UNDERSTANDING URBAN DYNAMICS: THE USE OF VECTOR TOPOGRAPHIC DATABASES AND THE CREATION OF SPATIO-TEMPORAL DATABASES

Julien Perret, Anne Ruas
Laboratoire COGIT, Institut Géographique National
73, avenue de Paris, 94165 Saint-Mandé cedex, France
Julien.Perret@ign.fr Anne.Ruas@ign.fr

Annabelle Boffet Mas
Laboratoire CEDETE, Université d'Orléans
10 rue de tours, BP 46527, 45065 Orléans cedex 2, France
Annabelle.Mas@univ-orleans.fr

Abstract

Understanding urban dynamics is an increasingly important challenge that can be answered in several ways. For many years, an important body of research has been tackling this problem in the literature of regional science, urban and economic geography, geomatics and related fields. But, whatever the field of research, in order to better understand urban dynamics, one needs to study the actual changes geographic objects undergo over time. Studying such changes requires the constitution of spatio-temporal databases containing data for each step of the considered evolutions. Unfortunately, building such spatio-temporal databases from existing topographic databases is a difficult process that requires tracking topographic objects in consecutive databases and creating links between them. The evolution of topographic objects which evolution has been tracked can be studied and then simulated. The observed changes the geographic objects undergo help up understand each topographic object's dynamics, their particularities on specific geographical areas as well as their evolutions in different time periods. Simulations can then be used to better understand the specific evolutions taking place on a given territory. Eventually, such simulations might be able to help the planning and policy making processes. The methodology presented in this paper includes the construction of spatio-temporal databases, the capture of historical topographic data from existing topographic maps, and a rich hierarchy creation process. Finally, the bases of our vector-based simulation platform are given.

1. Introduction: the GeOpenSim project

This paper deals with the presentation of the GeOpenSim research project which aims at making an open source system to study the evolution of urban space. Compared with other approaches, we aim at building evolution patterns based on the analysis of existing vector topographic data and to use these evolution functions to simulate town extension based on the agent paradigm. In our approach, the simulation aims at

validating identified knowledge on urban extension.
The project, funded by the French Research Agency (ANR), is a collaboration between four French research laboratories: 2 Geography laboratories from Strasbourg and Orleans, a Computer Science laboratory from Strasbourg and the COGIT laboratory of the French National Mapping Agency who leads the project. All developments are performed on GeOxygene, an open source GIS developed at the COGIT laboratory.

An overview of the GeOpenSim modules:
The GeOpenSim system should first allow for the analysis of existing topographic data in order to build realistic evolution functions. Indeed, evolution functions are based on the analysis of existing topographic databases at different temporalities and provide rules to simulate the evolution of urban areas. By comparing a large set of data, we wish to create rules such as: *If a geographical space has such geometrical properties, then X % of this kind of objects might change in such a way during this time period*. An instance of such a rule could be: *in small and low density detached house areas, 15% of new houses are created between 1980 and 2000.*
A simulation is characterised by a geographical area and a time period. If the starting point of the simulation lies in the past, the result of the simulation can therefore be compared with reality. Otherwise, the result of the simulation constitutes a new representation of the geographical space that should be as realistic as possible.
The construction of such a system relies on a specific data schema adapted for urban simulation and a set of six modules. The GeOpenSim data schema allows the representation of temporal topographic databases and their simulation. We first distinguish the representation of one geographical entity as one geographical agent that has a life span, and the representation of topographic objects which are snapshots of agents at a specific time. Thus, each geographical entity is represented by one geographical agent and several topographic objects (see section 3). This schema also includes meso objects such as urban blocks that are built by means of specific functions (module 2 below).

