
Geo Data Mining and Knowledge Discovery

Monica Wachowicz

Technical University of Madrid
ISTI, CNR

Chiara Renso

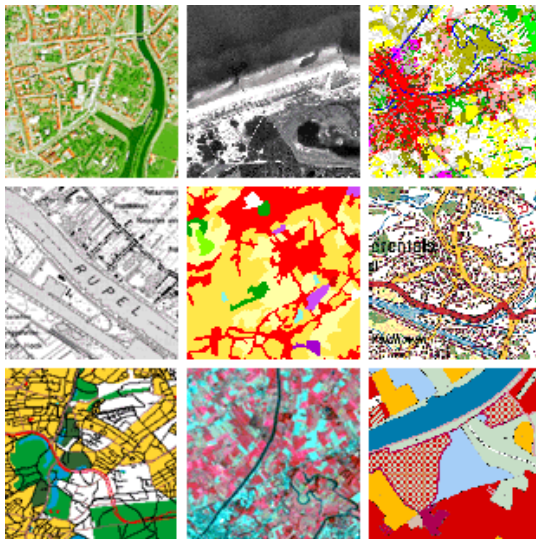
KDDLab,

Introduction

Geo Data Mining

How we can explore millions of records, hundreds of fields, and find patterns in spatial databases?

200 mining
tools !!!!



BUT

several barriers
to apply them to
georeferenced
data

What is special about geo data mining?

- the need for spatial data representations
 - Raster: gridded space
 - Vector: point, line, polygon
 - Graph: node, edge, path
- the visual data exploration process to generate a new hypothesis
- the role of spatial interest measures in finding previous unknown spatial patterns

What is special about geo data mining?

- the need for spatial data representations
 - Raster: gridded space

First Law of Geography

"All things are related, but nearby things are more related than distant things. [Tobler, 1970]"

to generate a new hypothesis

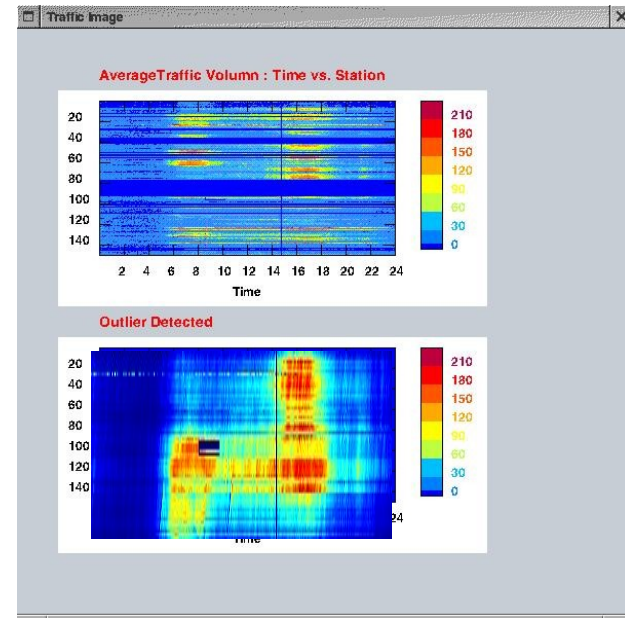
- the role of spatial interest measures in finding interesting spatial patterns

