# Investigating a bottom-up approach for extracting domain ontologies from urban databases

Christophe Chaidron[1], Roland Billen[1] & Jacques Teller[2]

[1] University of Liège (Belgium)
[2] CERMA (France)

# Outline

- Forewords
- Urban Spatial Databases (SDBs)
- Spatial objects and relationships
- Bottom-up approach for extracting domain ontologies
- DB's reengineering case study
- Role of spatial relationships
- Conclusions

# Forewords

- This study is a ***on-going reflective analysis*** based on DB reengineering works
  - Ontologies in Spatial DBs
  - Ontologies extraction issues
  - Ontologies have not been formally extracted
  - Key role of spatial relationships

# Urban SDBs (1)

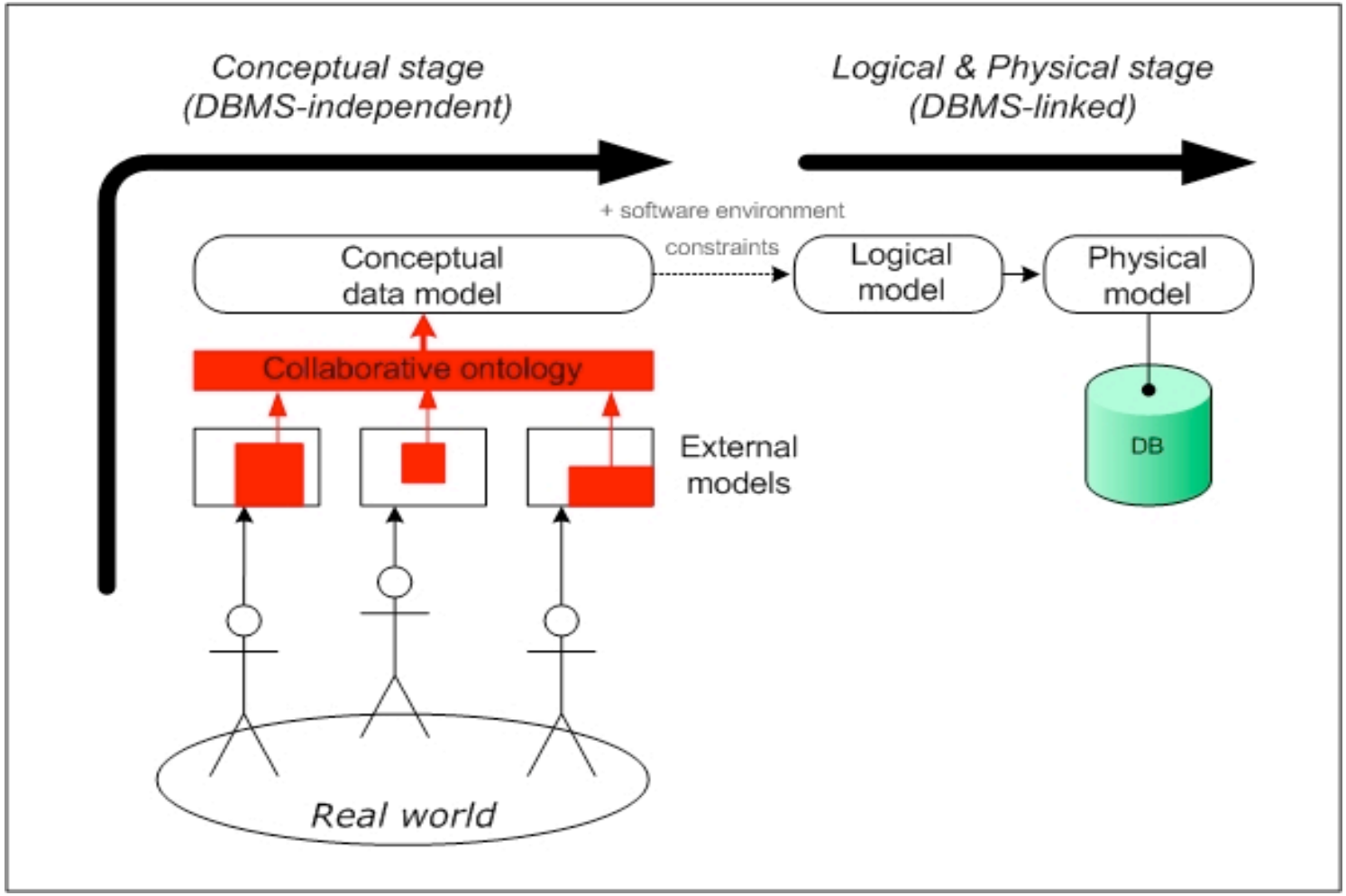
- Geographical (spatial) information about urban areas is more and more stored in GIS or in SDB
  - storing spatial data (geographical entities) described by
    - *attributes* (alphanumerical data or images, sounds, binary attributes…)
    - associated to some *geometric information* (position, shape, topological relationships, etc…)

