

EXTENDING THE UML-GEOFRAME DATA MODEL FOR CONCEPTUAL MODELING OF NETWORK APPLICATIONS

Sergio Murilo Stempliuć, Jugurta Lisboa Filho, Marcus V. A. Andrade
Departamento de Informática – Universidade Federal de Viçosa (UFV)
36570-000 – Viçosa, MG, Brasil

Karla A. V. Borges
Prodabel – Empresa de Informática e Informação do Município de Belo Horizonte
Av. Pres. Carlos Luz, 1275, 31230-000 – Belo Horizonte, MG, Brasil

Keywords: Geographic databases, Conceptual data model, Network applications.

Abstract: This paper presents an extension of the UML-GeoFrame data model that includes a set of new constructors to allow the definition of conceptual schemas for spatial database applications whose elements relationship forms a network. Also, it is discussed how the GeoFrame conceptual framework is changed with the inclusion of new metaclasses and the corresponding stereotypes related to network elements. The extension proposed in this paper is evaluated using a class diagram for a water distribution company.

1 INTRODUCTION

Conceptual data models have been refined to meet particular details found during the conceptual modeling of geographical applications, allowing the abstraction of geographical phenomena according to spatial and temporal dimensions. These models also allow the representation of geographical and non-geographical entities, as well as the specification of spatial, temporal and semantic integrity constraints.

Among several models created for Geographical Information Systems (GIS), it is important to stand out the balance between coverage and simplicity of these models (Friis-christensen et al., 2001). As pointed out by Kösters et al. (1997) and Spaccapietra et al. (2006), some models define a large set of elements trying to be general enough to be applied in several domains, but usually, they are too complex to be used by users and developers. Other models are simpler and easier to use but are restricted and it is necessary to include some additional textual information to describe a specific domain.

A subject in GIS domain that has not been receiving the necessary attention from authors of these models is network applications, that is, those

applications in which the elements relationship should be described by a network, such as water, road, hydrographical, telecommunications and energy distribution networks. One of the first conceptual models trying to address this problem in detail was the GeoOOA model (Kösters et al., 1997). Other more recent models as OMT-G (Borges et al., 2001) also provide simple constructors for network modeling, but do not address the problem at the same level of detail as the GeoOOA model.

The objective of this work is to extend the UML-GeoFrame data model (Lisboa Filho and Iochpe, 2008) to provide constructors for the development of class diagrams involving network elements. The new constructors will allow the modeling of geographical databases for domains that require mapping the connectivity of their elements.

The article is organized as follows: Section 2 presents a general description of the GeoFrame framework and the conceptual UML-GeoFrame data model; Section 3 describes the proposed extension to UML-GeoFrame for network modeling; Section 4 shows a class diagram of a water distribution company; and Section 5 presents our conclusions.