**What geo data
mining techniques
have been**

**successfully applied
?**

Some examples include

❖ Spatial Outliers and Discontinuities

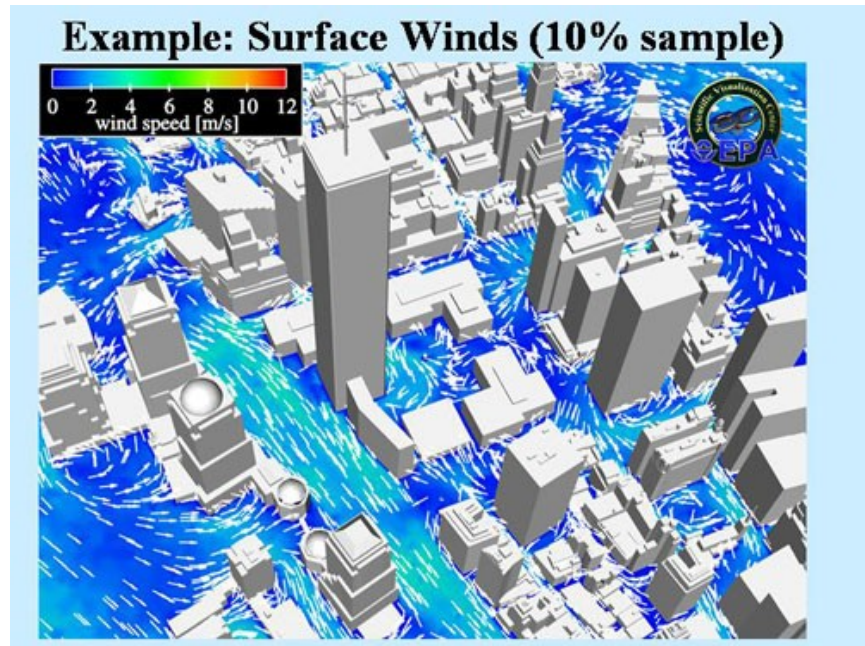


**JSDM Version 0.1 A Java Based Spatial
Data Mining Software**

Some examples include

❖ Spatial Outliers and Discontinuities

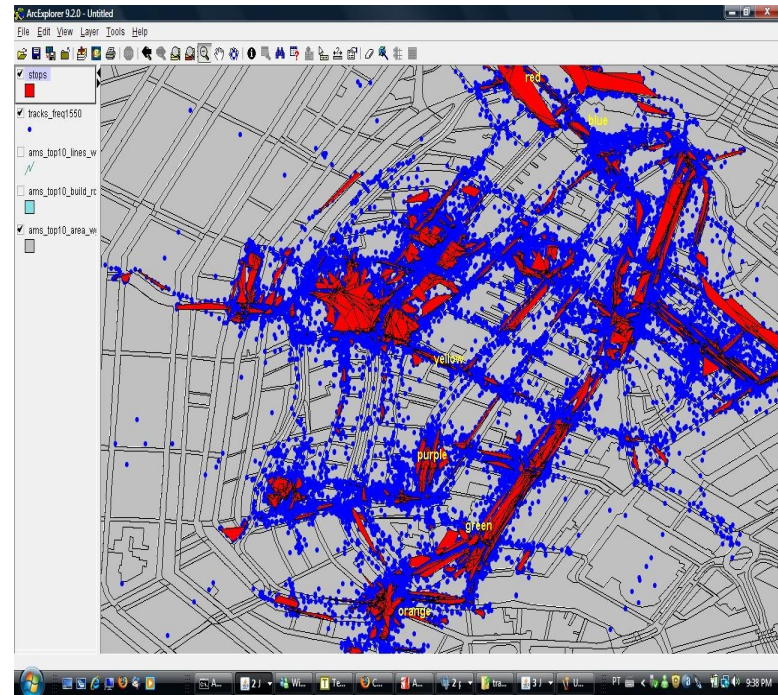
❖ **Spatial Slicing**



Flow field details for a horizontal slice at the surface. (EPA)