The GeOpenSim modules:
Module 1, Loading data: This module loads a classical topographic database into the GeOpenSim data schema (see section 3).
Module 2, Spatial analysis and data enrichment: This module creates meso objects such as urban areas, urban blocks, building clusters or street sub-network from the initial data. It also allows characterising micro and meso objects by means of spatial analytical tools. For example, urban blocks are characterised by their density, their size, their shape and the nature of buildings. The characterisation is used to associate the appropriate evolution function to each urban situation. It lays on classical analytical tools as well as supervised machine learning algorithms for urban block classification.
Module 3, Simulation: This module allows the simulation of urban extension at the specific time and for a specific duration. It uses the temporal data schema, some evolution rules built by module 5 and populating functions that modify the space. A populating function, for example, adds a new house within an urban block composed of

houses, while respecting the spatial organization of existing houses.
Module 4, Building spatio-temporal databases (STDB): In order to build the evolution functions, we analyse real topographic databases at different times. As such databases rarely exist, we propose methods to build them from current databases and other geographical sources such as photographies and maps.
Module 5, Building evolution functions: As soon as STDB exist, a module analyses spatial changes such as house densification or street extension. Machine learning algorithms are used to generalize examples into rules.
Module 6, State Evaluation: The aim of simulation is to study (and therefore improve) our knowledge on urban temporal evolution. It is thus necessary to develop a module that allows state comparisons: either to compare a simulation with real data (when the simulation is done on the past) or to compare the results of two different simulations with two rule sets.
Section 2 introduces our proposal to built STDB from topographic databases (module 4). Section 3 presents our agent based approach for simulation (modules 2 and 3). Conclusions and future work are discussed in section 4. The evolution functions, the state evaluation modules and the populating functions will be presented in an upcoming paper.

2. Creating Spatio-temporal databases from existing topographic databases

This section describes a methodology to create spatio-temporal databases (STDB) databases at the scale of the city. Most of the numerous researches lead on this subject (see http://www.fas.harvard.edu/~chgis/work/design/) aim at describing the evolution of space through time by analysing the evolution of statistical information on administrative units and/or the geometric modifications of the limits or shape of such administrative units.
In our project, we aim at describing the geometric evolutions of the built environment by analysing the modifications of each geographic object (at different scales) composing the studied area. No statistical information on human and society is used, but topographical information (databases and maps) produced by the French national mapping agency (IGN). Note that information from several other sources was not used in order to simplify the process.

Existing maps and data

Among other products, the IGN produces the BD Topo®, a vector geodatabase with metric precision adapted to the scale of 1:5 000 to 1:25 000. This database is structured into thematic categories: *Communications axes (roads, paths), Railway, Energy transport (electric network), Hydrography (rivers, lakes), Buildings, Relief (contour lines), Administrative limits, Vegetation*. Since 1996, the IGN 1:25 000 paper maps are derived from the BD Topo®. Before that, the best geometric precision on IGN products could be found on 1:25 000 paper maps.
One of the two test areas for the GeOpenSim project is centred on the city of Orleans, whose topographic database was updated in 2007. Thus, historical information is

collected on old maps. For the city of Orleans, maps established in 1957, 1978, 1989 and 1999 are selected.

Creation of spatio-temporal databases

This section shows how to derive spatio-temporal databases from old existing maps and a topographic database. The process has to be simple and fast to reproduce since the creation of STDB is not the main objective of the GeOpenSim project. The methodology is based on the most accurate information available: the last produced topographic database (established in 2007 in our case). Then, STDB are created using a manual "down-dating" of the most recent topographic database: a copy of the 2007 database is modified ("down-dated") using the 1999 map by removing, creating or modifying topographic objects (buildings and section of roads in a first time). Each STDB is created sequentially from the previously down-dated database, enabling a step by step handling of each object's evolutions.

Another possibility for this process is to create each STDB from the 2007 topographic database. Such a solution speeds up the process (each STDB can be produced simultaneously) and avoids the repetition of mistakes. Nevertheless, it also requires a final correction step in order to match analogous objects from different databases.

Tracking objects in time

In order to track the evolution of each object of the city, the creation of a links between objects in each time version is required. In the case of a single database that describes the type of evolution and the succession of objects in time such as (Mas, 2008), links are managed by a virtual object to which several geometries and semantic statutes are attached. On the contrary, the presented approach is based on the multiplicity of STDB in order to grow up rapidly and easily, no link between objects exists by default: it must be created.

The proposed link is based on the identifier (*idGeo*) given to the object in the 2007 database. From an agent model point of view, this *idGeo* is unique for each agent object during its life span: a building created in 1980 and still existing in 2007 will have the same *idGeo* in all subsequent databases (1987, 1999 and 2007).

