An Extended Entity-Relationship Model for Geographic Applications [*]

Thanasis Hadzilacos
Computer Technology Institute, University of Patras
Box 1122, GR-26110, Patras, Greece
thh@cti.gr

Nectaria Tryfona
National Center for Geographic Information and Analysis
Univ. of Maine, Orono, ME 04469-5711, U.S.A
nectaria@spatial.maine.edu

Abstract. A special-purpose extension of the Entity-Relationship model for the needs of conceptual modeling of geographic applications, called the Geo-ER Model, is presented. Handling properties associated to objects not because of the objects' nature but because of the *objects' position*, calls for dealing -at the semantic modeling level- with space, location and dimensionality of objects, spatial relationships, space-depending attributes, and scale and generalization of representations. In order to accomplish this in the framework of ER and its derivatives, we introduce special entity sets, relationships, and add new constructs. The rationale as well as examples of usage of the Geo-ER model from actual projects are presented.

1. Introduction

Is everybody special or are we all alike? Should we develop applications according to a special methodology for each class of applications, such as medical, business process and geographic, or should we use a single blanket approach for all? Personal preferences and philosophical discussions aside, it does seem that the general purpose side has won most battles in computer science, from hardware to programming languages to relational data models. On the other hand, successful special-purpose tools do exist, from signal processors to web browsers, and in any special case the decision is an engineering one, balancing the additional productivity stemmed from the special crafted tool versus the extra effort of developing, disseminating and learning how to use one.

One successful method for handling the general vs. special dilemma, is the layered approach: use a general purpose engine on top of which a special purpose tool is built. Relational DBMS's have been a popular such general engine on top of which a large variety of animals breeds. In our case the question is whether geographic applications have sufficient peculiarities to warrantee a special approach in conceptual designing applications, and, if so, what is this approach. The position of this paper is that while indeed, general-purpose conceptual models are not adequate for geographic applications, what is in fact needed is *additional*, not substitute; thus it can be incorporated on existing models, such as the Entity-Relationship (ER).

The ER model [4], simple to understand yet powerful for modeling, is the most widespread and time-persistent of all conceptual models; with innumerable extensions has gained the acceptance of developers, users and software companies. While our proposal does not depend on ER specifically, and the extensions proposed can work with other semantic and object-oriented models (see [18] [20]), spelling out the details for ER is significant in showing the generality and practicality of our approach.

This work builds on (a) classic and recent research on conceptual modeling by using ER [5][14][16], (b) the study of spatial needs [8][10][22], and (c) previous and theoretical research on the subject at CTI [17][18][19]. Its contribution starts with the identification of the particular needs of geographic applications which are not adequately served by ER (Section 2). The main goal is the satisfaction of these requirements without resorting to a fundamentally different model, thus ensuring reusability in software and training. Section 3 contains the central technical contribution, which is Geo-ER, a semantic model suitable for geographic applications, based on the fundamental concepts of entities and relationships. Finally, Section 4 contains a substantial practical example taken from a real project [21].

Previous works related to the topic of geographic modeling by using ER take two main approaches: (a) some either use ER [14] or its extensions proposed for business applications to capture spatial semantics [7] and (b) others extend ER based on the study of some spatial peculiarities [16]. None of these proved to be mature enough to represent information of space: the former are not semantically rich for spatial needs, while the latest are driven by the philosophy of the target software, involving concepts of physical design (see for example [16] for the use of specific geometric characteristics supported by the Arc/Info model). Additionally, the fact that none of all these proposals have been adopted and used extensively by others shows that the problem has not yet been solved and a deep understanding of semantics of space has to precede.

2. What is special about Geographic Conceptual Modeling?

Modeling geographic applications entails tackling of specific and unique problems [4][8][10]. These include the duality of objects and fields, reasoning with incomplete information in several forms such as indeterminate boundaries, fuzzy geographical phenomena, imprecise measurements, scaling and map generalization, spatial relationships, and technical issues such as raster vs. vector choice. All these special needs bring up the question as to whether modeling at a highly abstract level of representation (i.e., conceptual), calls for additional fundamental semantic constructs. The answer must come from a balancing design decision: a model with many

---

[*] Part of this work was supported by a scholarship from C.T.I and grant from the European Union Research Program UtilNets Brite/Euram 7120 while the second author was affiliated with C.T.I.

"special" constructs allows for closer representation of the world but is more difficult to use.

Our thesis is that for the conceptual modeling of geographic applications we only need two additional concepts: *position* of geographic objects and *space-depending attributes*. This corresponds to the object/field dichotomy, which has counterparts in physics and philosophy and is most likely based on human cognition which "appears to make use of both the object and field views, but at different geographic scales and for different purposes" [4]. For the conceptual data modeling of geographic applications the designer should have at his disposal both views; by combining them we can model *space-depending attributes of geographic objects*.

