

# MECOSIG Adapted to the Design of Distributed GIS

Fabien Pasquasy, François Laplanche, Jean-Christophe Sainte,  
and Jean-Paul Donnay

Unit of Geomatics, University of Liège, allée du 6 Août, 17 (B5),  
Sart Tilman, 4000 Liège, Belgium  
{pasquasy, laplanche, sainte,  
donnay}@geomatique-liege.be  
<http://www.geo.ulg.ac.be>

**Abstract.** For more than ten years MECOSIG has been used as a method for GIS design and implementation in various national and international projects achieved in our laboratory. During a decade, the method has been progressively improved and extended without modification of its basic principles. However the emergence of distributed GIS, implying several organizations capable to play various roles, requires the reappraisal of the methodology. New concerns are identified and a collection of new tools must be deployed. Taking the most of various recent researches completed for public authorities in Belgium, this paper presents some significant adaptations of the original MECOSIG method in order to cope with a distributed GIS environment.

## 1 Introduction

MECOSIG (*M*ethod *e* de *C*onception de *S*ystèmes d'*I*nformation *G*éographique) is one of the first methods devoted to GIS design and implementation. Developed in the context of PhD thesis [13] it has been published in 1996 [18]. The method considers a GIS as an information system and, according to the systemic approach, as a component of an organization. The methods dedicated to the design of information systems prevailing at that time (e.g. MERISE in France or SDDADM in U.K.) showed various deficiencies: inflexibility (Cartesian succession of predetermined steps to follow), lack of specific formalisms for spatial data modelling, etc. MECOSIG endeavours to fill in these gaps. It is based on a systemic analysis of the organization concerned by the implementation or the computerization of the GIS and the organization itself is at the core of the method, not only a single database project. Since then this viewpoint is shared by the few published GIS methodological guides [11] [22]. However, MECOSIG is likely the first to cover the complete lifecycle of the GIS and it comes with genuine tools supporting the method which are partially presented in the next paragraphs. The reader interested by the advances and benefits of MECOSIG will find a comprehensive list of references at the end of this paper.

Data, processes and data flows are of course in the list of the main concerns of the method, but the functional structure of the organization and integration controls are added at the same level of concern. During the design process these classes are analysed at different abstraction levels: from the descriptive level to the logical-physical one, corresponding to the implementation as such. The complete design

protocol can be summarized in a monitoring matrix (table 1) where every cell relates back to specific tools (formalisms, models, prototyping, etc.) and semantic concerns through metadata and documentation. The designer can adopt the way he/she will follow the steps suggested in the matrix according to the special features of the design strategy.

**Table 1.** The monitoring matrix of MECOSIG

| <i>Classes of concern</i> | <i>Organization</i> | <i>Data</i> | <i>Data flows</i> | <i>Processes</i> | <i>Integration controls</i> |
|---------------------------|---------------------|-------------|-------------------|------------------|-----------------------------|
| <b>Abstraction levels</b> |                     |             |                   |                  |                             |
| <b>Descriptive</b>        |                     |             |                   |                  |                             |
| <b>Conceptual</b>         |                     |             |                   |                  |                             |
| <b>Organizational</b>     |                     |             |                   |                  |                             |
| <b>Logical-physical</b>   |                     |             |                   |                  |                             |

Since the first publications dealing with MECOSIG, several improvements and adaptations have been proposed, notably in the framework of graduate and PhD theses. The most significant advances concern the tools available to conceptualize the classes of concern. In this regard it is worth noting that MECOSIG already came with an original and rich formalism for geographical data modelling – so called CONGOO for *CONceptualisation Géographique Orientée Objet* [15] – and a sophisticated typology of topological rules. Improvements have been suggested for instance in order to integrate geographic objects with indeterminate boundaries into conceptual data models [19] and to extend the formalism to spatio-temporal features [21]. Two major advances consisted, on the one hand, to cope with 3D spatial information and relationships [3] [4] and, on the other hand, to introduce UML schemas at different abstraction levels of the design methodology [9]. At the same occasion, the topological rules were simplified and re-formalized. MECOSIG is flexible enough to put up with all these additions and adaptations as long as they do not modify the principles underlying the methodology. In concrete terms, MECOSIG was able to support numerous projects achieved in various public organization (e.g. [14] [16] [17]).