Whereas Hornsby and Egenhofer (1997) present a complete classification of change over space and time in GIS, six cases are managed in our case. A special attention to the nomenclature of the object's life and the DB creation process is required. Between 1999 and 2007, the possible changes would be (see figure 1):

a. **Stability** of a building: the object is the same in both BD – *idGeo* is preserved.

b. **Creation** of a building: the object is removed from the copy of the 2007 DB to create the 1999 DB.

c. **Destruction** of a building: an object is created in the copy of the 2007 DB to create the 1999 DB. A new *idGeo* is created.

d. **Modification** of a building: the object is manually modified in the copy of the 2007 DB to create the 1999 DB (the type of modification is computed in post processing by comparing the geometric attributes). As the object still exists, *idGeo* is preserved.

e. **Restructuring of order 1-n**: (e.g. substitution of a collective housing by several individual buildings) several removals and a creation of object are operated on the copy of the 2007 DB to create the 1999 DB. A new *idGeo* is computed for the created 1999 object.

f. Restructuring of order n-1: (substitution of several buildings by a single one, like the destruction of individual buildings replaced by a collective housing unit) one removal and several creations of objects are operated on the copy of the 2007 DB to create the 1999 DB. Several new *idGeo* are computed for the created 1999 objects

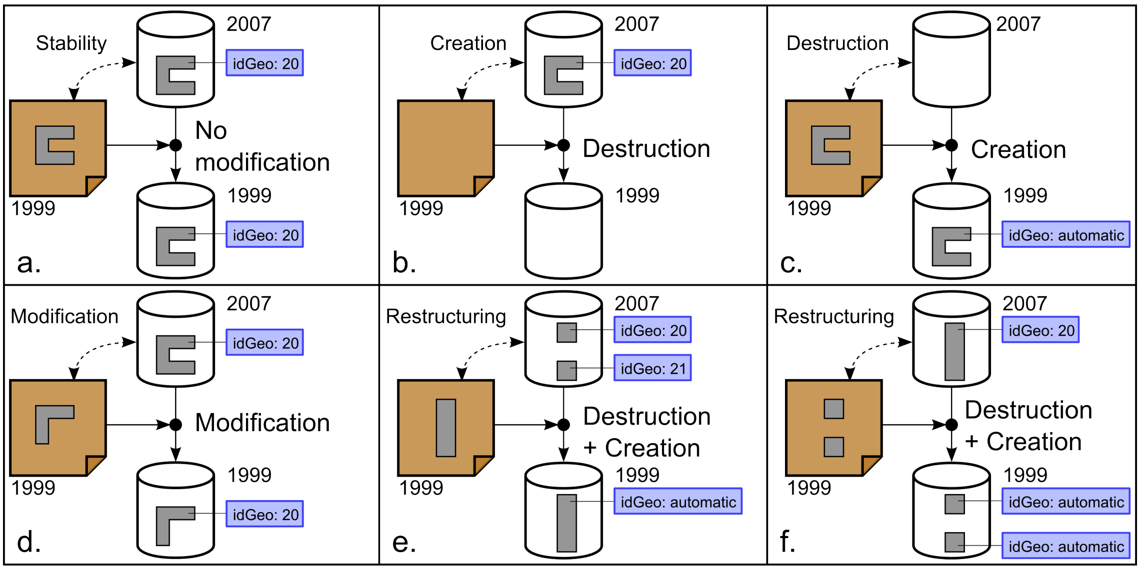

Figure 1. Tracking an object between 1999 and 2007.

3. Representing spatio-temporal data and simulation

This section presents our application schema for STDB and the simulation of urban dynamics. The integration of time in Geographical Information Systems has been widely addressed (Langran, 1993). Nevertheless, it is getting even more attention as National Mapping Agencies update their topographic databases and begin to create their first STDB. The issue of integrating time in GIS has slid from a theoretical problem to a practical one. The solution presented here is practical: it proposes to tag each topographic representation with the date coming from the data used to capture it (its source date). As shown in the previous section, each topographic database can therefore be seen as a snapshot of the system under study at a given date. These snapshots, along with their identifiers, can be used to identify the elements constituting the system: namely, its constitutive geographic agents. Each geographic agent is thus linked to one or more representations which can be used to study its evolution, its dynamics. Figure 2 shows a geographic agent (the Louvre) and 4 of its topographic representations at 4 different dates (1589, 1643, 1871 and 2007).