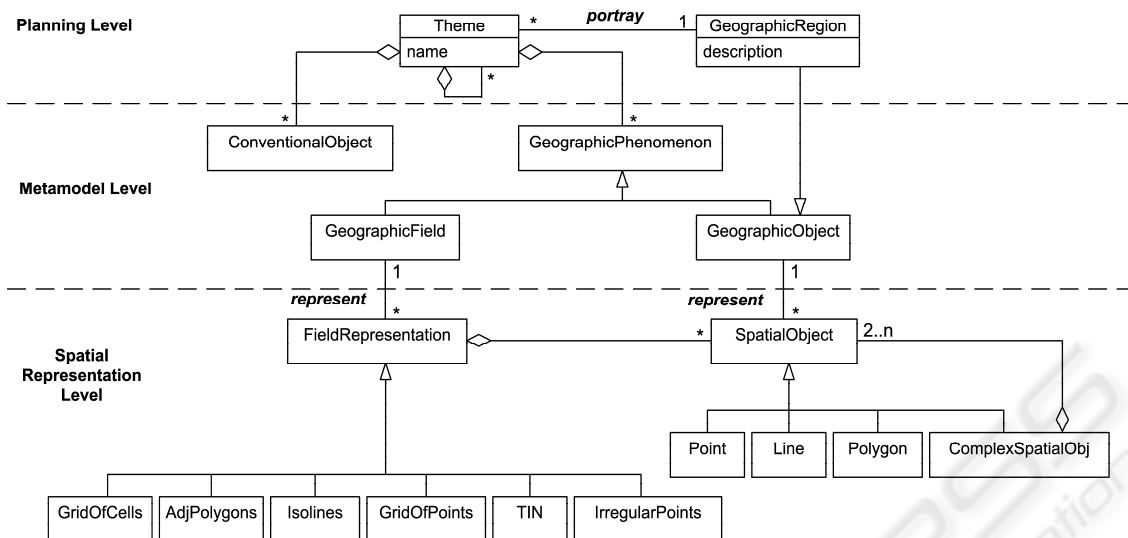


Figure 1: The GeoFrame Framework (Lisboa Filho and Iochpe, 2008).

2 UML-GEOFRADE: A CONCEPTUAL DATA MODEL FOR GIS APPLICATIONS

The UML-GeoFrame data model (Lisboa Filho and Iochpe, 2008) proposes a set of stereotypes, forming a UML profile (Unified Modeling Language) for conceptual modeling of geographical databases. Constructors of UML-GeoFrame data model are based on the class hierarchy defined in the GeoFrame framework that was originally proposed in Lisboa Filho and Iochpe (1999). Section 2.1 presents an overview of GeoFrame where only the elements needed to understand the new network constructors are described. Section 2.2 shows the use of the UML-GeoFrame stereotypes.

2.1 GeoFrame Overview

The GeoFrame framework (Figure 1) consists of a hierarchical class structure that describes the fundamental elements for the modeling of any geographical database.

At the planning level, the geographic regions corresponding to the areas of interest (*GeographicRegion* class) are defined and, for each region, it is defined the associated themes (*Theme* class), such as hydrograph, transport, relief, etc. A theme can be subdivided in a hierarchy of sub-themes.

The metamodel level is composed of metaclasses that reflect how the reality is interpreted and it can be represented by conventional data (*Conventional-*

Object) or geographic phenomena (*Geographic Phenomenon*). The latter is still specialized in metaclasses for field (*GeographicField*) and objects (*GeographicObject*) views.

Also, the level of spatial representation reflects how the reality is represented by designers and users more abstractly in relation to the representation within the database. The *SpatialObject* class generalizes spatial representation classes observed in objects view, including *Point*, *Line*, *Polygon* and *ComplexSpatialObj*. This last one represents a spatial object formed by two or more spatial objects. The *FieldRepresentation* class generalizes the field view classes, including *GridOfCells*, *AdjPolygons*, *Isolines*, *GridOfPoints*, *TIN* (Triangular Irregular Network) and *IrregularPoints*. Multiple representations are supported in the two views.

2.2 UML-GeoFrame Overview

The UML's package constructor corresponds to the first concept incorporated in the UML-GeoFrame data model. It is used at the planning level to identify themes related to the geographic region, allowing a top-down approach to the problem.

The UML-GeoFrame data model defines a set of stereotypes which are important for the application classes representation avoiding that these classes are represented as subclasses of the metamodel level. The aim is to avoid overloading the application diagram and hindering its understanding. Figure 2 presents some classes examples with stereotypes of *ConventionalObject*, *GeographicObject* and *GeographicField* metaclasses, respectively.

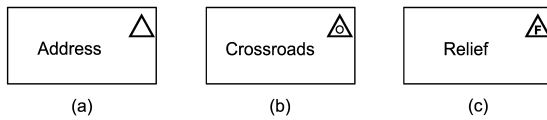


Figure 2: Stereotypes for generalization/specialization.

