	DTHEME_E1	d3

Meta\_star

IDF	NAME	IDD	NAME	DATES	DATEE
f2	FNEWS_E1	d5	DTIMES_E1	2010	2014
f2	FNEWS_E1	d4	DGEOGRAPHY_E1	2010	2014
f2	FNEWS_E1	d6	DTHEME_E1	2010	2014

Meta\_Fact

ID	NAME	IDP
f1	FNEWS	
f2	FNEWS_E1	f1



# Plan

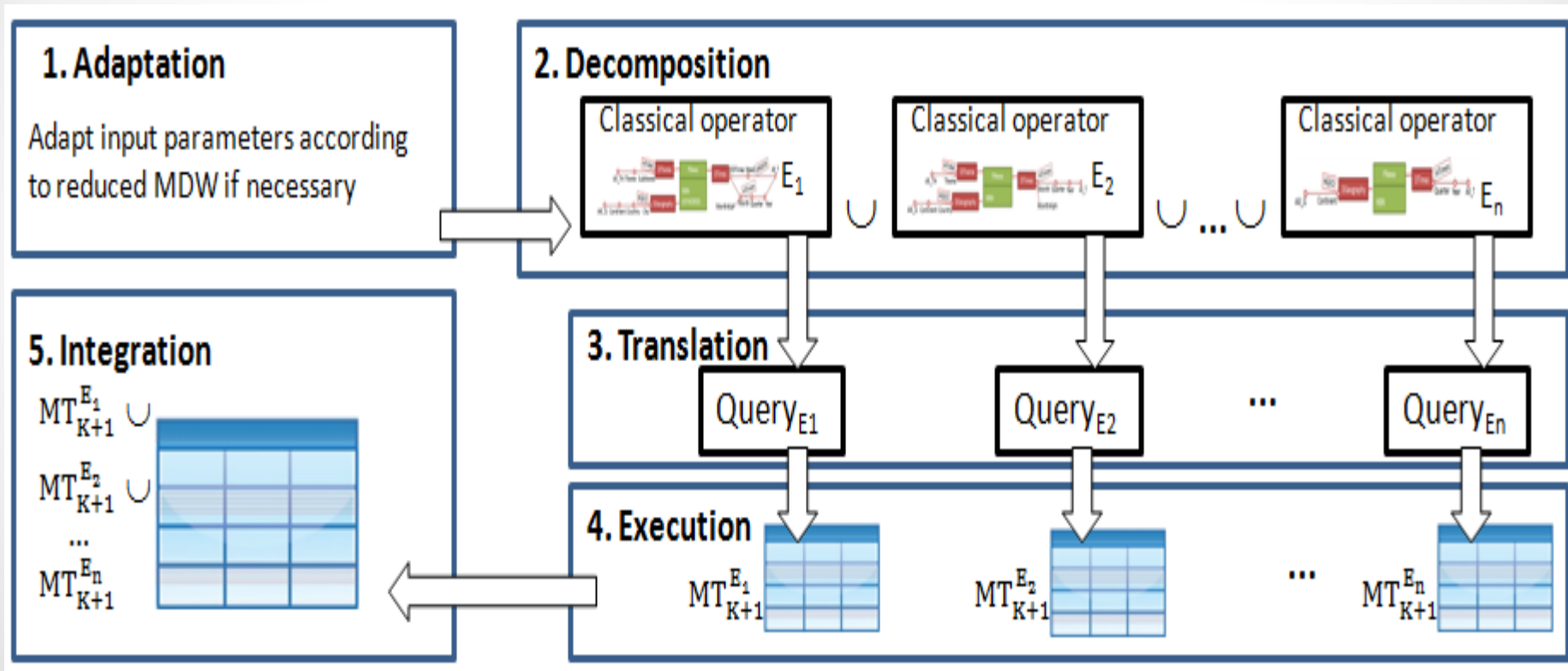
---

1. Context
2. Multi-states analyses
  - 2.1 Data management
  - 2.2 Analysis processing**
3. Implementation
4. Conclusion