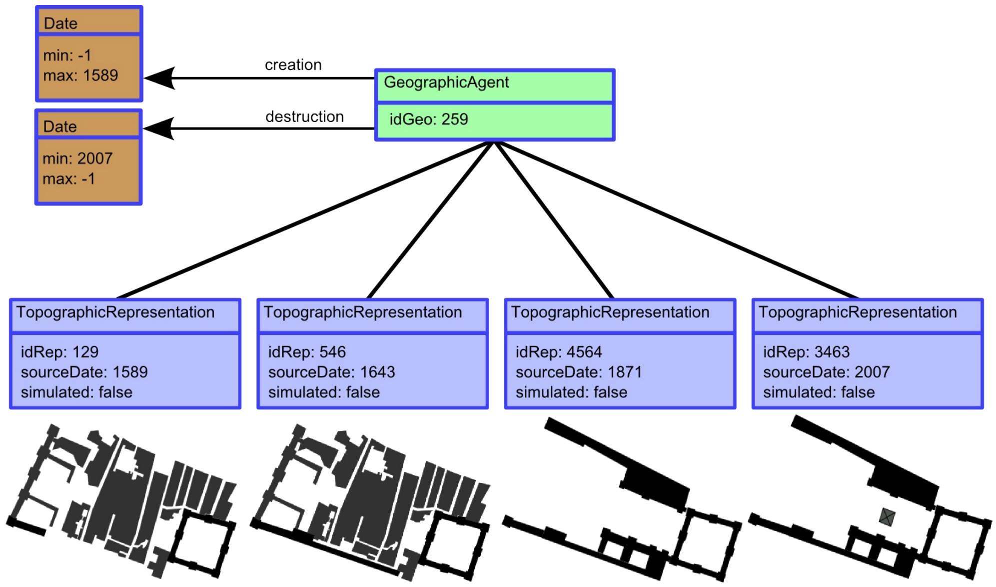

Figure 2. A geographic agent and its topographic representations.

Representation of space and time

Apart from its identifier *idGeo* (see section 2), a geographic agent also owns multiple topographic representations, an estimated date of construction and one of destruction (cf. figure 3). Topographic representations can be simulated (created through the simulation process) or not (when extracted from topographic databases). Each representation has an identifier *idRep*, a geometry, a source date, and a link to the geographic agent it represents. Representations are Features according to the OGC definition (OGC, 1999).

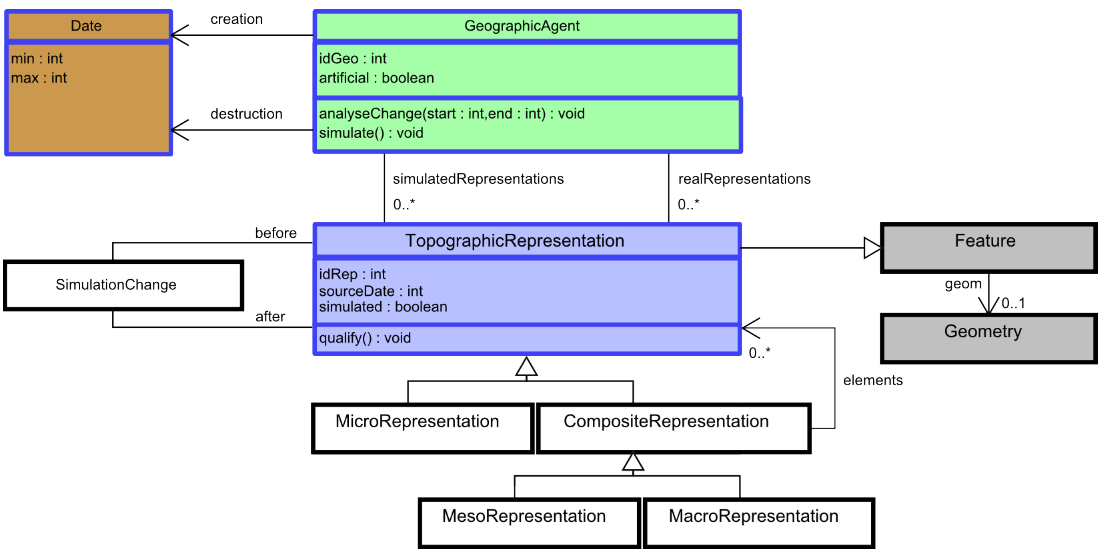

Figure 3. Representation of geographic agents and topographic representations.

