

## Spatial Data Infrastructures and Data Mining – An Introduction

Environmental Informatics and Industrial Ecology  
Leuphana University of Lüneburg  
September 10-12, 2008

Nguyen Xuan Thinh & Gotthard Meinel



## Content

- 1 **Introduction**
- 2 **Spatial Data Infrastructures – a short Overview**
- 3 **Spatial Data Mining**
  - 3.1 Three most popular data mining techniques  
Predictive modelling, Association analysis, Cluster analysis
  - 3.2 Components and specific features of spatial data mining
- 4 **Examples of application of spatial data mining at the Leibniz Institute IOER**



### 1. Introduction

- Spatial data = Items of information related to a location on the Earth (e.g. topography, place names, height data, land cover, hydrology, cadastre)
- Approximately 80 % of governmental data has a spatial basis
- Two major forces driving the development of spatial data
  - Growing need for government & businesses to improve their decision-making
  - Advent of cheap, powerful information and communication technology
- Growing need to share spatial data at all spatial levels and across all sectors
  - To avoid duplication of expenses
  - The more partners share data, the more the savings and the greater the efficiency



### 1. Introduction

- Spatial data infrastructures (SDIs) & data mining a key challenge of a modern society
- Access to spatial information & to reach information by data mining for better decision-making one key to a sustainable future
- This is not possible without SDIs and data mining technology



## 2. Spatial data infrastructures - overview


H. J. Streuff 2008

**What?** Fundamental spatial data, application related spatial data, geographically referenced data, meta data  
→ **Fundamental Dataset**

**Klenke et al.:** German Spatial Meta Data Infrastructure

**Bischof & Bauer-Messmer:** Semantic Enhancement of Environmental Metadata

**Components of Spatial Data Infrastructure**



## 2. Spatial data infrastructures - overview

H. J. Streuff 2008


**What?** Fundamental spatial data, application related spatial data, geographically referenced data, meta data  
→ **Fundamental Dataset**

**Wherewith?** Services (search, presentation, download, transformation)  
→ **Geoportals, Web Mapping Services**

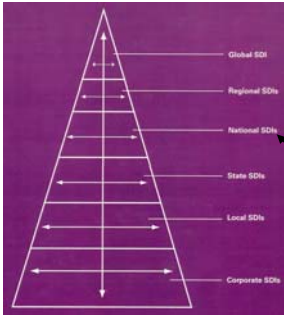
**How?** Technical and semantic standards  
Legal regulations  
→ **Standards: ISO, OGC**

**From where?** Administrative structures, networks of experts and people  
→ **Access Network**

**For whom?** **Business, administration, science, citizen**




## 2. Spatial data infrastructures - overview



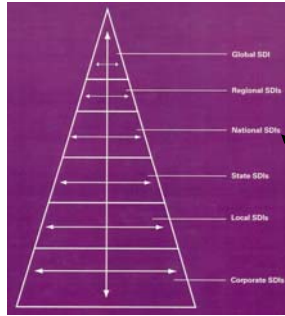
Policy at a global level has only direct impact on regional and national SDIs

Major role in forming fundamental datasets created from the lower levels and in building the other levels of SDIs

Relationships within and between different levels




## 2. Spatial data infrastructures - overview



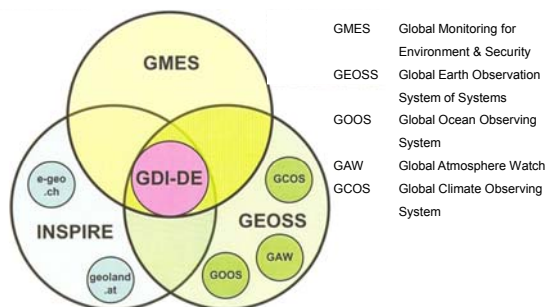
INSPIRE  
Infrastructure for Spatial Information in Europe

GDI-DE      German SDI  
Geoland.at      SDI in Austria  
E-geo.ch      Swiss SDI  
For example

Relationships within and between different levels (Williamson et al. 2003)



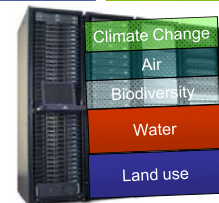
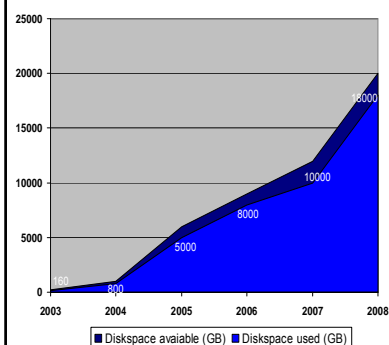
## 2. Spatial data infrastructures - overview



Spatial Data Infrastructures and Observing Systems (Streuff 2008)



## Development of the amount of the EEA data (Jensen 2007)



Page rank 7 on Google  
6 Million user sessions / year

28 Terra byte on data  
80 Servers  
20Mbit internet line

80 European spatial datasets  
415 Statistical datasets  
>1.000.000 landscape photos



## 3. Data Mining

- The current development of SDIs
- The continuous increase of the amount of digital spatial data
- The complexity of spatial data types, spatial relationships, and spatial autocorrelation



Spatial Data Mining getting more and more important



## 3. Data Mining

- **Predictive modelling**  
Goal is to develop classification models, capable of predicting the value of target variable as a function of other variables (explanatory variables)  
For example: predicting consumer spatial behaviors
- **Association analysis = pattern discovery**  
Goal is precisely to discover patterns that describe strong correlations among features in the data or associations among features that occur frequently in the data
- **Cluster analysis**  
Goal is to partition a data set into groups of closely related data in such a way that the observations belonging to the same group, are similar to each other, while the observations belonging to different clusters are not .  
→ **Behnisch**: Spatial similarities in urbanisation and regional diversity