# Multi-states analysis

- Analysis Engine  
Analysis processing



# Multi-states analysis

- Analysis Engine

Algebraic operator: Drilldown<sup>multi-states</sup>

**Algorithm 1:** *Drilldown<sup>multi-states</sup>* ( $I$ ;  $MT_k, D_i$ ;  $P_{inf}$ )

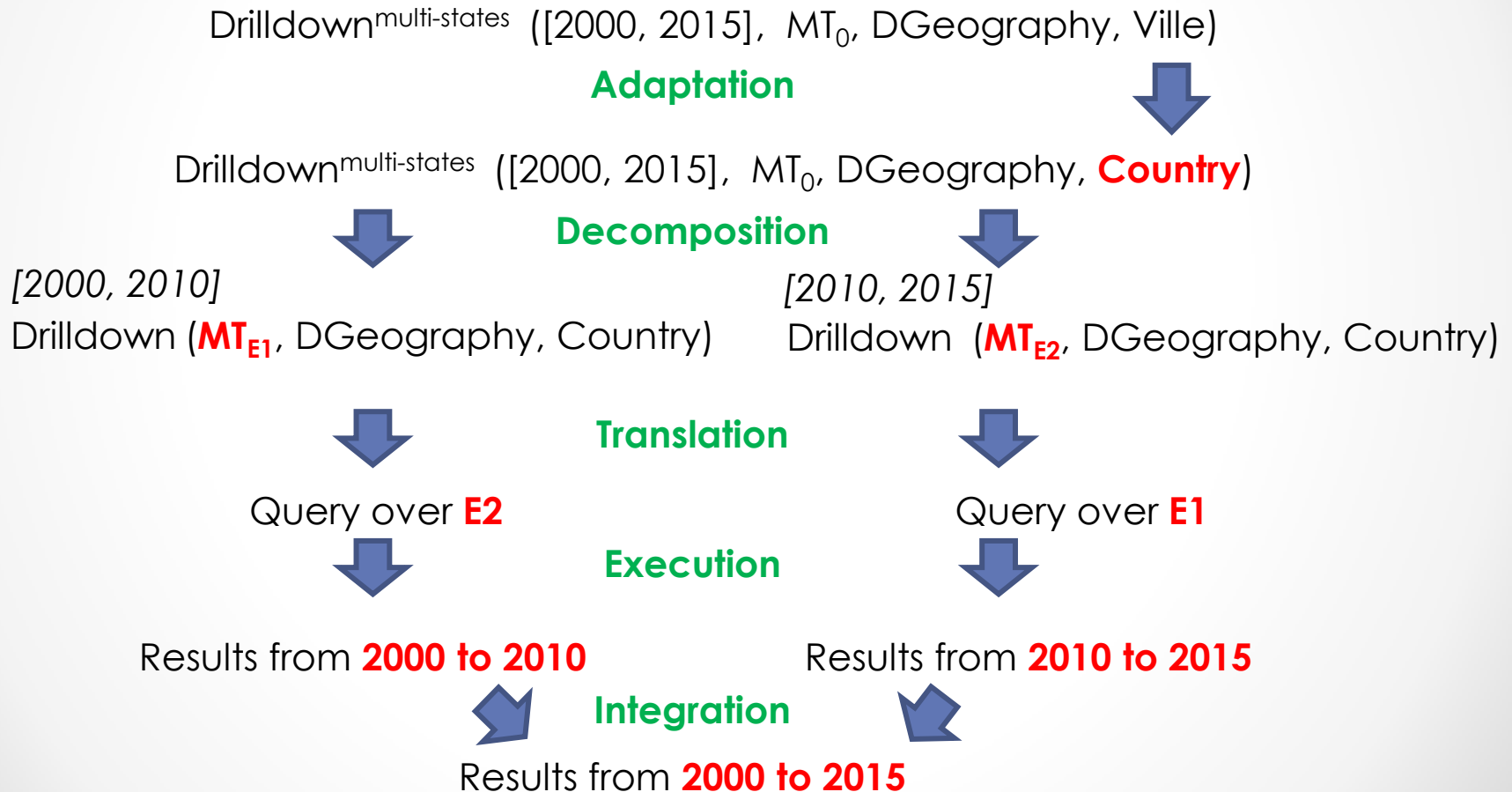
Input: Set of temporal intervals  $I$ , displayed multidimensional table  $MT_k$ , displayed dimension  $D$ , parameter  $P$ . Output: new multidimensional table  $MT_{k+1}$

```
1  Let  $H_{actual}$  be the actually displayed hierarchy
2  Let  $P_{actual}$  be the actually displayed parameter
3  If  $P_{actual} \prec^{H_{actual}} P \vee ALL_D \prec^{H_{actual}} P$  then
4    Impossible operation
5  Else
6    Find the subset of states  $E_i \rightarrow \forall E_j \in E_i | I_{E_j} \in I \vee I_{MT_k}$       -Adaptation
7    Let  $P_{Drilldown} = P$ 
8    Let  $r = FALSE$ 
9    While  $ALL_D \prec^{H_{actual}} P_{Drilldown} \wedge r = FALSE$ 
10     If  $\forall E_j \in E_i | P_{Drilldown} \in A_{E_j}^D$  then
11        $r = TRUE$ 
12     Else
13        $P_{Drilldown}$  increases one granularity level
14     End if
15   End While
16   If  $r = FALSE$  then
17     Impossible operation
18   Else
19     For  $E_j \in E_i$       - Decomposition
20       Let  $MT_K^{E_j}$  be the part of MT in states  $E_j$ 
21       Translate  $Drilldown(MT_K^{E_j}, D, P_{Drilldown})$  into query  $Q$       - Translation
22        $MT_{K+1}^{E_j} = \text{Result of query } Q$       - Execution
23        $MT_{K+1} = MT_{K+1} \cup MT_{K+1}^{E_j}$       - Integration
24     End for
25   End if
26 End if
```