A multi-level representation and its construction

The creation of different geographic levels is an important process which allows for the study of change and urban dynamics at several levels. For each topographic database, three main levels of representation are considered: micro, meso and macro (Ruas 2000). Micro-representations include buildings, roads, rivers, railways, etc. Meso-representations are aggregations of micro-representations such as groups of buildings, elementary areas, aggregated areas, built areas, communication networks (built from linear micro-representations such as roads, rivers and railways), etc. Finally, macro-representations are populations of micro and meso-representations: they contain all the objects of a certain type (e.g. a population of buildings contains all the buildings from a topographic database).

Most of the micro-representations are directly extracted from topographic databases. Others are created using other topographic objects by means of spatial analytical tools (see module 2 in section 1). Such objects include empty-spaces (built from elementary areas and buildings), crossroads (built from the communication network), etc. The structuring of urban space from micro representations is composed of 5 main steps (from micro to meso) illustrated in figure 4.

Step 1: Built areas (urban areas) are created using building buffers.

Step 2: Built areas are split into elementary areas (urban blocks) using the communication network (composed of roads, railways and rivers).

Step 3: Adjacent elementary areas similar according to their characterisation (density, function, etc.) are merged into aggregated areas (districts).

Step 4: Elementary areas are further decomposed into structures (e.g. building clusters).

Step 5: Empty spaces are created inside elementary areas by subtracting building buffers to the elementary areas' surface.

An example of the hierarchy creation process is given in figure 5 on 2007 topographic data from Orleans.

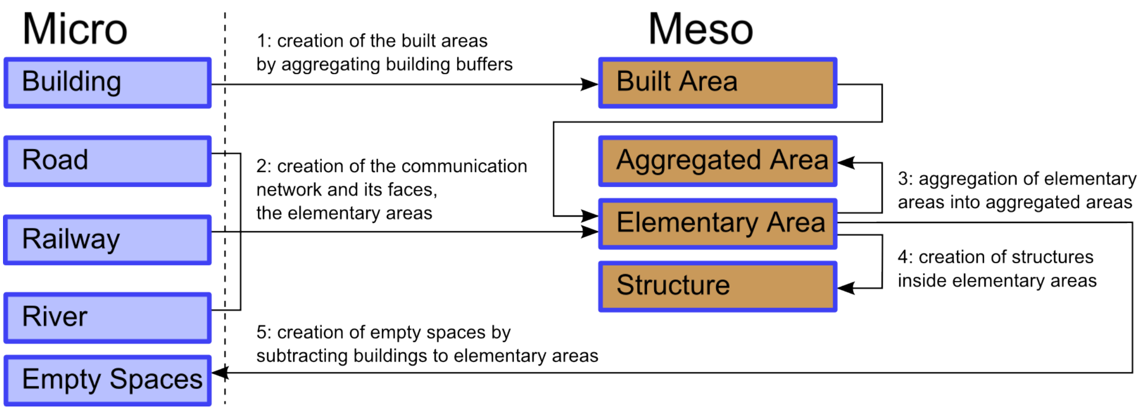

Figure 4. The hierarchy creation process and its 5 main steps.