Some examples include

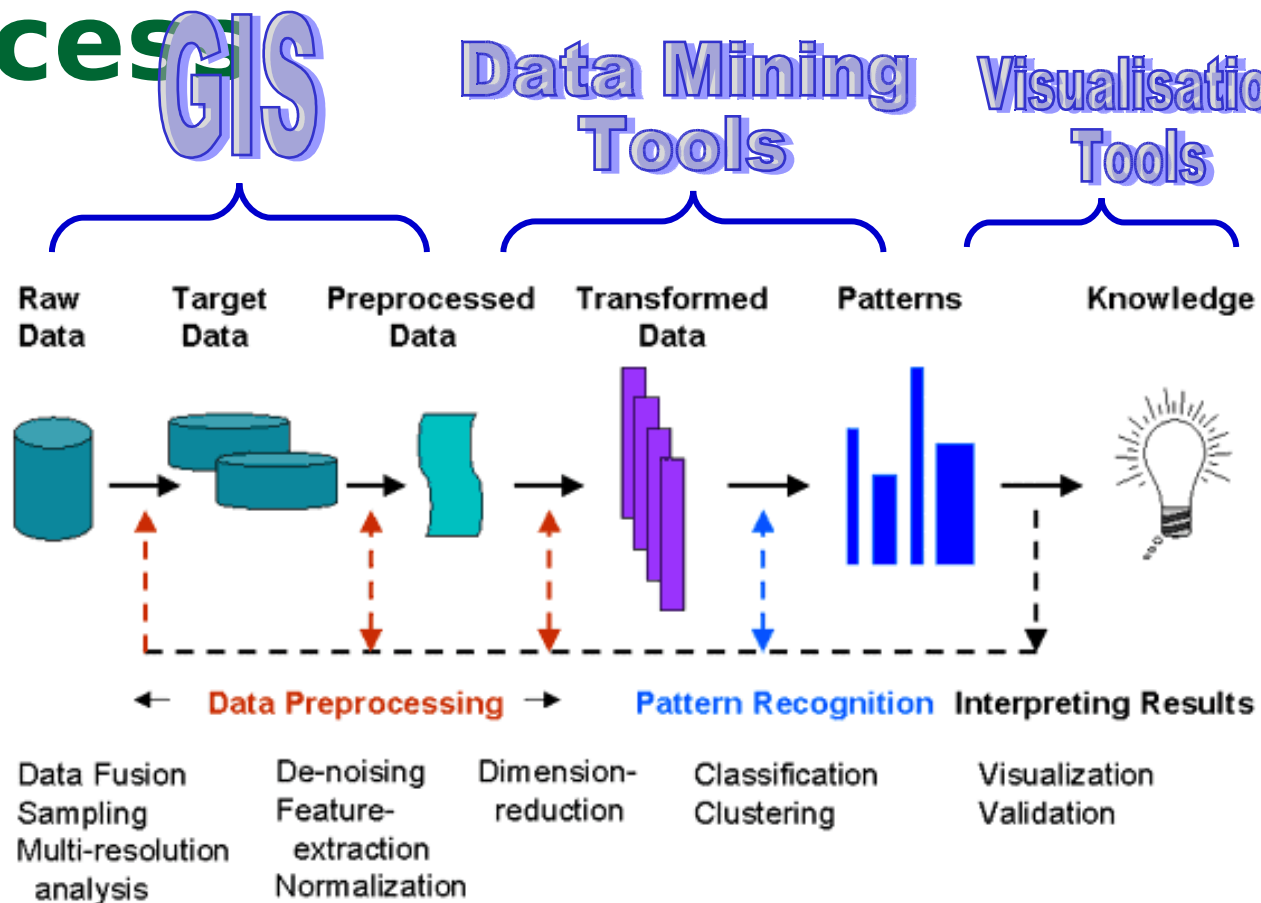
- ❖ Spatial Outliers and Discontinuities
- ❖ Spatial Slicing
- ❖ **Spatial clustering**



Vania Bogorny and Wachowicz, M. (2008). A Framework for Context-aware Trajectory Data Mining. In: "Data Mining for Business Applications", Springer (at press).

Geo Knowledge Discovery

Process



An iterative and interactive process

**What are the
scientific grand
challenges to which
GKD could
contribute?**



MOBILITY

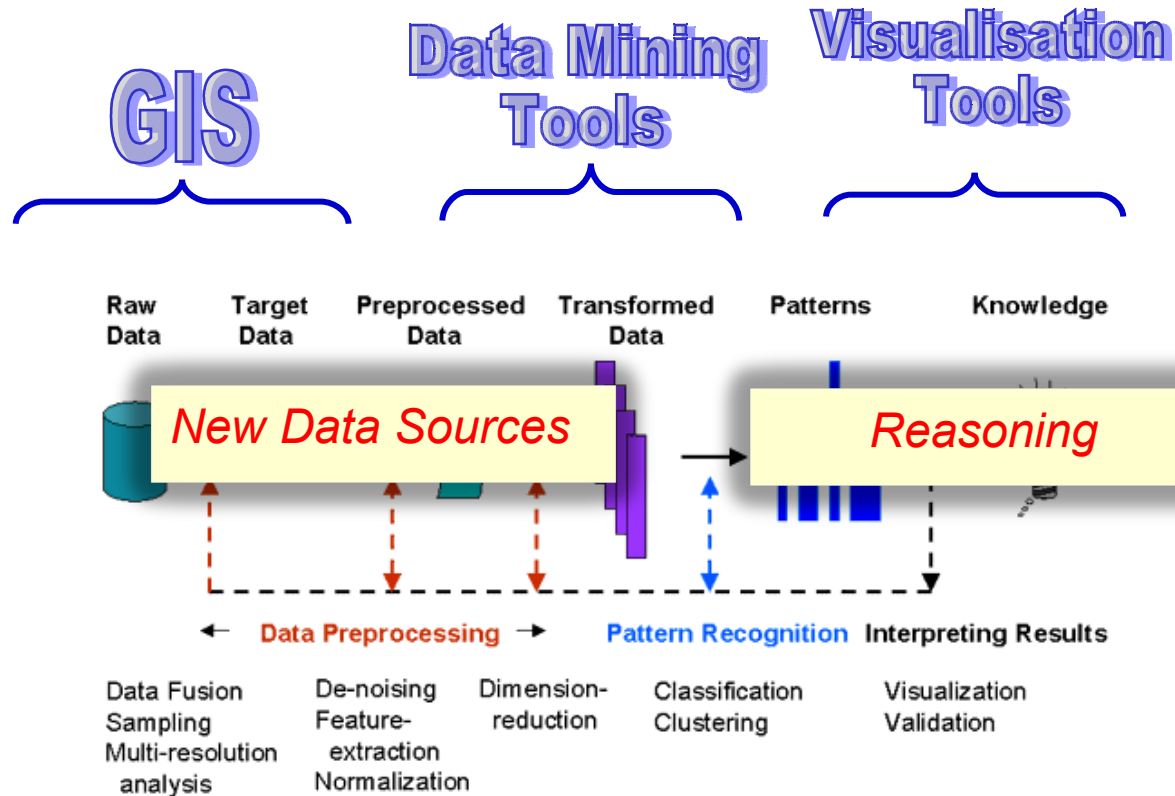
- ❖ **Is the dependence on cars changing?**
- ❖ **What are the accessibility patterns?**