# Multi-states analysis

- Analysis Engine

Example:  $MT_0$  - Number of published news by Continent from 2000 to 2015



Context

Multi-states analysis

Implementation

Conclusion

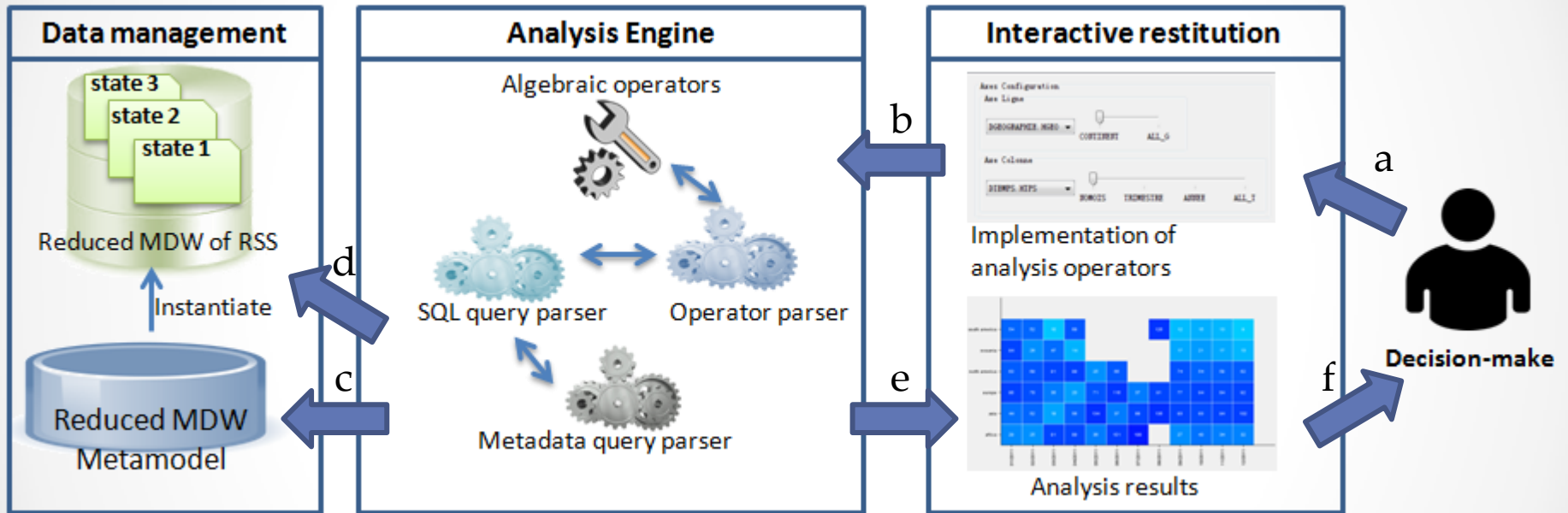
# Plan

---

1. Context
2. Multi-states analyses
  - 2.1 Data management
  - 2.2 Analysis processing
- 3. Implementation**
4. Conclusion