The emergence of distributed spatial databases alters the foundations of information system design. It is particularly true with MECOSIG which puts the organization at the root of the design issue. Nowadays many projects dealing with GIS in Belgium require interoperability and other distributed capabilities, while organizations are involved in spatial data infrastructures at the regional and national levels [5]. Facing with such issues, it was necessary to adapt MECOSIG and to provide new tools fitting the requirements of distributed environments. This paper takes the most of various recent researches that we have completed for public authorities in order to develop common GIS platforms, notably in the framework of the SIGMaTE project [2].

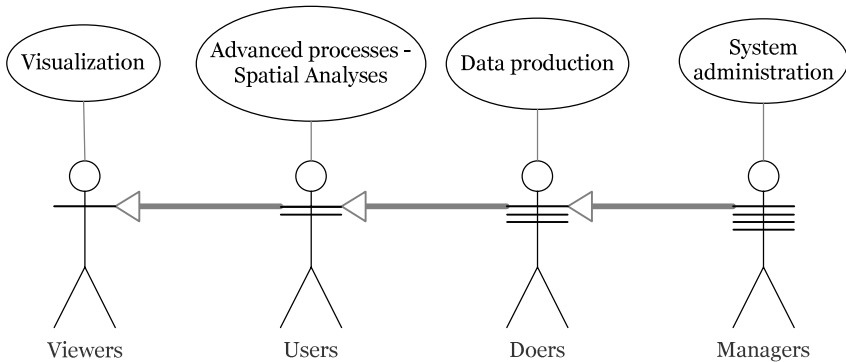
## 2 Adaptations at the Descriptive Level

Some changes were introduced when compared to [18]. MECOSIG proposed to go twice through the monitoring matrix, first for the analysis phase and secondly for the

conception step. For clearly pragmatic reasons the formal current system analysis and the requirement analysis are henceforth performed at the descriptive level appearing during the phase of conception. It is justified because – practically – the course of the matrix at the analysis level generally resulted in textual descriptions only and hence it generates a waste of time due to the redundancy with the descriptive level of the phase of conception.

It is worth noting that the design of distributed GIS is always based on existing systems which work in a more or less effective way. Distributed GIS implementation is the opportunity to re-conceptualize the spatial databases which have to be distributed. Moreover a common irredundant model of these databases must be proposed on the basis of a single spatial reference framework.

These adaptations affect the classes of concern of the monitoring matrix. The tools used to describe the current situation and to formalize the requirements are the same and they are presented below. This enables to quickly identify the current gaps of the systems and to propose a conceptual solution of the problems in a quite short time.



**Fig. 1.** The four categories of actors and the inheritance of their privileges

In order to describe organization(s) – there could be more than one organization in case of distributed GIS – actors of this or these organization(s) must be identified. These actors are classified in four categories: viewers, users, doers and managers. A set of privileges (figure 1) and specific use cases match every category of actors in the organization(s). Particular symbols can be introduced to distinguish the categories (figure 1). Moreover a specialization relationship exists between these groups. The actors belonging to the higher levels inherit the properties from the lower levels (figure 1). For example, the users inherit the rights granted to the viewers. This classification of actors is not strictly fixed and can evolve according to the cases. Every category can be subdivided to match as well as possible the modeled organization.

The distinction made between doers and managers constitutes an essential characteristic of GIS. Indeed data production and system management are two well differentiated spots which are generally entrusted to different GIS actors in an organization.

The class of concern called "processes" is strongly related to the previously identified actors and in particular crossed processes are specific to distributed GIS. In order to describe them the interactions between the four groups of actors and the system are identified in terms of use cases and they are modeled with UML use case diagrams [12].

Concerning data, there are some changes due to the distributed aspect of the system. Data are still described in a catalogue and their quality must still be analysed. In this respect the use of the standards ISO 19110 [7] and ISO 19115 [8] is suggested. However the observance of standards is not a guarantee of success in a distributed environment as long as data and metadata are concerned. They are essential for specifying the form of the "packaging" but interoperability requires further agreements on the "content" and its meaning. That is in this very place that a significant improvement of the MECOSIG method must be achieved in order to incorporate semantic rules in the spatial data domain. A first step in this direction is to identify data according to the use cases in which they are implied and finally to gather data into batches. Every batch constitutes an intuitive group maintaining a certain number of relations (semantic and/or topological) and intervening together in peculiar activities. Yet the description of the relationship is not mandatory at this level of detail. It will be taken into account at the stage of conceptual modelling. As an example, a specific batch consists in the spatial data of reference which are common to the whole distributed spatial database.