# Urban SDBs (2)

*IDEALLY …*

- **Design** of such systems or DBs requires creation of standardised documentations
  - Feature data catalogues
  - Conceptual data models, …
- It forces the community of data producers, developers or even future users to (re-)think about the “basic geographic entities of their world” regardless of any database system.
- It corresponds somehow to domain ontology design.

# Urban SDBs (3)


Conceptual stage
(DBMS-independent)
Logical & Physical stage
(DBMS-linked)
+ software environment
Conceptual
data model
constraints
Logical
model
Physical
model
Collaborative ontology
External
models
DB
Real world

# Urban SDBs (4)

- It is quite common to have no record of ontologies in current urban information systems
- Urban GIS and SDB are therefore a huge source of “hidden” urban “domain” ontologies
- If they have not been formally recorded, it is necessary to extract them

  → *Bottom-up approach*
- Why extracting them? (see Jacques’s COST C21 objectives)
- How can we extract them?
- Extraction processes depend on type (and quality) of available documentation

# Spatial Objects and Relationships (1)

- Types of objects
  - Geo-object representation
  - Vector / raster
  - 0D, 1D, 2D, 3D, 4D
  - Scale, granularity, …
- Spatial relationships
  - Qualitative vs Quantitative
  - Topological relationships (“connectivity” concepts widely used in GIS)
  - And others …

# Spatial Objects and Relationships (2)

- Qualitative vs Quantitative spatial relationships
  - saying that the city of Liège is {*disjoint of, not far from, east of*} the city of Brussels, is a qualitative statement,
  - when saying that the city of Liège is *at 95 km* from the city of Brussels is a quantitative statement