Stereotypes are also defined at the spatial representation level to represent different abstractions of the spatial shape of geographic phenomena. For example, in Figure 1, the *GeographicField* class is associated to *FieldRepresentation* class and *GeographicObject* associated to *SpatialObject*. Based on these associations, it is possible to model the spatial characteristic of a geographic phenomenon. The stereotypes used for the spatial representation were defined on these associations (Figure 3). Then, two stereotypes are used for each class, one reflecting the metaclass type and another reflecting the spatial representation type. A complete description of this model and examples of use can be found in Lisboa Filho and Iochpe (2008).

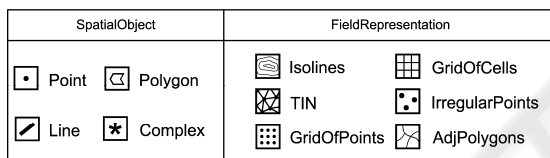


Figure 3: Stereotypes for spatial representation.

3 EXTENSION OF THE UML-GEOFRAME MODEL FOR NETWORK MODELING

This section describes the extension of the UML-GeoFrame data model so that it will provide specific constructors for modeling elements involved in a network. Four stages of the GeoFrame framework evolution are presented, identifying advantages and disadvantages found in its previous versions. The final version proposed is described in section 3.4.

3.1 Inclusion of Network Properties into Geometric Classes

The first idea of GeoFrame extension was limited to the inclusion of network properties into the spatial *Point* and *Line* classes. The advantage of this approach is to use the geometric characteristics of point and line, considering the geometric representation for the elements node and arc

respectively, although they do not depend on information such as shape, length, position, orientation and transformation. Another advantage is the non-creation of new stereotypes to represent these elements, only including attributes referring to connectivity would be necessary.

Demirel (2002) mentioned the reduction in requirements of data storage and increase in performance of query processing as advantages of this approach, but he pointed out the disadvantages in relation to data maintenance. Although topological data do not depend on geometric properties, changes in object geometry would necessarily cause the maintenance of both geometric and topological data.

3.2 Specialization from Geometry

The second alternative for GeoFrame extension was to avoid the problems found in the first proposal through specialization of *Node* and *Arc* classes from *Point* and *Line*, respectively.

This solution has the same advantages as the previous proposal, except for the inclusion of new stereotypes, and solves the problem relative to topological data maintenance when there is change in geometric data. However, as similarly discussed in Lisboa Filho and Iochpe (2008), a flaw exists in this reasoning, as the generalization/specialization concept does not recommend that different things relate in such way.

3.3 Class Hierarchy for Network Modeling

As the solution using the generalization/specialization concept was shown inappropriate, it was included a new hierarchical class structure to provide the fundamental elements for the network modeling. Figure 4 shows this new approach.

A new specialization was added from the *GeographicPhenomenon* class at the metamodel level, making it possible to interpret reality elements as network elements through the *NetworkObject* class. At the spatial representation level, the *NetworkRepresentation* class generalizes the classes of network representation, including *Node*, *Arc* and *ComplexNetworkObj*. The last one represents a network object formed by two or more network objects.

Despite the hierarchical similarity with the field and objects views, network modeling has some particularities. The first refers to the multiplicity in