# Implementation

- Architecture of analysis framework



Context

Multi-states analysis

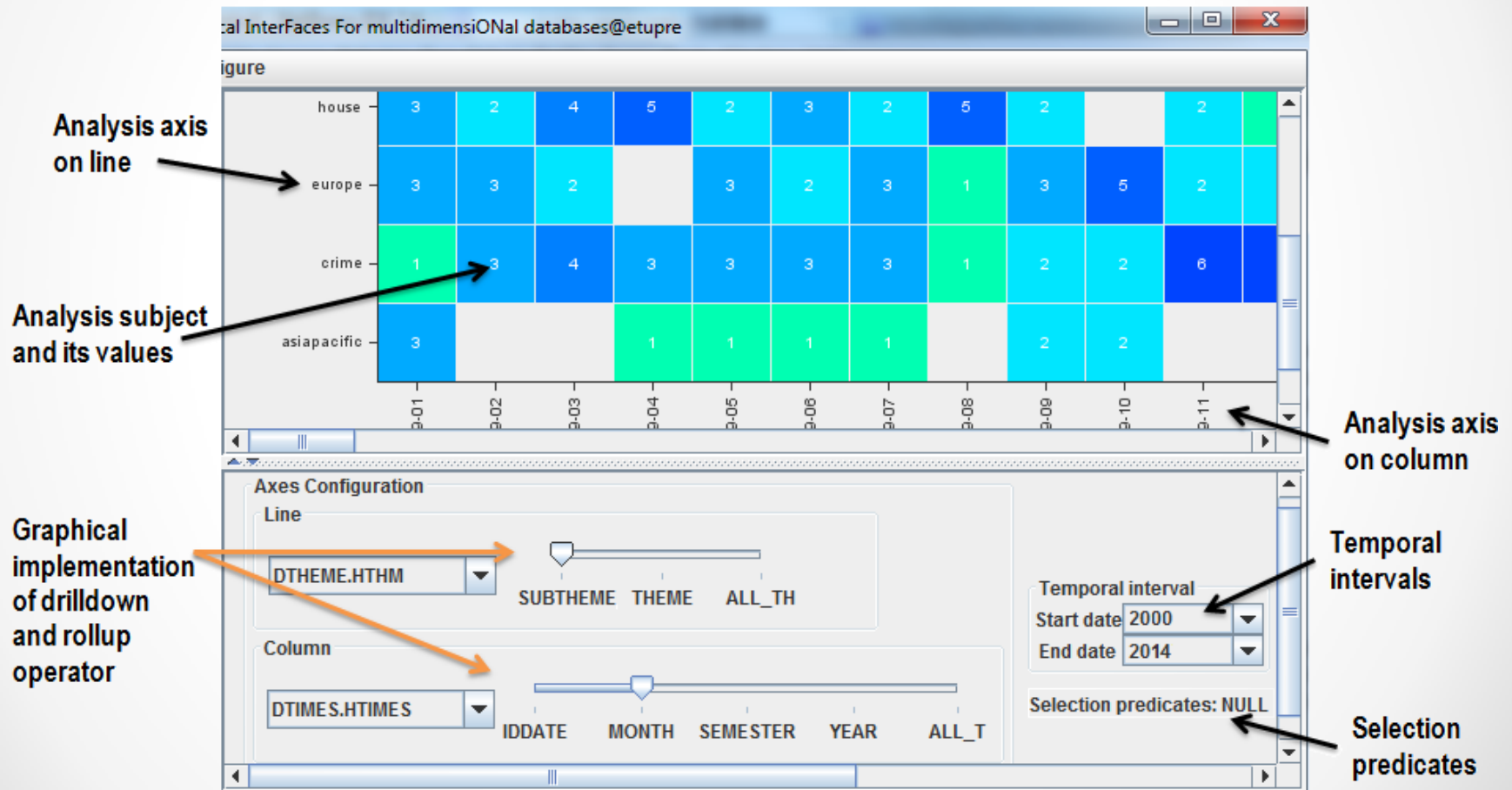
Implementation

Conclusion

# Implementation

- Interactive restitution

Graphical interface



Context

Multi-states analysis

Implementation

Conclusion

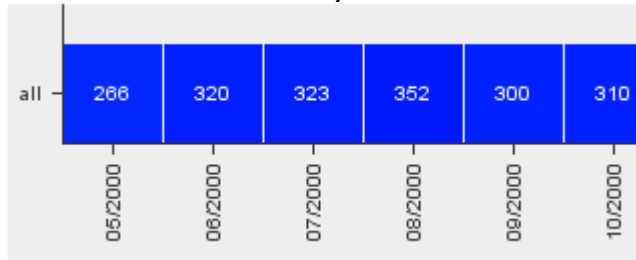
# Implementation

- Interactive restitution

Example:

published news **in the world** by month from 2000 to 2014

Drilldown<sup>multi-states</sup>



published news by month and **by continents** from 2000 to 2014

```
SELECT SL.FACT_NAME AS FACT,
       SL.DIMENSION_NAME AS DIML,
       SC.DIMENSION_NAME AS DIMC
FROM META_STAR SL,
     META_FACT F,
     META_DIMENSION DL,
     META_DIMENSION DC,
     META_STAR SC
WHERE SL.NAME = F.NAME
      AND SC.NAME = SL.NAME
      AND F.NAME_PARENT = 'FNEWS'
      AND DL.NAME = SL.DIMENSION_NAME
      AND DC.NAME = SC.DIMENSION_NAME
      AND DL.NAME = 'DGEOGRAPHIE'
      AND DC.NAME = 'DTEMPS'
      AND SL.DATES <= TO_DATE
        ('30-12-2014', 'DD-MM-YYYY')