New Perspectives

Driving Forces

- **Mobility and sensor services** are merging: from macro to micro Geography.
- **Knowledge** about the behaviour and patterns of sensors, terminals, users, and sessions.
- A vast range of benefits but also threats. **Privacy and control** are the most evident examples.

Thinking outside the box...



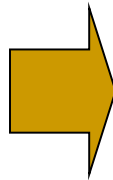
An iterative and interactive process

A concrete example: Reasoning on Mobility Patterns



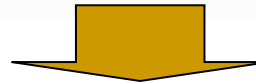
<http://www.geopkdd.eu>

GKDD Process

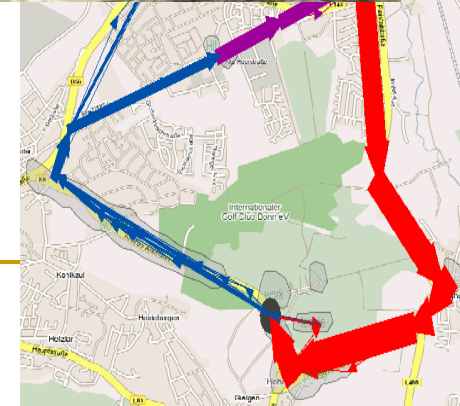
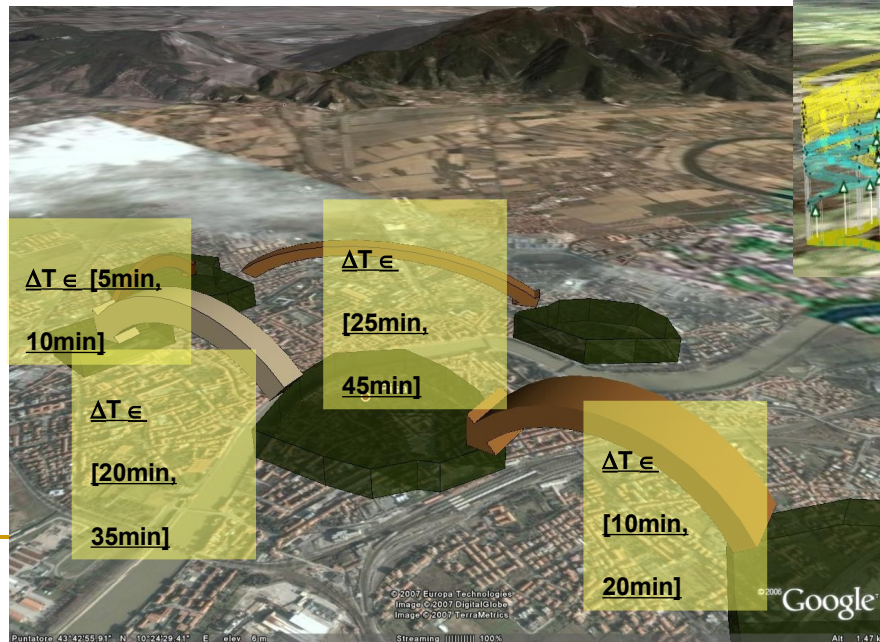
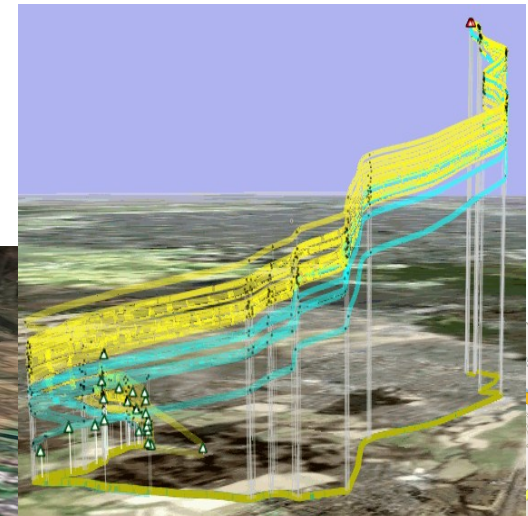


```

name|date|y|x
Prinzessin|08.20.1998|52.118|12.087
Prinzessin|08.23.1998|51.019|15.309
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|08.31.1998|38.715|32.165
Prinzessin|09.01.1998|37.195|35.255
Prinzessin|09.03.1998|32.979|36.021
Prinzessin|09.05.1998|28.513|33.437
Prinzessin|09.06.1998|23.961|32.937
Prinzessin|09.07.1998|19.418|33.446
Prinzessin|09.12.1998|15.823|34.094
Prinzessin|10.11.1998|14.685|32.848
Prinzessin|11.03.1998|11.510|32.591
Prinzessin|11.24.1998|13.888|35.667
Prinzessin|12.08.1998|12.562|34.777
Prinzessin|12.10.1998|9.124|35.644
...
    
```