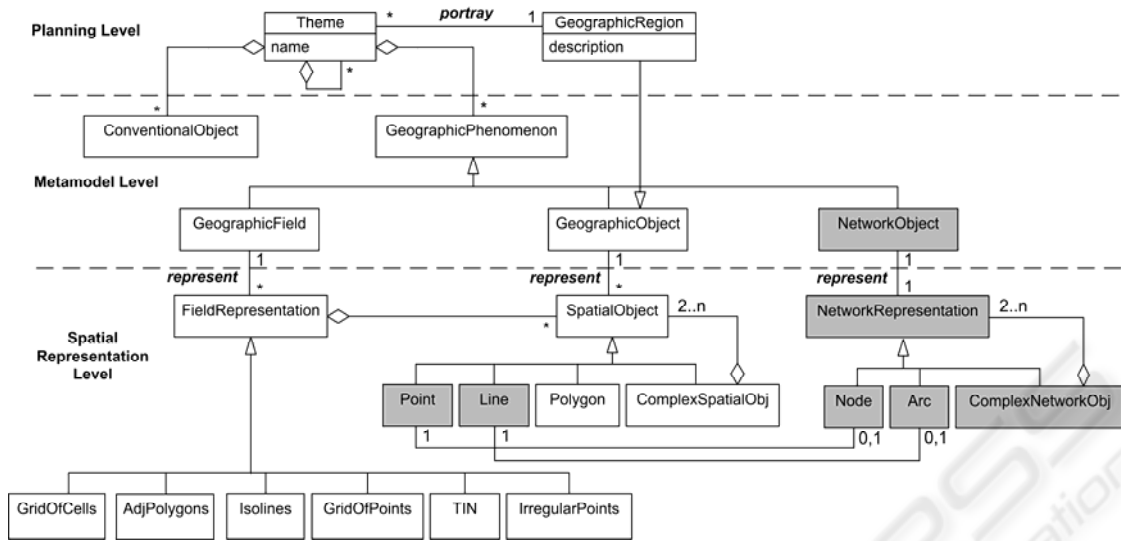


Figure 4: Class hierarchy for network modelling.

the association *represent* between the *NetworkObject* and *NetworkRepresentation* classes (Figure 4). As the association has multiplicity one to one (1:1), an instance of the *NetworkObject* class should be represented by only one instance of the *NetworkRepresentation* class, i.e., there are no multiple representations.

The second particularity refers to the association between *Node* and *Point*, as well as *Arc* and *Line*, to represent the geometric aspect of the network classes. This characteristic imposes the restriction that in any diagram using GeoFrame for network modeling, there should be a class of the type *Point* associated with a class of the type *Node*, in the same way that a class of the type *Line* should be associated to a class of the type *Arc*. These relationships become optional when the application involves only geographic phenomena in field and/or object view.

The third characteristic refers to the *ComplexNetworkObj* class. In spite of the apparent similarity with the *ComplexSpatialObj* class, this class has a different meaning. Each application of network representation will always use the *ComplexNetworkObj* class, since a network is an aggregation of nodes and arcs. Being an aggregation, the same node or arc may belong to several networks. The *ComplexSpatialObj* class in its turn is not always present in the modeling of an application using the object view.

This solution, although meeting the basic requirements for modeling of network applications, still has some problems to be solved. The first problem refers to the association between objects of

geometric and topological representation. As several geometric characteristics are not pertinent to the *Node* and *Arc* classes, their association with *Point* and *Line* makes it possible to specify queries starting from the topological classes with access to characteristics that are not about connectivity. One example is to query the attributes length and thickness of a line associated with an arc.

Another problem refers to the network representation through the *ComplexNetworkObj* class. Since it is essential in network modeling and does not add any spatial attributes to aggregated objects, a new solution became necessary to replace this class by a more abstract concept.

3.4 Final Proposal of GeoFrame Extension for Network Modeling

To solve the problems of section 3.3, the GeoFrame was changed as showed in Figure 5. The association between *Node* and *Point* was removed, as well as between *Arc* and *Line*, so that only the essential geometric characteristics are used by network elements. Besides, network elements can be associated with different types of spatial objects. As an example, both a power plant represented by a polygon and a pole represented as a point can play the role of nodes in an electric energy network.

However, the most important change was the replacement of the *ComplexNetworkObj* class at the spatial representation level by the *Network* class at the metamodel level. As in the previous version (Figure 4), only alphanumeric information