The use cases make an intensive use of data flows. Because the description of the use cases is already performed in the "data" and "processes" classes of concern, the analysis of a specific class of concern devoted to the data flow can be withdrawn in many applications.

### 3 Adaptations at the Conceptual Level

The UML class diagram is now used as a surrogate for the CONGOO formalism originally proposed with MECOSIG. There are several reasons to explain this transition.

First UML is internationally admitted as standard for system modeling and many case tools are already using it.

Then UML seems to be more suitable for the design of distributed systems. Indeed it offers more than only one diagram to represent the reality and, based on its philosophy, the design can combine more than one model of diagrams to depict a context.

The design is based on the batches identified at the descriptive level. A class diagram is suggested for every batch. That will allow thereafter a more effective distribution of the data between the various spatial databases.

Due to the use of UML instead of CONGOO some specificities of the spatial data design are withdrawn while others become available. The first and probably the most important one is the ability to identify the objects' geometry. A lot of studies (e.g. [1]) exist to cope with this problem and moreover some mechanisms of extensions are defined in UML. Stereotypes or tagged values permit to adapt UML to the domain of

interest [12]. Another specificity of spatial data is worthy of interest: the management of topological relationships. UML class diagram does not allow the definition of topological relationships like CONGOO does. Moreover these added formalisms can make the model more complex. Introduction of the topological relationships into the model such as suggested in CONGOO is therefore given up. On the other hand, the concept of topological matrix is kept and adapted. One or more topological matrixes can be defined for every data model. Two types of topological matrixes are available: the traditional matrix [18] and the strong topological matrix [9] (figure 2). The latter proposes to define topological constraints having to be respected by all objects belonging to a class.

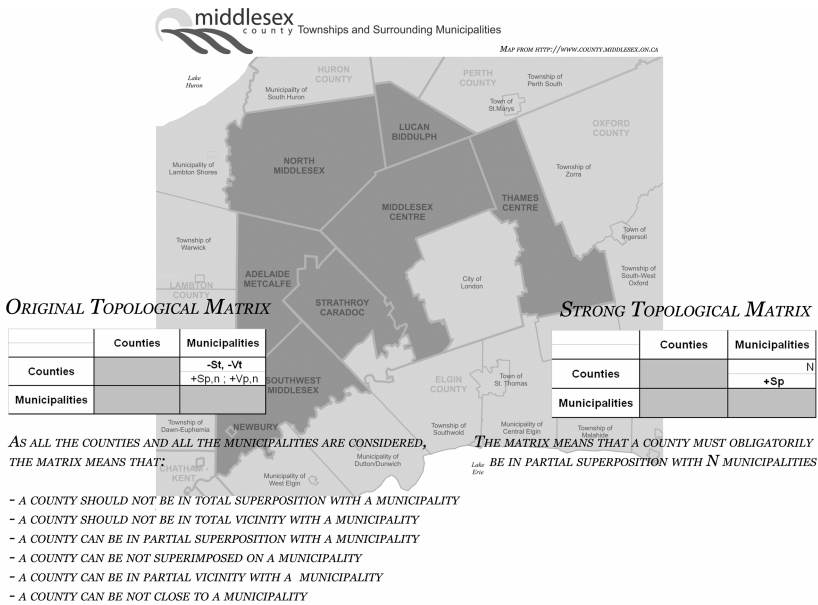


Fig. 2. The two kinds of topological matrixes

Concerning the processes design, every use case is extracted from the descriptive level and detailed using UML activity diagrams [12]. The formalization is independent of the implementation. Nevertheless, the activity is replaced within the subsystem of the organization in which it is carried out. The distributed aspect of the system starts to appear in a coarse way. On this level the data catalog can also be enriched by the mention of the activities making use of each data.

Finally sequence diagrams are also used to conceptualize some specific scenarios of activity diagrams [20] (figure 3). It consists in typical sequences describing the interactions between the data and the actors of the organization. The classes of concern "data flow" and "organization" are thus taken into account by the designer.