This proposal is fueled from the fact that studying space-depending properties of geographic objects, such as "temperature" or "soil_type", means dealing with properties whose value depends on objects' position in space; so we talk about properties of space, and entities residing at some place inherit values for those properties from such *spatial fields*. Therefore, primarily there is the need to study and model *space*.

2.1 Modeling space

Space is a set. The elements of space are called *points*. Any set will do for space, and very important intuition and interesting theories come up from non-standard spaces, even non-numeric ones; however for practical purposes of current spatial applications, space is modeled as a subset of R^3. The following are the most common subsets used in practice: finite subsets, Z^3, Z^2, Z, R^2, and R. All specific discussions and examples in this paper use R^2 as space.

2.1.1 The concept of *space-depending attribute*

A fundamental peculiarity of spatial information systems is that some properties of interest do not properly belong to any particular object. For example, "soil_resistivity" in a utility application [21]. Although one application view may regard the "resistivity" of the soil around pipes as an attribute of the pipe, it is clear that: (a) it is defined whether or not a pipe exists at that point in space, and (b) when a pipe is moved, it will not keep the value of "its" attributes; rather it inherits new values from the new positions.

Informally, space-depending attributes are properties of space which indirectly become properties of objects situated at some position in space. Overlapping objects share the same values for these attributes. *The value of a space-depending attribute depends on object's position only*, and not on the object itself. Formally, a *space-depending attribute* is a function whose domain is space and range is any set. Examples of such attributes are the "temperature" (a function from space to R), or "degree_of_erosion" (a function from space to the set $\{high, low, medium\}$).

2.1.2 The concept of *position*

In the real world, most objects have a position which is the object's link with space. In information systems we are only interested in position for *some* objects: those are the *geographic* objects of the application. The position of an object is any finite subset of space, also called *geometric figure*, and it includes its *location* (centroid)*, shape, size* and *orientation.* Geometric operations on objects, (such as length, area, or distance) are defined in an analogous way; for example *distance* is a function from pairs of points to Reals. *Translation, rotation, similar-shape expansion* and *reduction,* and *symmetry* are functions from geometric figures to geometric figures (i.e. transformations). Each geometric figure divides space into its *interior*, its *boundary* and its *exterior*. *Topological relationships* such as inside, touches and others, are defined between geometric figures (objects' positions) based on the content of the intersections of the interiors, boundaries and exteriors in all combinations [6].

2.2 Modeling spatial relationships

Furthermore, dealing with entities at a conceptual level means dealing with relationships among them. Describing relationships among entity sets is the key of capturing semantics. Relationships among geographic entities (objects) are actually conditions on objects' position and are called *spatial* relationships. Topological relationships are special cases of the spatial ones. Spatial relationships are translated into spatial integrity constraints of the database. The definition of a "square", in a cadastral application, as a "land parcel which is not contained in any building block" is an example of using spatial integrity constraints. Conceptual geographic models should lead to straightforward solutions for explicitly storing topology in the logical and physical levels -a common practice despite topology being derivable from object's positions [11].

2.3 Modeling different views

An important issue in geographic applications is the need for modeling different views of the same excerpt of the real world. This, does not only stem from the fact that scale changes (multi-resolution representations), but also from the requirements of a spatial database: the user wants to be able to see and refer to a *city* either as a *point* or as a *region* or both. Apart from issues dealing with assessing consistency among these different perspectives, we are also invited to integrate all different views in one single conceptual schema as well as to preserve uniformity of successive results when dealing with them; for example, how is defined the distance between two cities *A* and *B* when they captured (a) both as points, (b) *A* as a point and *B* as a region?

3. The formal Geographic Entity-Relationship Model

ER is arguably the first conceptual model that appeared in the literature. Its main advantages are the ease of use that provides and the minimum set of supported constructs. Its basic elements are the *entity sets,* the *attributes*, and the *relationships* among entity sets. Several extensions of ER have been proposed in order to capture special needs from different research areas. Amongst these, the most popular are the relationships *part_of* and *member_of* which are used to capture the meaning of *aggregation* and *grouping* respectively [5][14].

3.1 The Geographic Entity-Relationship Model

Based on the study of spatial aspects that call for special modeling techniques and constructs for the conceptual design of geographic applications we develop and present the Geographic Entity-Relationship Model (Geo-ER); it is based on the standard ER, and on the Extended ER models of [5] and [14] which integrate the concepts of aggregation and grouping. Geo-ER includes special entity sets and relationships to express the semantics of space, geographic entities' position, entities' space-depending attributes and spatial relationships. Two new constructs are added to express the spatial dimension of complex geographic entity sets: *spatial aggregation* and *spatial grouping*.