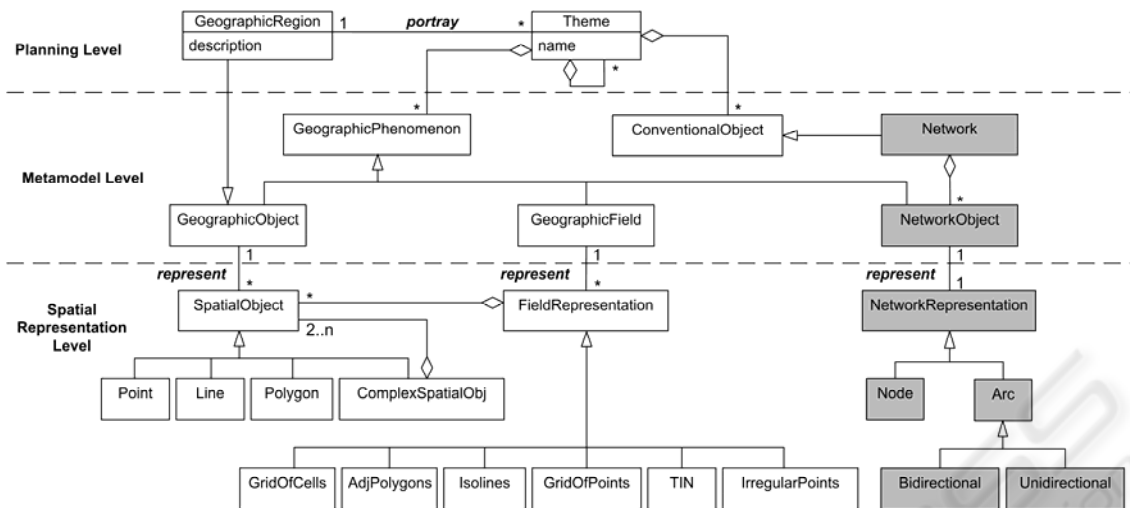


Figure 5: Final proposal of GeoFrame extension for network modelling.

was added to aggregated network objects, the interpretation of a network became the same of a conventional class, but maintaining the aggregation of network objects. To make the interpretation of the *Network* class clearer, as it was conventional class, and show that each instance of the *Network* class should belong to a package (*Theme* class), we specialize the *Network* class from the *ConventionalObject* class.

The aggregation between *Network* and *NetworkObject* implies that in the application diagram an instance of *Network* will be an aggregation of *NetworkObject* instances, allowing network handling as a composed object and allowing that an instance of *NetworkObject* belongs to more than one network.

A last modification was made in the hierarchy proposal in order to specify bidirected and unidirected networks. Although data structure for nodes and arcs are different in these two types of networks, the representation, on the other hand, only differs by the arc shape. For the GeoFrame, the need to specify the type of a network and to maintain simplicity of model is based on representation, therefore the idea consisted in specializing the *Arc* class into the *Bidirectional* and *Unidirectional* classes.

3.5 Network Stereotypes

With new classes added to the framework, new stereotypes were also defined. They follow the same UML-GeoFrame principles, with a stereotype defined for the specialization of the geographic phenomenon and another for spatial representation. The *Network* class uses the conventional object

stereotype. Figure 6 shows three examples of classes with the generalization stereotype for the *NetworkObject* metaclass, and representation stereotypes for the *Node* (a), *Bidirectional* arc (b) and *Unidirectional* arc (c) metaclasses.

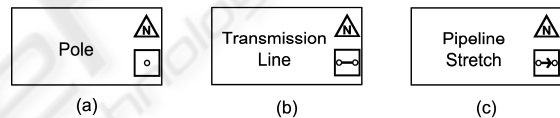


Figure 6: Stereotypes for network modelling.

3.6 Definition of Network Representation Elements

Simplicity and expressivity are characteristics of the UML-GeoFrame data model that make it easy to understand the modeled reality. On the other hand, they imply in additional textual information to build a complete diagram concerning the problem domain. UML-GeoFrame models, which provide simple constructors for developing simple diagrams, also have textual information to facilitate concept understanding.

Textual information incorporated to models is useful to define each of the classes or even to establish restrictions common to several domains. The *Node* class, e.g. is defined as a spatial representation in network object view used when the end or cross between lines is represented by a point with connectivity. The *Arc* class is also a network object used to express the connectivity between one or two nodes through a line. Both are independent of the real position or transformation applied to the element of the portrayed reality.

4 A CASE STUDY

A conceptual data schema was designed for a Water Distribution System of a city to demonstrate the use of network constructors proposed. The example was based on Barros et al. (1995) and was supported by the staff of the Viçosa Municipal Water and Sewer Service (SAAE, 2008). We present only the fundamental elements of the system that was restricted to consider only the part that conveys water from the treatment plant to house hydrometers.