# Spatial Objects and Relationships (3)

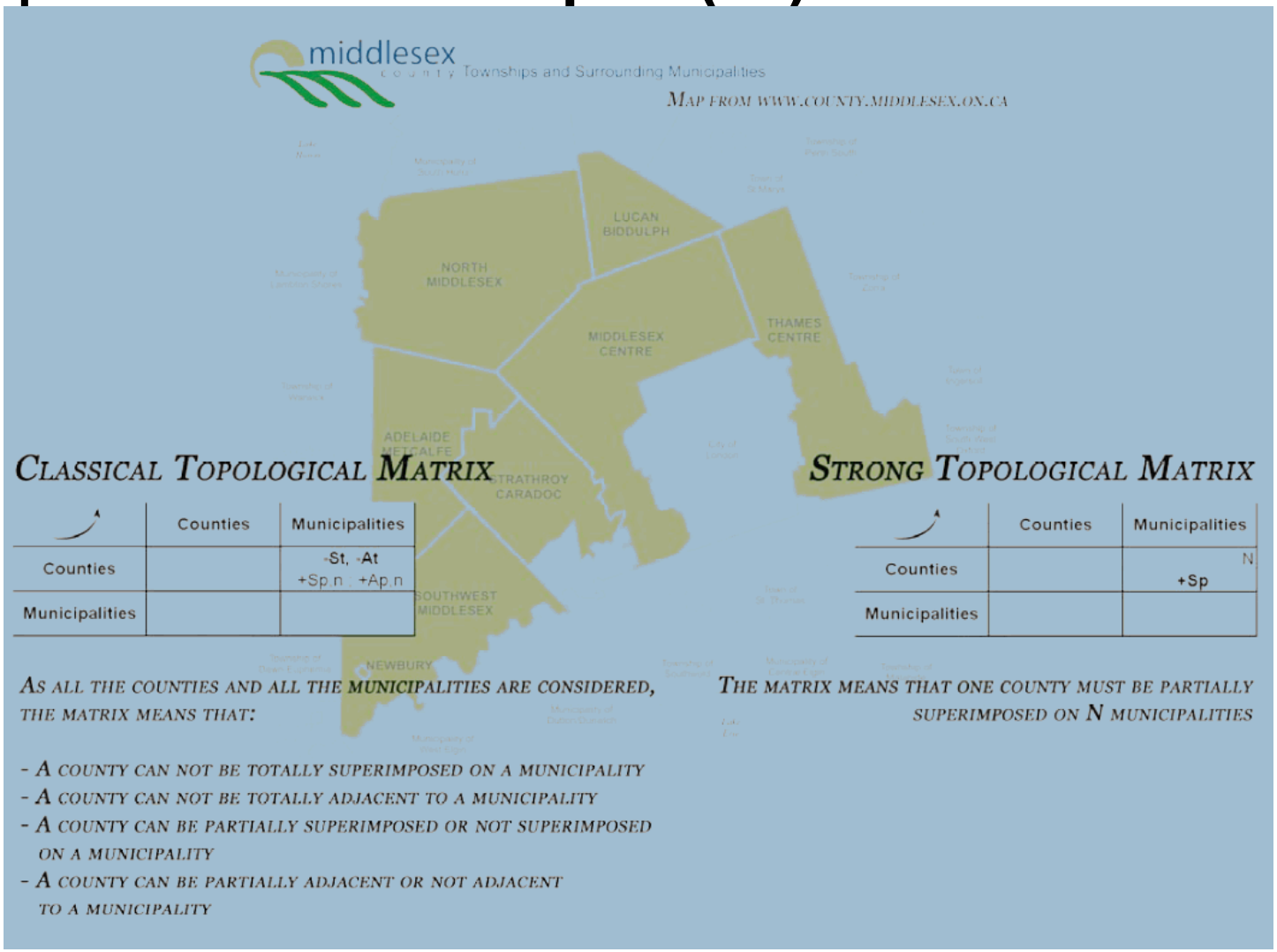
**Topological relationships**

- Such as “disjoint”, “overlap”, “included”, …
- Formally defined (Egenhofer, Clementini, …)
  → Open GIS standards

*Another way to “consider” them*

- CONGOO formalism considers two relations (Superimposition (S), Neighbourhood (N)) with three application levels: total (t), partial (p), non existent (ne).
- For example, saying that Liège and Brussels are disjoint could be stated as: Liège S ne N ne Brussels
- Interesting concept of *Topological Matrices*

# Spatial Objects and Relationships (4)



# How can we extract ontologies from DBs?

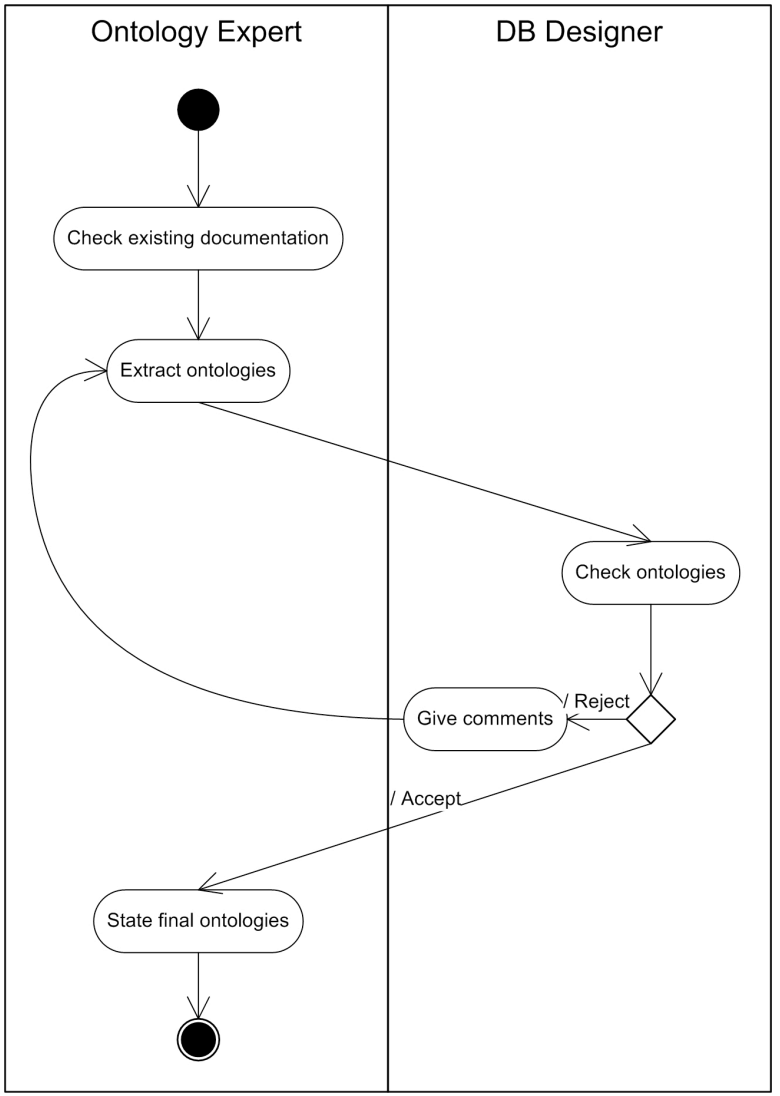
- Case 1: Ontologies have been recorded …