Which are the tourist activities?



The need for semantics and reasoning

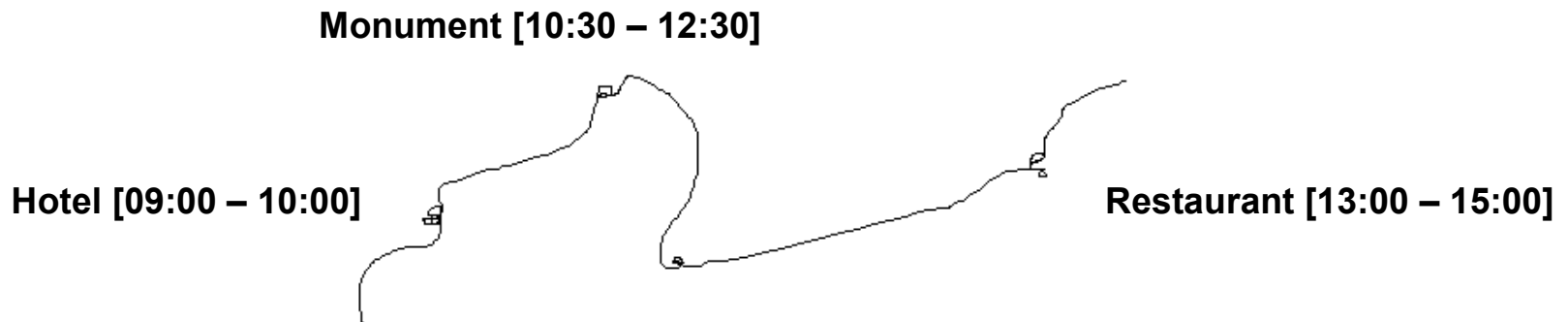
- may support the user in the **interpretation** of trajectories and patterns as people **behaviour or activity**.
 - may infer suspicious behaviour or tourist activity from a motion pattern.
-

Semantic enrichment process

Semantic Trajectory defined as sequence of Stops and Moves.