Two networks of water distribution were identified in a city: the main network, which conveys water from the treatment plant to the connection points with secondary pipelines, passing through reservoirs; and the secondary, formed by the distribution grid and consumer's direct water supply.

The reservoirs provide water to supply the demands of consumption, any emergency demands and also, they keep the minimum or constant pressure within the network. Regarding network location, reservoirs can be located upstream of the distribution network, for normal supply; or downstream, to store unused water during hours of lower demand and supply the network during hours of higher consumption (Barros et al., 1995).

The water main stream is the largest-diameter pipeline responsible for conveying water from the treatment plant to the upstream reservoir. The main pipelines are responsible for conveying water from the reservoirs to the several connection points with ramifications. This way, each stretch of the main pipeline may have a reservoir and a connection point or two connection points at its ends.

The main connection is the link between the two networks, controlling the connection between a main pipeline (larger diameter) and the network ramifications (smaller diameter). Each stretch of the ramification also has two more connection types, one for connection between ramifications and another one for connection between ramifications and building ramifications, known as water-taking device. The building ramification conveys water from the public network to the consumer's water meter (hydrometer).

When nodes and arcs represent spatial objects, a network is called spatially embedded. Examples of this type of network include: road, electric power, water and gas networks (Kösters et al., 1997). A counter-example is the construction of a network over adjacency relationships between neighbourhoods of a city.

The concept of multiple inheritances in diagram classes was chosen to elucidate the relationship between the elements of a network and the representation of spatial objects. To generate this multiple inheritances cases, a class should be a specialization of two order classes, for example, *GeographicObject* and *NetworkObject*, and also, it should associate the possible spatial representations of both views. The replacement of these specializations and associations are carried out by using generalization and representation stereotypes. It is still possible the use of multiple spatial representations for objects view.

Figure 7 shows the UML-GeoFrame class diagram designed for the water distribution system in which most classes have four stereotypes. The *TreatmentPlant* class, e.g., plays the role of node in the main network with polygonal representation as geographic object. The use of four spatial stereotypes, in spite of simplifying diagram visualization, did not avoid its overloading.

To prevent this situation, the UML-GeoFrame has an important feature: the generalization stereotypes can be left with no representation since stereotypes of spatial representation for field, objects and network views are disjoint groups. So, a class can use only stereotypes of spatial representation. Superclasses also assist designers and users since they allow the identification of classes sharing common features including generalization and spatial representation.

The Figure 7 also shows that constructors for aggregation and composition allow the definition of the network(s) that the classes of type node and arc belong to. Furthermore, the relationships between arcs and nodes reduce the ambiguity during diagram interpretation. The *MainConnection* class, for example, aggregates to the Main and Secondary networks, and relates to the *Ramification* and *MainPipelineStretch* classes, connecting these two networks.

Finally, Figure 7 also shows the use of stereotypes for representation of unidirectional arcs. Using this stereotype, it is possible to verify that both networks in the example have a flow for water distribution. However, it is not necessary the presentation of some details such as initial and final nodes. These issues are considered only in the implementation phase, since the important thing modeling this type of application is to determine the type of network and its nodes and arcs.