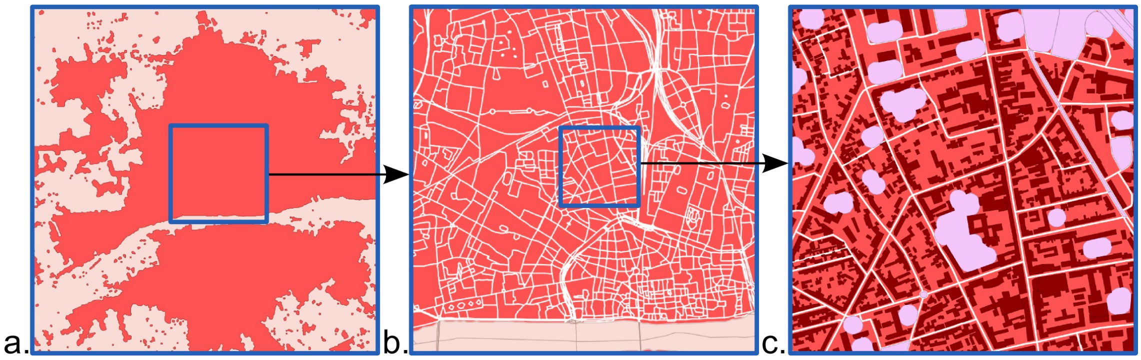

Figure 5. Different steps of the hierarchy creation process: **a.** built areas (in red) created during step 1, **b.** elementary areas (in red) created from the communication network (in white) during step 2, **c.** elementary areas, buildings (in dark red) and empty spaces created during step 5.

Representation of change

Once the hierarchies are created, the study of the evolution of geographic agents can begin. When two consecutive representations of the same geographic agent have been identified, the change taking effect between them can be analysed. A list of the most common changes is given in section 2. The list of all changes between two databases constitutes a differential database allowing the system to analyse the most common changes for each type of objects on the entire database, but also to identify specificities of each geographic partition of the database (see module 5 in section 1).

From its topographic representations and their changes, we can guess the dates of creation and/or destruction of a geographic agent. For example, in the example shown in figure 2, the creation of the building can be determined to have happened before 1589 and its destruction not to happen before 2007. Note that the list of actions that can be identified from consecutive databases is the same as the one used for the simulation.

Agents and simulation

Urban simulation is an active field of research (Barros, 2003, Batty, 2005, Benenson, 1997, Hammam, 2007, O'Sullivan, 2001) where most existing approaches are based on cellular automata or graph cellular automata whereas our approach is vector-based. Moreover, our approach is based on a multi-agents system where each autonomous geographic agent has a context (neighbourhood) it can perceive. This perception is essential to its decision making process since it influences the satisfaction of the agents' constraints. Agents have two types of constraints: macro-constraints concern all agents of the same type (e.g. buildings cannot overlap), meso-constraints are local to the meso objects the agent belongs to (e.g. the final density of the elementary area should be 0.5). Finally, all agents have to comply with the evolution rules presented in section 1.

Each agent is given a list of actions it can apply. At each step of the simulation process, an agent evaluates its satisfaction (determined by its respect of its own constraints) and all its possible actions. It then selects the best action (if any) and applies it. This agent life-cycle is illustrated in figure 6. A complete simulation begins with the activation of the topmost agents: built areas, which, in turn, activate the aggregated areas they are composed of its elements. This process is applied recursively until all agents are activated. At the end of the activation of all its agents and as they all ended their life-cycle, meso-agents check for potential conflicts between the actions taken by their elements. Meso-agents actually play the role of referee and decide which agents are allowed their actions and which are not. Eventually, meso-agents can decide the removal of some of their elements in order to allow other elements to apply their actions or simply to respect their own constraints (e.g. final density) or evolution rules.