## Spatial Data Mining

- **The process of discovering**
  - Interesting, potentially useful, non-trivial patterns
  - From large spatial datasets
- **Components of Data Mining**
  - Input: table with many columns
  - Statistical Foundation
  - Computational process: algorithms
  - Output: patterns and interest measures (e. g. outliers, predictive models, associations, clusters)

## What's NOT Spatial Data Mining

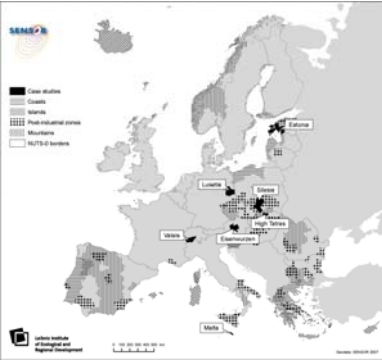
- **Simple querying of spatial data**
  - Find neighbors of Lüneburg given names and boundaries of all municipalities
  - Find shortest path from Lüneburg to Berlin in a map
- **Testing a hypothesis via primary data analysis**
  - Search space is not large
- **Uninteresting or obvious patterns in spatial data**
  - Common knowledge: Nearby places have similar rainfall
- **Mining of non-spatial data**
  - Pastry sales and beer sales are correlated in evening

## 4. Examples of applications of spatial data mining at the Leibniz Institute of Ecological and Regional Development (IOER)



Cluster analyses for the 116 German cities and subsequent analysis of the structural characteristics of clusters

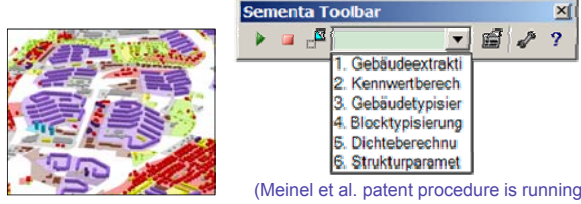
→ extrapolate benchmark values and recommendations for sustainable urban development



Identify generalisable and specific sustainability characteristics in seven sensitive European areas. EU integrated project SENSOR

LEIBNIZ Institute of Ecological and Regional Development

### SettlementAnalyzer - SEMENTA®

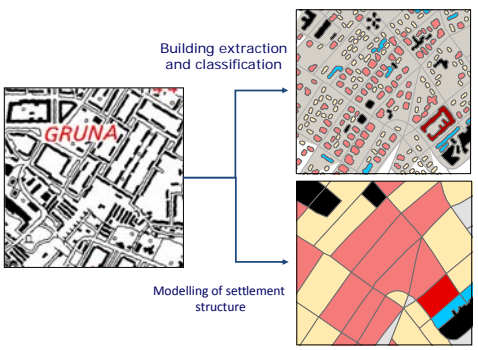


(Meinel et al. patent procedure is running)

- Classification of each building and urban block
- Calculation of building- and block-oriented parameters (e.g. population, structure density, number of apartments)

LEIBNIZ Institute of Ecological and Regional Development

### Classification



Building extraction and classification

Modelling of settlement structure

**Legend Building Type**

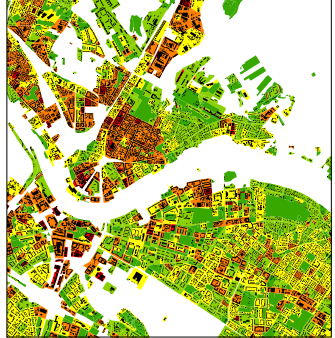
- 17 - Multi-family houses, traditional in closed
- 18 - Multi-family houses, free-standing
- 19 - Multi-family houses, traditional in rows
- 31 - Detached and semi-detached houses
- 41/42 - Non-residential

**Legend Urban Structure Type**

- 11 - Heavy built-up blocks
- 12 - Detached elements
- 21 - Terraced housing / linear development
- 31 - Clusters Detached single family housing
- 41/42 - Non-residential
- Non-built-up areas

LEIBNIZ Institute of Ecological and Regional Development

### Results for Dresden - Footprint Density



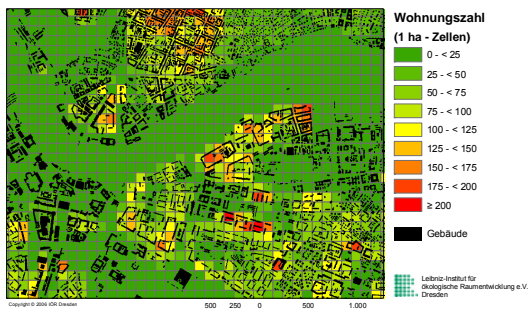
**Grundflächendichte**

- ≤ 0.10
- > 0.10 - ≤ 0.20
- > 0.20 - ≤ 0.30
- > 0.30 - ≤ 0.50
- > 0.50 - ≤ 1.00

500 250 0 500 1.000 1.500 Meter

LEIBNIZ Institute of Ecological and Regional Development

## Result for Dresden – Flat Density (Dresden Center - 100m-Raster)



## Conclusion

- Spatial Data Infrastructures and data mining a key challenge of sustainable future
- SDIs as well as Spatial Data Mining should become a topic of the annual conference EnviroInfo

## Contact

Nguyen Xuan Thinh & Gotthard Meinel  
Leibniz Institute of Ecological and Regional Development

[ng.thinh@ioer.de](mailto:ng.thinh@ioer.de), [g.meinel@ioer.de](mailto:g.meinel@ioer.de)

Thank you very much for your attention!