Otherwise, using a Bottom-up approach

- Case 2: From existing documentation
- Case 3: From a DB reengineering
- Case 4: From a populated DB only?

# A bottom-up approach

- Available documentation
  - Feature catalogue
  - CDMs
  - DB itself …
- Extraction
  - (semi-)automatically (feature catalogue, CDMs, etc.)
  - Manually (others documents, brain storming, etc.)
- Check - validation
  - With “experts” …?



# From Feature catalogues

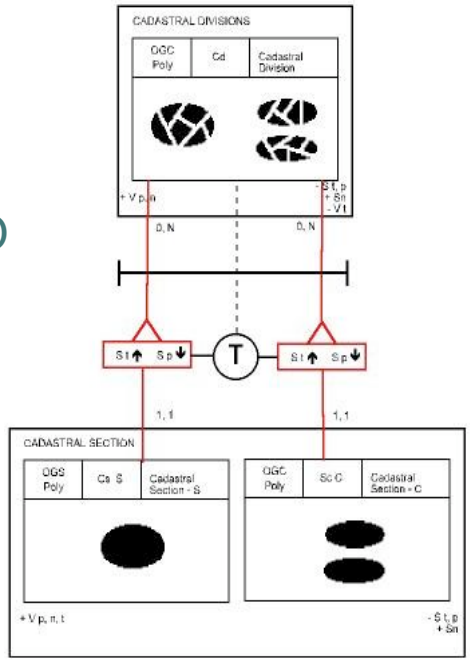
Selection of relevant information

# From CDMs

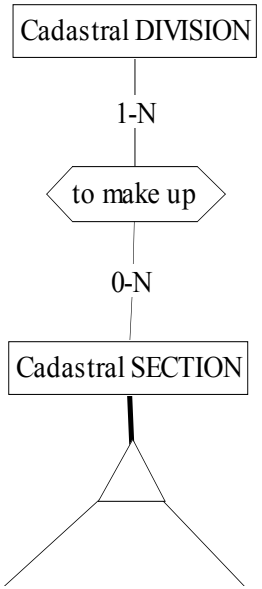
- Conceptual Data Models
- Created using (geo)-formalisms
  - ER, Geo-UML, MADS, …, CONGOO
- … could be « simplified » into semantic nets