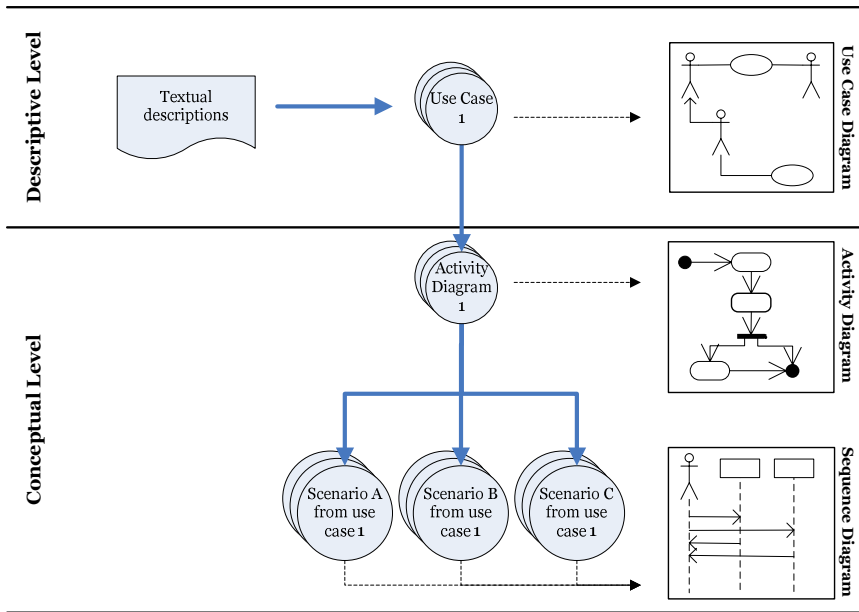


Fig. 3. The tools used at descriptive and conceptual levels

#### 4 Adaptations at the Organizational Level

Facing the implementation of a distributed infrastructure, the components of the system evolve according to their own requirements. From the implementation point of view important organisational constraints must be met. First, the missions of all the actors should be assessed and possibly redefined. Then new tasks are introduced in order to guarantee maintenance, permanent working and growth of the distributed system. Representative committees in charge of coordination tasks will arise for this purpose. These committees will also aim at coordinating actors from the various organizations in-bedded in the distributed system.

Data exchange must be performed easily and in a transparent way for all users of the distributed GIS. The advised use of the data cannot be guaranteed without a free access for all possible users to data about data, i.e. metadata. In this regard again, it is suggested to refer to the ISO standards [7] [8]. Particularly, the metadata related to the quality play a major role in the data usability and in the appraisal of the final results essential in decision-making.

As one of the main objectives of a distributed system is to share data from different organizations, new data flows are generated and others have to be adapted. New applications based on this infrastructure may appear and they can be distributed between all the actors.

Data and/or applications access can be restricted or subject to authorizations. On the other hand acquisition policies for common data could be considered. All these agreements must be considered at the level of the distributed organization and will fall

under the responsibility of the representative committees in charge of the coordination tasks. Some formalization can also be introduced to depict the committees' tasks thanks to the previously proposed tools: UML use case, activity or sequence diagrams according to the steps illustrated on figure 3.

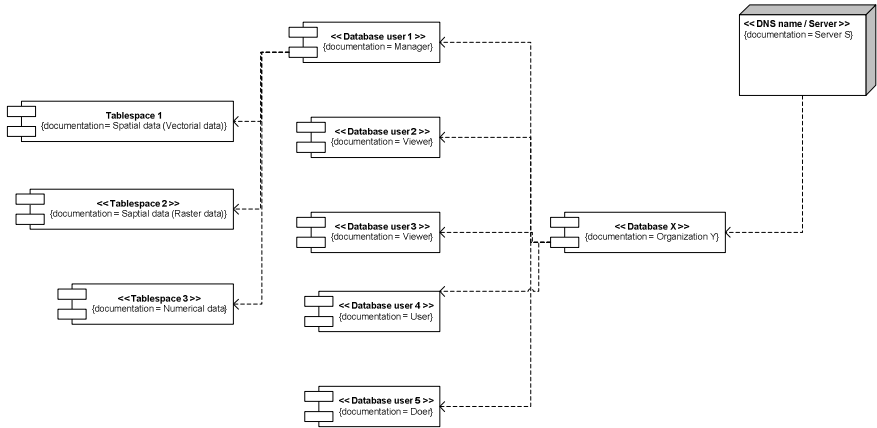
## 5 Adaptations at the Logical-Physical Level