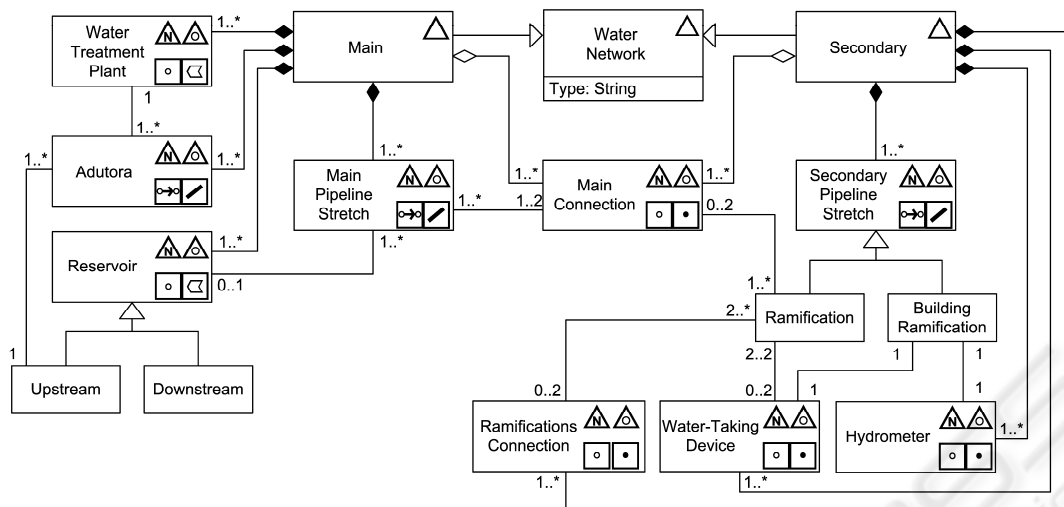


Figure 7: Example of class diagram of a water distribution system.

5 CONCLUSIONS

This work presented the extension of the UML-GeoFrame data model for conceptual modeling of network elements. The proposed solution was based on a set of specific requirements for these applications identified by Kösters et al. (1997), which include: (a) to show the incidence between arcs and nodes; (b) to show which networks the arcs and nodes belong to; (c) to make clear the class type of the arcs, nodes and network without identifying this types by class name; (d) to identify the classes with common features (superclass); (e) to make clear the class that represents the connection between two or more networks.

Thus, comparing the solution proposed in this article with the solution presented by the GeoOOA model, it is possible to say that the UML-GeoFrame, besides the good features listed above, it allows the representation of specific associations that can occur between elements of a network represented by nodes and arcs. As well as, it facilitates the aggregation of these elements with the composed object, which is the own network.

To conclude, we believe that the conceptual data schema can be read and interpreted with no ambiguities even by non-specialist users and, at the same time, it gives a larger freedom to specify accurately the requirements of the application.

ACKNOWLEDGEMENTS

This project is partially financed by CAPES, FAPEMIG and CNPq/MCT/CT-INFO

REFERENCES

Barros, Raphael Tobias de V. et al. 1995. Saneamento. Manual de saneamento e proteção ambiental para os municípios, UFMG. (In Portuguese).

Borges, K. A. V.; Davis Jr., C. A. and Laender, A. H. F. 2001. OMT-G: An object-oriented data model for geographic applications. *GeoInformatica*, Vol. 5, No 3, pp. 221-260.

Demirel, H. 2002. An Integrated Approach to the Conceptual Data Modeling of an Entire Highway Agency Geographic Information System (GIS). Library of Berlin Technical University.

Friis-christensen, A., Tryfona, N. and Jensen, C. S. 2001. Requirements and Research Issues in Geographic Data Modeling. In *Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems*, pp. 2-8.

Kösters, G. et al. 1997. GIS-Application Development with GeoOOA. *International Journal of Geographical Information Science*, Vol. 11, No 4.

Lisboa Filho, J. and Iochpe, C. 2008. Modeling with a UML Profile. In *Encyclopedia of GIS*, Eds. Shekhar, S. and Xiong, H., Springer-Verlag, pp. 691-700.

Lisboa Filho, J. and Iochpe, C. 1999. Specifying Analysis Patterns for Geographic Databases on the Basis of a Conceptual Framework. In *Proceedings of the 7th ACM GIS*, pp. 7-13.

OMG - Object Management Group 2006. Object Constraint Language. Needham.

SAAE Viçosa - Serviço Autônomo de Água e Esgoto do Município de Viçosa. 2008. (In Portuguese).

Spaccapietra, S.; Parent, C. and Zimányi, E. 2006. The MADS Data Model - Concepts to Understand the Structure of your Spatial and Temporal Data. In *International Workshop on Informative Modeling for the Architectural Heritage*.