# Semantic net from CDM

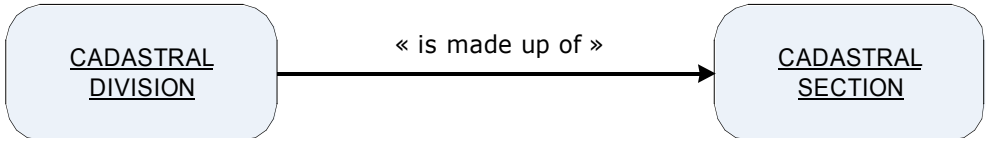
CONGOO




ER

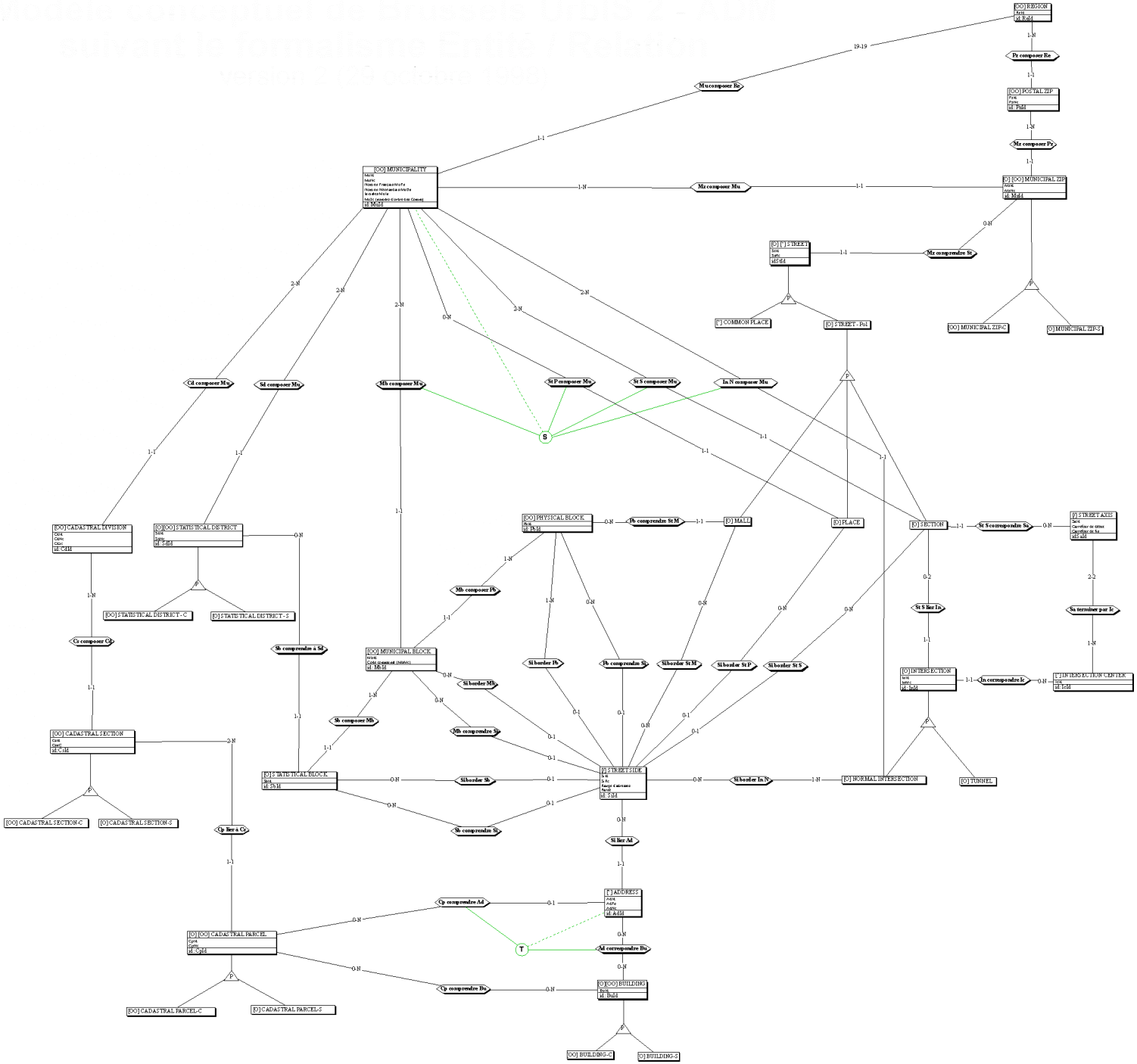



Semantic net



# Modèle conceptuel de Bruxelles-Urbis 2 - afin suivant le formalisme Entité / Relation

version 2.025 octobre 1999



# DB reengineering case study

- Context
  - Brussels regional reference geographical DB
  - Supporting DB reengineering processes
  - DB designed with very little formalisation
- Aims
  - Create a feature catalogue and CDMs
  - Further work on data quality
- (re)-definition of domain ontologies of the original database
- Not an explicit extraction of ontologies