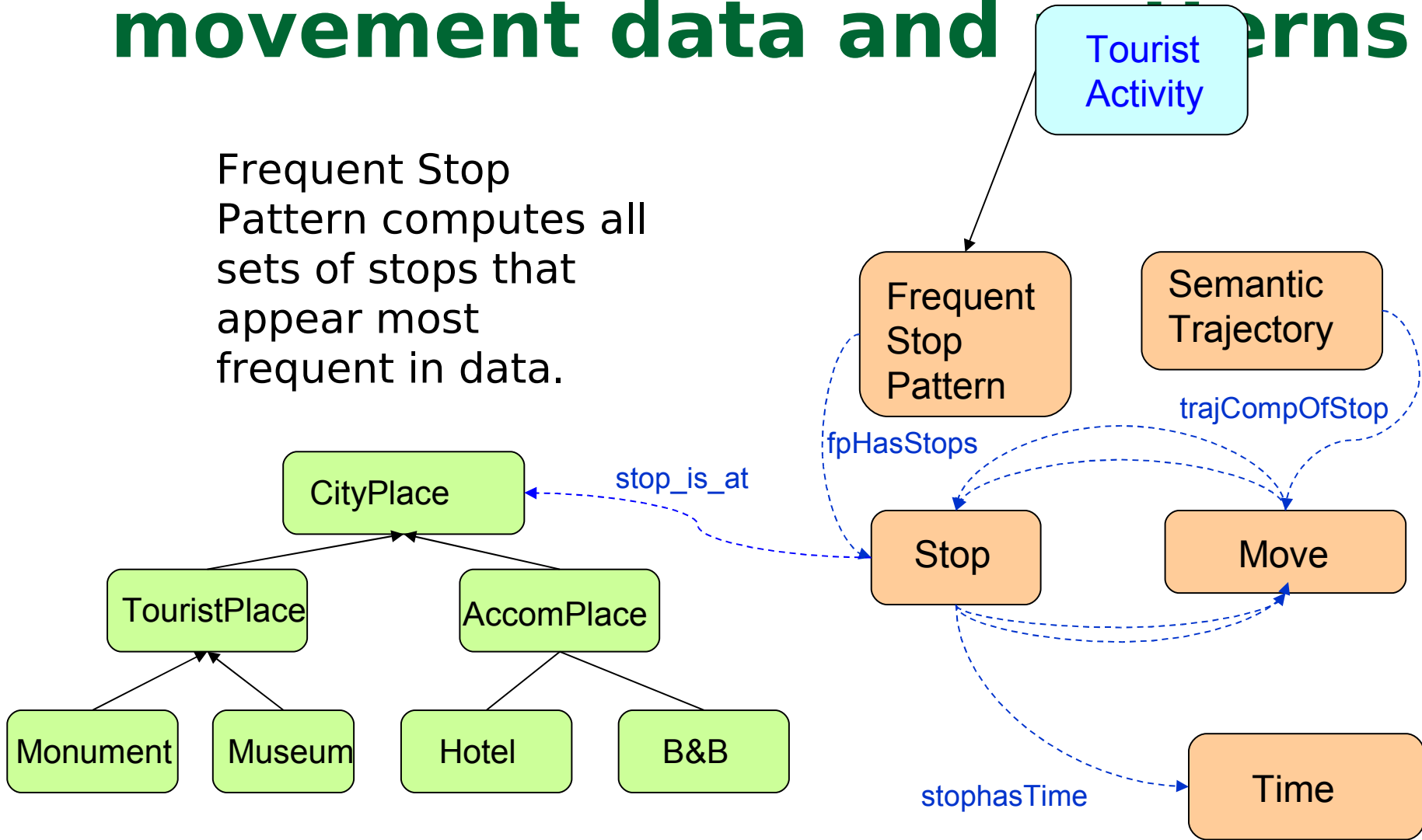
Stop is where the position of the object stays fixed

Move is the part of the trajectory where the position changes



Semantic representation of movement data and patterns

Frequent Stop Pattern computes all sets of stops that appear most frequent in data.



Reasoning on patterns

A **tourist activity** is represented by a **frequent pattern** that has some stops in Accomodation Places and some in Tourist Places.

The reasoning engine checks if an individual (a given frequent pattern) is an instance of a concept (tourist activity).

An axiom is a combination of logical operators that defines an implicit class

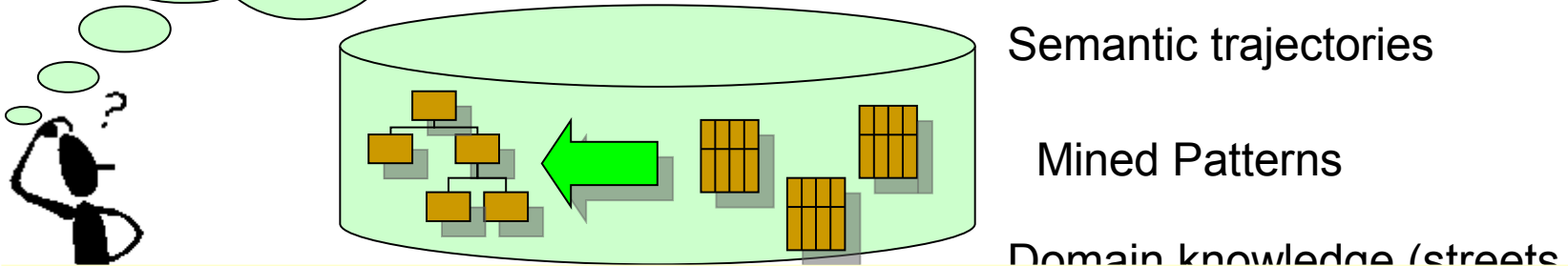
⊑ **TouristActivity**

⊑ (FpHasStop some (stop_is_at some AccomodationPlace) and FpHasStop some (stop_is_at some TouristPlace))

ATHENA

A prototype has been built to compute inferences over pattern extracted from semantic trajectories. The reasoning engine is Oracle 11g Semantic Technologies, with a subset of OWL called OWLPRIME.