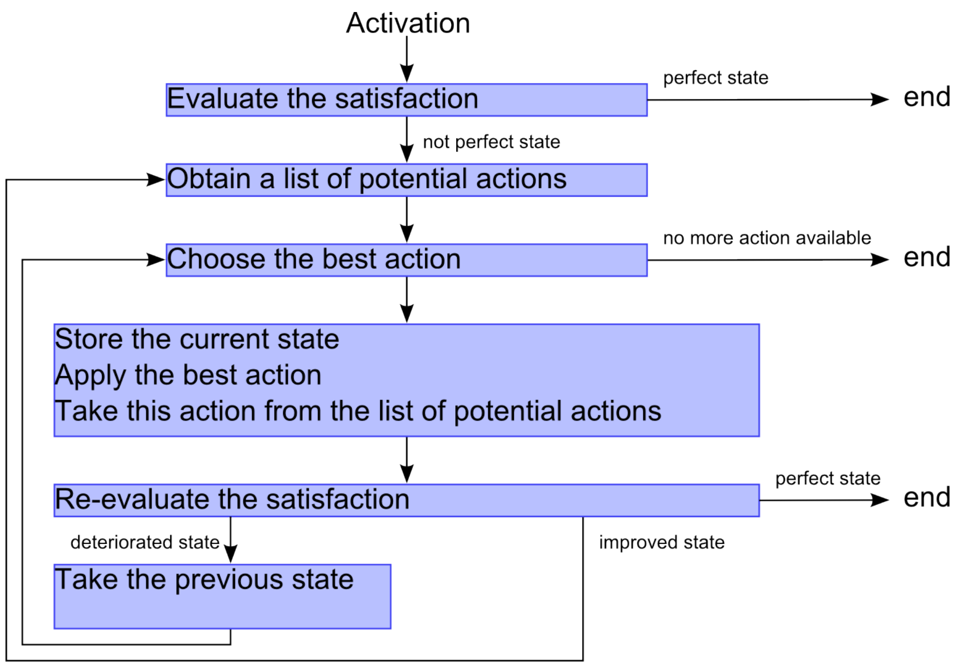

Figure 6. Life-cycle of an agent.

4. Conclusion, perspectives and future tasks

In this paper, a new approach to understand the urban dynamics has been presented. First, we proposed a methodology to create spatio-temporal databases from a single existing topographic database and several existing maps. This methodology, applied on a test area centred on the city of Orleans, allows us to track the evolutions of topographic objects in the consecutive created databases. Moreover, we introduced as well an application schema for the analysis and the simulation of urban dynamics and the bases of our simulation platform. In term, we hope that such simulations can provide new insight on urban dynamics.
Our first attempt to create STDB, presented in this paper, showed that building such databases is a costly task. Hopefully, partly thanks to INSPIRE European directive (http://inspire.jrc.ec.europa.eu/), topographic databases are getting better, especially by

integrating temporal information and unique identifiers for topographic objects. In term, one could hope for such identifiers to be made temporal-coherent and thus for the creation of STDB to be eased. Nevertheless, this will not solve the problem of acquiring older data.

In the context of the GeOpenSim project, using data matching to link several topographic databases automatically will be a focus to future work that should hopefully speed the creation of STDB up. Finally, we are still building the last bricks of our simulation platform, and further research is ongoing. A particularly interesting area of research in the model calibration of simulation rules with topographic data. For instance, using machine learning to identify and generalise rules from spatio-temporal databases will provide us with a rich variety of simulation rules and the tools to identify location-specific rules.

To conclude, using topographic databases for urban dynamics analysis and simulation is a real challenge which opens up the use of such topographic databases to broader geographical studies and applications.

5. References

J. X. Barros (2003). `Simulating Urban Dynamics in Latin American Cities'. In *proceedings. of the 7th International Conference on GeoComputation*, University of Southampton, United Kingdom.

M. Batty (2005). *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. MIT Press, Cambridge MA, USA.

I. Benenson & J. Portugali (1997). `Agent-Based Simulations of a City Dynamics in a GIS Environment'. In *COSIT '97: Proceedings of the International Conference on Spatial Information Theory*, pp. 501-502, London, UK. Springer-Verlag.

Y. Hammam, et al. (2007). `The dynamic geometry of geographical vector agents'. *Computers, Environment and Urban Systems* **31**(5):502-519.

K. Hornsby & M. J. Egenhofer (1997). `Qualitative Representation of Change'. In *COSIT '97: Proceedings of the International Conference on Spatial Information Theory*, pp. 15-33, London, UK. Springer-Verlag.

G. Langran (1993). *Time in geographic information systems*. Technical issues in geographic information systems. Taylor & Francis, London.

A. Mas (2008). `Temporal follow-up of morphological and social changes in urban zone'. In *AutoCarto'08: Proceedings of the 17th International Research Symposium on Computer-based Cartography.*

D. O'Sullivan (2001). `Graph-cellular automata: a generalised discrete urban and regional model'. *Environment and Planning B : Planning and Design* **28**:687-705.

OGC (1999). *The OpenGIS Abstract Specification - Topic 5: Features, version 4.* OpenGIS Project Document Number 99-105r2.

A. Ruas (2000). `The roles of meso objects for generalisation'. In *Proceedings of the 9th International Symposium on Spatial Data Handling*, vol. 3b, pp. 50-63.