      AND SC.DATEE >= TO_DATE
        ('01-01-2000', 'DD-MM-YYYY')
```



```
(SELECT SUM(NBN) AS FNEWS, DIML.COUNTRY, DIMC.NUMMONTH
FROM FNEWS_E1 FAIT, DGEOGRAPHY_E1 DIML, DTIMES_E1 DIMC, DTIMES_E1 DIMIT
WHERE FAIT.ID_DGEOGRAPHY_E1 = DIML.ID_DGEOGRAPHY_E1
      AND FAIT.ID_DTIMES_E1 = DIMIT.ID_DTIMES_E1
      AND TO_DATE(DIMIT.NUMMONTH, 'MM-RRRR')
        BETWEEN TO_DATE('01-01-2000', 'DD/MM/RRRR')
          AND TO_DATE('30-12-2014', 'DD/MM/RRRR')
      AND FAIT.ID_DTIMES_E1 = DIMC.ID_DTIMES_E1
GROUP BY DIML.COUNTRY, DIMC.NUMMONTH )
UNION
(SELECT SUM(NBN) AS FNEWS, DIML.COUNTRY, DIMC.NUMMONTH
FROM FNEWS_E2 FAIT, DGEOGRAPHY_E2 DIML, DTIMES_E2 DIMC, DTIMES_E2 DIMIT
WHERE FAIT.ID_DGEOGRAPHY_E2 = DIML.ID_DGEOGRAPHY_E2
      AND FAIT.ID_DTIMES_E2 = DIMIT.ID_DTIMES_E2
      AND TO_DATE(DIMIT.NUMMONTH, 'MM-RRRR')
        BETWEEN TO_DATE('01-01-2000', 'DD/MM/RRRR')
          AND TO_DATE('30-12-2014', 'DD/MM/RRRR')
      AND FAIT.ID_DTIMES_E2 = DIMC.ID_DTIMES_E2
GROUP BY DIML.COUNTRY, DIMC.NUMMONTH )
```