Which are the
Tourist
Activities?



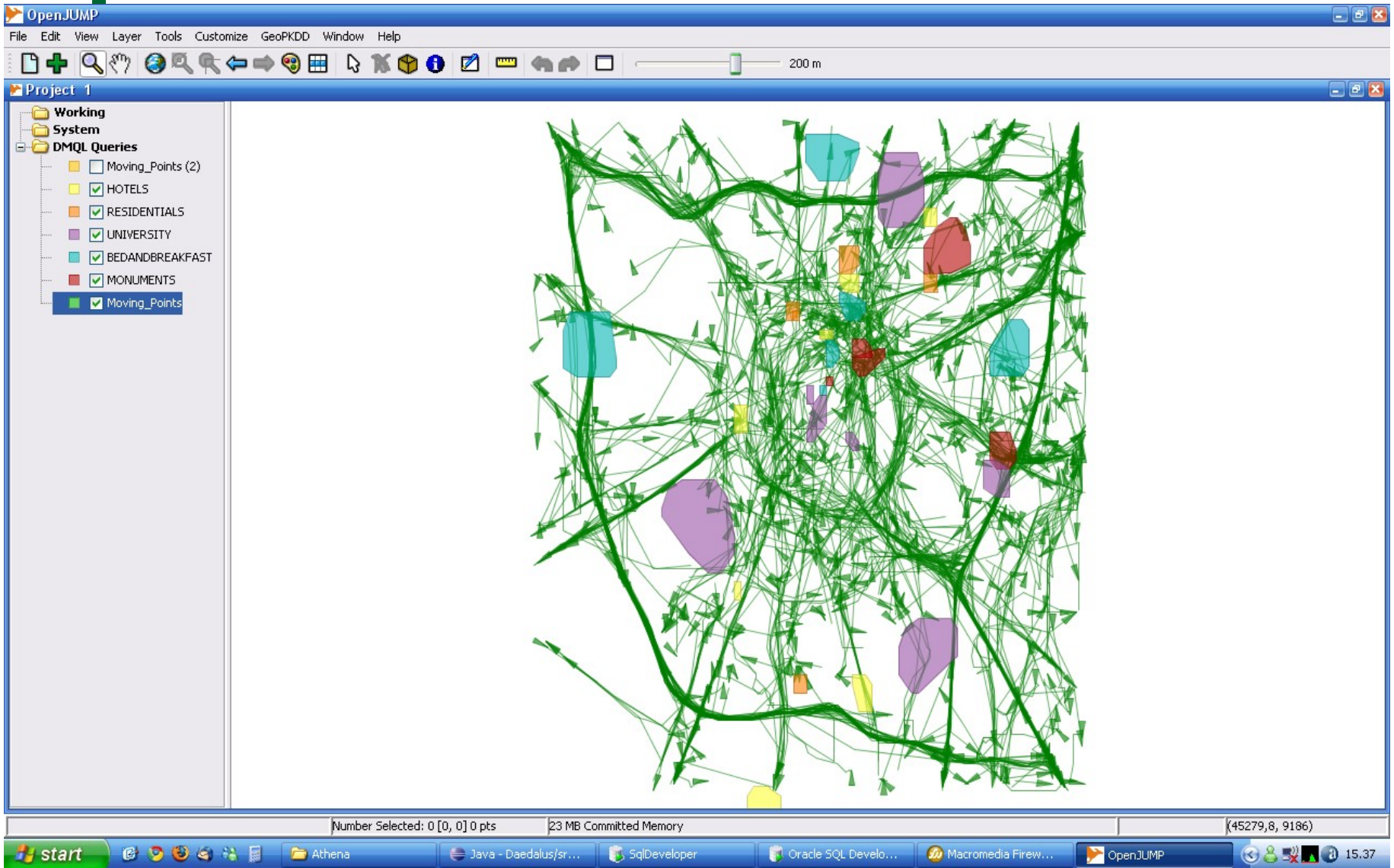
```
SELECT m FROM table(SEM_MATCH('(?m rdf:type  
:TouristActivity)',SEM_Models('modelgeopkdd'),SEM_  
rulebases('owlprime'),SEM_ALIASES(SEM_ALIAS(",'http://www.owl-  
ontologies.com/GeoPKDDOnto.owl#')),null));
```