Because of the numerous actors involved in a distributed system, the roles and tasks of each of them should be clearly specified. The specification and the possible reallocation of the human resources have not to be underestimated because they will condition the implication of the actors in the new system which will be set up. This level offers the opportunity to assess the tasks and the objectives or at the contrary to reorganize the work of each actor. If required, sensitizing and training sessions can be planned and organized in due time in order to facilitate the acceptance and the support for the re-engineered system. A specific attention will be given on the constitution of the coordination task group which must be competent, representative of the various departments or organizations concerned by the distributed spatial system and which must receive the capacity to act decisively. Basically all these tasks are not specific to the implementation of a distributed system however the degree of difficulty is magnified in a distributed framework.

From a technical point of view, some managerial economy and economy of scale could be achieved thanks to the implementation of the distributed system. The acquisition of common reference data and the globalisation of software licence costs are two examples between many others. However the objective of the distributed operation is certainly not the standardization of the hardware and software resources. GISs, DBMSs and data/application servers likely will remain in the different parts of the distributed organization, in order to avoid a complete disrupt in the respective businesses. The technical analysis which must be completed at the physical level concerns the add-on hardware/software needed to guarantee the required degree of interoperability specified during the design analysis. The main reason of a technical study is to identify precisely the distributed system components but also to assess the capability of the whole to provide the required level of service. The ability of a user to access consistently and coherently spatial data and processes across distributed databases depends on semantic interoperability [6]. Providing systems for cross-relating items of information across multiple sources is needed to solve problems.

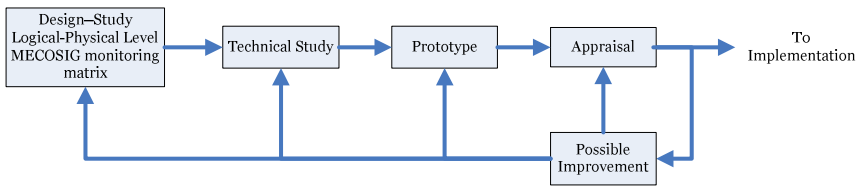
As early mentioned, UML class diagrams were used to build up a conceptual model allowing to note classes, attributes and associations needed to describe the distributed databases structure. The modelling process gives place to the design of many objects which must be deployed in a database environment and that claims many decisions concerning the logical-physical level [10]. In this connection, the design aspects are mainly related to the data storage structure. The design of the data storage and the knowledge of their physical location on servers, discs (even of multiple partitions) will facilitate the final management of the databases. Using UML to model tablespaces, to specify the discs containing them and their physical location are very important tasks which must be planned carefully and precisely. Components and deployment UML diagrams offer objects well suited to describe all these

functions and of course to expose them graphically. For instance (figure 4), one database of the distributed system is designed and is related to various tablespaces and users by some dependence relationships.



**Fig. 4.** Components diagram (UML) designing data storage structure

Activity and sequence UML diagrams produced at the conceptual level can also be improved by adding objects dealing with data flows, procedures and process according to GIS software, tools and DBMS specificities which will be used in the distributed system.



**Fig. 5.** The evolutionary project pilot

Lastly and recommended, whole of the logical-physical level is integrated in an evolutionary project pilot. A technical study of the numerous software components from the identified infrastructure will be considered. A prototype will be implemented as well on the basis of elements previously identified (figure 5).

In this connection, it is still necessary to find out the performing level of networks, DBMS and spatial data servers, GIS software, etc, in order to guarantee the implementation of an appropriate distributed system. The interoperability level required (data interoperability and/or GIS services interoperability) is decisive in the technical choice of distributed system components.



## 6 Conclusions

Distributed GIS and spatial data infrastructures have the special feature of being made over existing GISs and DBMSs. This does not imply that the design analysis of the future distributed system can be avoided. On the contrary, this kind of project requires a sound study from the conceptual to the physical level. MECOSIG proved to be flexible enough to support such design analysis provided that it is assisted by suited tools taking place in a revisited monitoring matrix. This paper shows that several diagrams issued from UML can offer interesting capabilities in the field.

Based on actual experiences requiring a comprehensive level of interoperability between distributed data, this paper is deliberately vague about the possible implementation of distributed processes. The tools dedicated to the processes (formal description...) presented in this paper could provide some elements in this direction. However a comprehensive review of this topic would require further experiences embedding likely a stronger dependency on software capabilities.