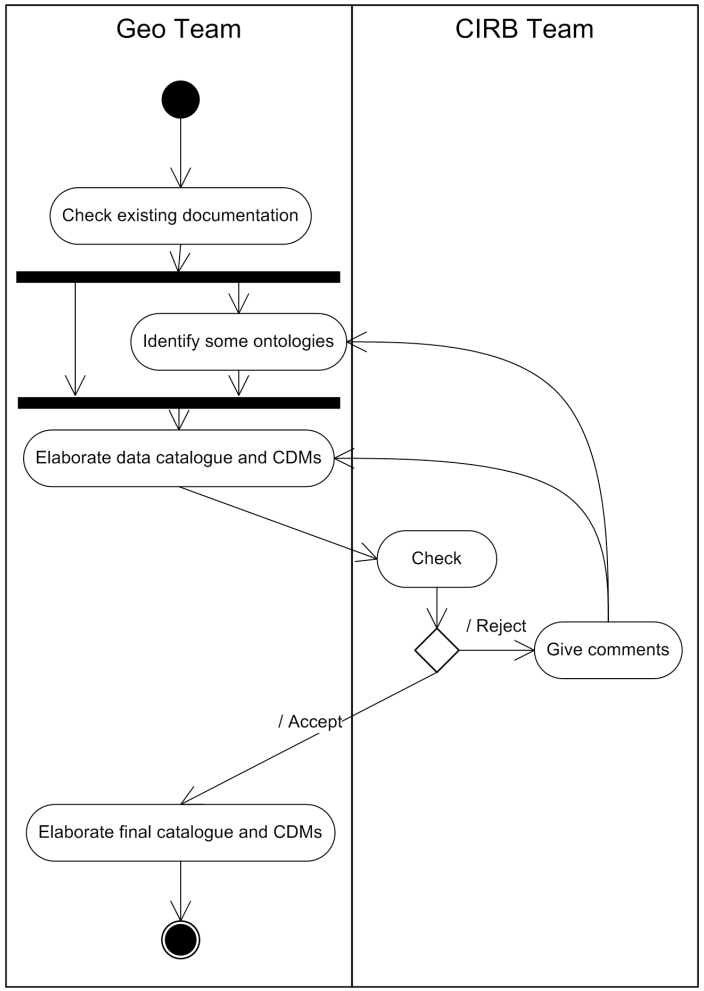
# Methodology

**Available documentation**

- Relational schemes
- Data list (different from a catalogue structure)
- Data acquisition specifications (for photogrammetric and land surveying)
- The DB's objects themselves…

**Outcomes**

- Feature catalogue
- Topological matrices
- CDMs
- Recommendations



# Role of spatial relationships

- Existing definition 1: *The « house » is the building extract out of the topographical survey*

  not satisfactory. Our own understanding of the objects leads us to the following definition (whose validity was checked against other DB's documentation):
- Definition 2: *The « house » corresponds to footprint of a building (including its annexes)*

  This definition appeared to correspond to the designer "ontologies". However, did not match with topological matrix
- Definition 3: *The « house » corresponds to building's footprint, including annexes and all other construction such as church, chapel, monument, school, fountain, greenhouse, bus stop, etc.*
- This definition is quite odd and not satisfactory conceptually. However, it corresponded to the reality of the DB and had been included in the feature catalogue.