M

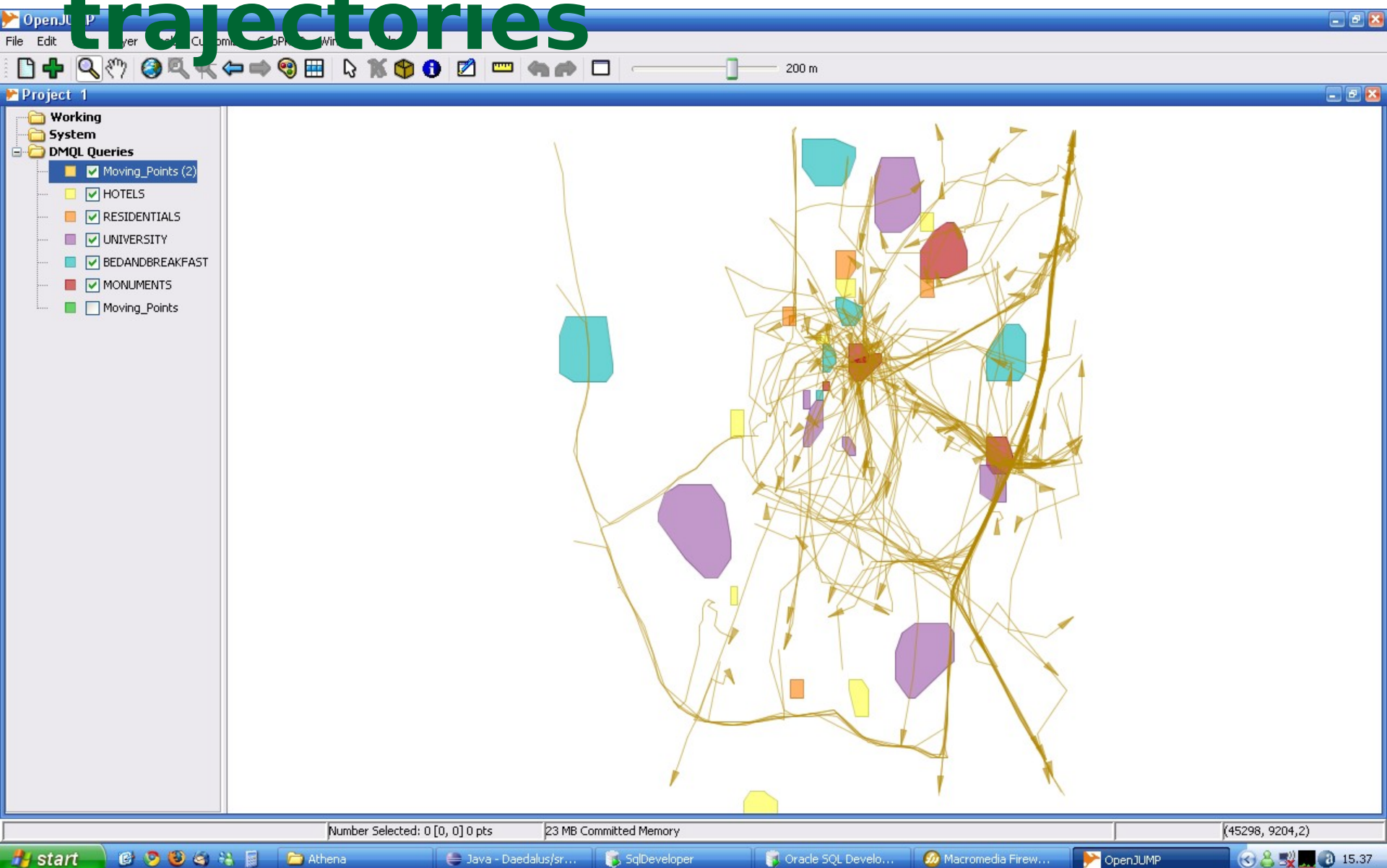
<http://www.owlontologies.com/GeoPKDDOnto.owl#FrequentStop3>

<http://www.owlontologies.com/GeoPKDDOnto.owl#FrequentStop6>

Athena: Trajectories and city



Athena: Tourist trajectories



Conclusions

-
- **New data sources** will provide a different pool of data and patterns that will require new geo data mining algorithms.
 - **Reasoning** will support GKDD process through interactive and explanatory inference tasks.
-