**Metamodel**

**Model RSS Stream**

Context

Multi-states analysis

Implementation

Conclusion

# Implementation

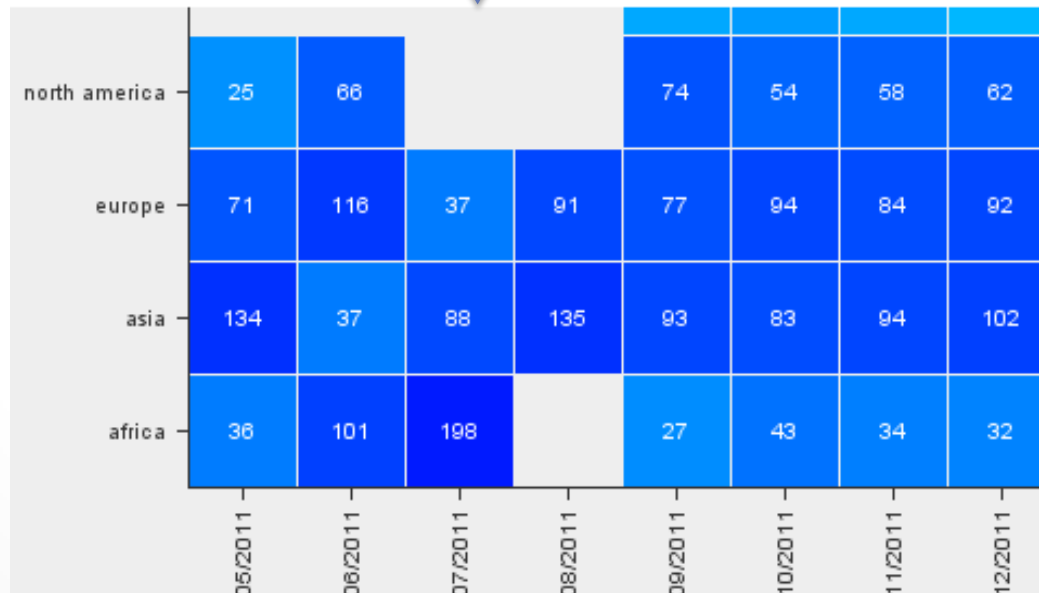
- Interactive restitution

Example:

published news **in the world** by month from 2000 to 2014

Drilldown<sup>multi-states</sup>

published news by month and **by continents** from 2000 to 2014





# Plan

---

1. Context
2. Multi-states analyses
  - 2.1 Modeling solution
  - 2.2 Analysis processing
3. Implementation
4. **Conclusion**

# Conclusion

- A MDW model composed of a set of states over time.
  - ✓ Decreases data volume
  - ✓ Facilitate decision-makers' tasks
- A generic multi-states analysis framework
  - ✓ Graphical interface: interaction with decision-makers
  - ✓ Analysis engine: algebraic operators and its execution algorithms
  - ✓ Data management: meta-model and its instances
- Implementation : graphical multi-states analysis tool
  - ✓ Automatic generation of queries
  - ✓ Transparency of data reduction
- Future work
  - ✓ Influence of data reduction over pre-aggregated data
  - ✓ Other analysis operators