## References

1. Bédard, Y, S. Larrivée, M.J. Proulx, M. Nadeau: Modeling Geospatial Databases with Plug-Ins for Visual Languages: A Pragmatic Approach and the Impacts of 16 Years of Research and Experimentations on Perceptory, S. Wang et al. (Eds.): COMOGIS Workshops ER2004, LNCS 3289, (2004) pp. 17–30.
2. Bels F, Capoen E., Pasquasy F., Swennen C.: Projet d'assistance technique, méthodologique et scientifique dans le domaine de la cartographie et des S.I.G., en vue d'implémenter un prototype d'infrastructure d'information géographique répartie entre la DGATLP et la DGRNE du Ministère de la Région wallonne, Final report, unpublished, (2003) 310 p.
3. Billen R.: Introduction of 3D information in urban GIS: a conceptual view. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, part B3, Amsterdam (2000), pp. 73-78.
4. Billen R., Zlatanova S.: 3D Spatial relationships model: a useful concept for 3D cadastre?, Proceedings of the International workshop on "3D Cadastre", Delft University, (2001) pp. 1-17.
5. Donnay J.-P.: Distributed GIS for sharing large scale data between public agencies. A case study in Belgium. GIS 2002 International Conference Proceedings, The Bahrain Society of Engineers, Bahrain, (2002). pp. 47-59.
6. Fonseca F., Egenhofer M.: Ontology-Driven Geographic Information System, 7<sup>th</sup> ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, C. Bauzer Medeiros (ed.), (1999)
7. ISO/TC211 19110: Geographic Information: Methodology for feature cataloguing (2004)
8. ISO/TC211 19115: Geographic Information: Metadata (2003) 140 p.
9. Laplanche F.: Conception de Projet SIG avec UML. Bulletin de la Société géographique de Liège, n° 42, (2002) pp. 19-25.
10. Naiburg E., Maksimchuk R.: Bases de données avec UML. Collection Référence, CampusPress, Gap, (2002) 293 p.
11. New York State Archives, Calkins H.: Geographic Information Systems Development Guides (1996)
12. [http://www.archives.nysed.gov/a/nysaservices/ns\\_mgr\\_active\\_gisguides.shtml](http://www.archives.nysed.gov/a/nysaservices/ns_mgr_active_gisguides.shtml)
13. Object Management Group: UML Specification version 2. <http://www.uml.org>. (2004)

14. Pantazis D.: Analyse méthodologique des phases de conception et de développement d'un système d'information géographique. PhD thesis unpublished. University of Liège. (1994)
15. Pantazis D.: Le développement de la base de données géographiques du Ministère de l'Aménagement du Territoire du Grand Duché de Luxembourg, actes du colloque Les systèmes d'Information en Géographie, Org.: Université de Genève, Université de Fribourg, Chexbres, Suisse, (1995)
16. Pantazis D.: CON.G.O.O.: A conceptual formalism for geographic database design, Geographic Information Research, Bridging the Atlantic, Ed. M. Craglia & H. Coucledis, Taylor & Francis, (1997) 348-367.
17. Pantazis N. D., Cornélis B.: Designing and implementing a GIS in an international context, Transactions in GIS, vol. 1, n° 4, (1997) 301-320
18. Pantazis D., Cornélis B., Billen R., Sheeren D.: Establishment of a geographic data dictionary: case study on the Brussels regional government GIS. Computer, Environment & urban Systems. 26(1), (2002) pp. 3-17
19. Pantazis D., Donnay J.P.: Conception de S.I.G.: Méthode et formalisme. Collection Géomatique, Hermes, Paris, (1996) 352 p.
20. Pantazis D. & Donnay J-P.: Objets géographiques à limites indéterminées. Modélisation et intégration dans un modèle conceptuel de données. Revue internationale de géomatique, 7(2), Hermes, (1998) pp. 159-186.
21. Roques P.: UML par la pratique: étude de cas et exercices corrigés. Eyrolles, Marsat. (2001) 302 p.
22. Sheeren D.: La conception de bases de données spatio-temporelles, problématiques et solutions dans le cadre du formalisme CONGOO, Graduate thesis unpublished, University of Liège, (1999) 113 p.
23. Somers R.: Developing GIS Management Strategies for an Organization. Journal of Housing Research, Vol.9, n° 1, (1998) pp. 57-77