# Objects' Clustering

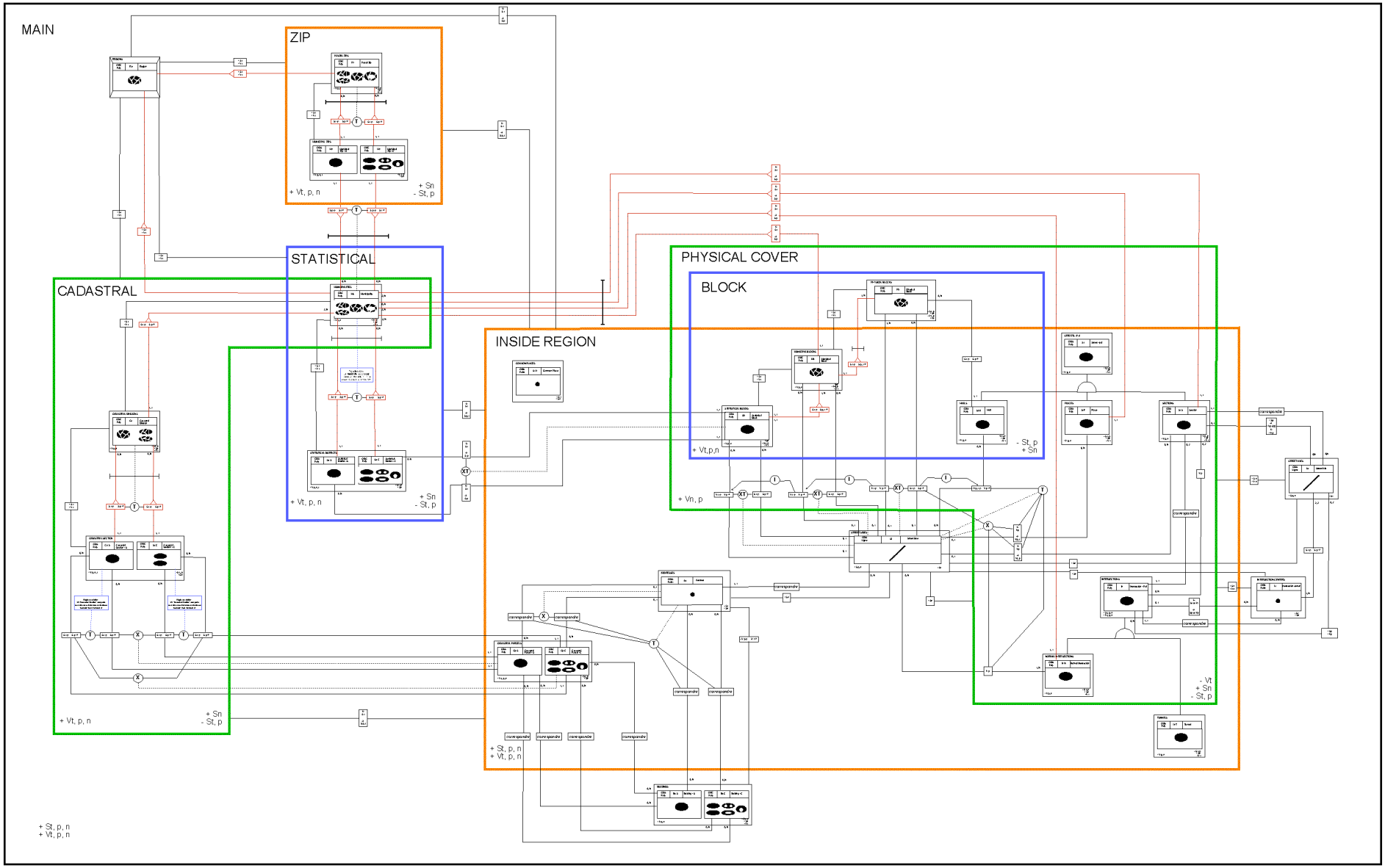
- Using Topological Matrix

| | Object A | Object B | Object C | Object D | Object E |
|---|---|---|---|---|---|
| Object A | | | | | |
| Object B | | | | | |
| Object C | | | | | |
| Object D | | | | | |
| Object E | | | | | |

- Based on topological relationships similarity

| | Object A | Object D | Object E | Object C | Object B |
|---|---|---|---|---|---|
| Object A | | | | | |
| Object D | | | | | |
| Object E | | | | | |
| Object C | | | | | |
| Object B | | | | | |

- Topologically coherent layers


MAIN
ZIP
+ Vt, p, n
+ Sn
- St, p
STATISTICAL
+ Vt, p, n
+ Sn
- St, p
CADASTRAL
+ Vt, p, n
+ Sn
- St, p
PHYSICAL COVER
BLOCK
+ Vt,p,n
- St, p
+ Sn
INSIDE REGION
+ Vn, p
- Vt
+ Sn
- St, p
+ St, p, n
+ Vt, p, n
+ St, p, n
+ Vt, p, n

# … a populated DB only

- Could build some king of topological matrices
- This could provide a first "understanding" of the spatial entities

  … what are the allowed or forbidden spatial relationships between objects

# Conclusions

- Do not underestimate ontologies design in DB design
- Ontologies could be extracted from DBs
  - from data catalogues, CDMs, …

    if they exist
- Spatial relationships play a key (crucial?) role in ontologies retrieval