A. Specializations of ER

I. A special entity set is introduced: *SPACE*. In this paper *SPACE* is any set homomorphic to R^2.

II. Entity sets whose position in space matters are called *geographic* (depicted by bold-margin rectangles). Since ER uses entity sets and attributes thereof, we model *position* as a special entity set with fixed meaning: *POSITIONS* is a function on all and only on geographic objects and returns for each geographic object *a part of space*. In other words, *POSITIONS* is 2^{R^2}, the power set of space as *position is any subset of space*. So, the domain of *POSITIONS* is a finite subset of sets of *geometric figures*, i.e. points, lines and regions. A position is fully and non-redundantly determined by four elements [19]: *shape*, *size*, *location* (centroid) and *orientation*; these are *part_of* *POSITIONS*.

SPACE is related to *POSITIONS* by:

$$\forall p(p \in POSITIONS \Leftrightarrow p \subseteq SPACE) \qquad \textbf{(1)}$$

This approach places no restrictions on the model of space. Also, all geometric operations and topological relationships are considered to be pre-defined for the domain of *POSITIONS*.

In order to represent *shape*, the special entity sets $0-Dimensional$, $1-Dimensional$, and $2-Dimensional$ are introduced and are called *geometric types.* These sets are related to *POSITIONS* by an ISA hierarchy. $0-Dimensional$, $1-Dimensional$, and $2-Dimensional$ are (ISA) *shapes* of an entity. As the position of a complex geographic entity can be any combination of points, lines and regions in space, *shape* (as well as position itself) is determined by the higher dimension of geometric figures that constitute entity set's position (*dimensionality*). For example, a "high-tension tower" of an electricity company in a 2d map is represented by a set of discrete points (usually 3 or 4) and its shape is $0-Dimensional$.

Figures 3.1 represent entity's position in space. When only the *shape* matters, geometric types are directly connected to *POSITIONS*.

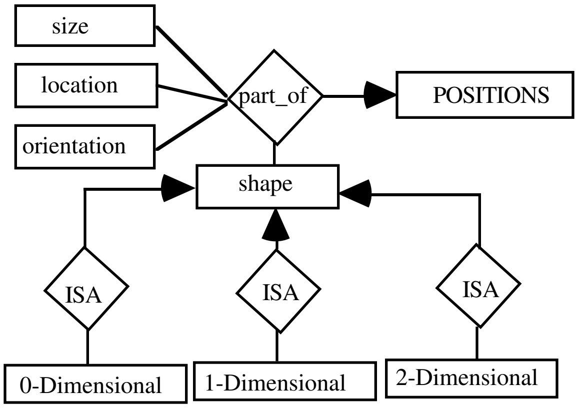


Figure 3.1: Modeling entities' position in space.

III. A special 1:M relationship *is_located_at* relates each geographic entity set to its -not necessarily unique-position in space. There are several reasons why the position of a geographic entity might be represented in more than one ways in the same application (most likely belonging to different application views). For example, a "city" may be represented as a point or a region depending on the scale of the map; going from one representation to the other is not always automatic and we may need to store both in the database. This is why we cannot restrict *is_located_at* to be a 1:1 relationship.

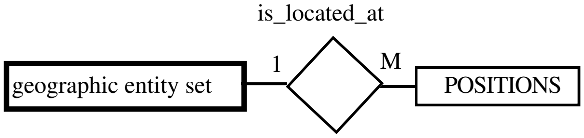
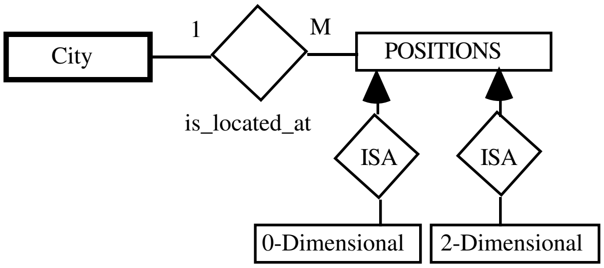
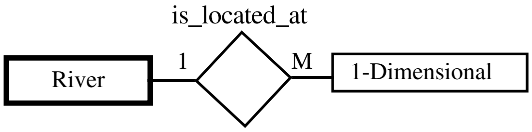


Figure 3.2: $is\_located\_at : geog\_entity\_set \rightarrow POSITIONS$.

Only the ISA entity sets of *POSITIONS* that are meaningful appear in the conceptual schema. Additionally, when a geographic entity set is represented uniquely then *POSITIONS* is omitted. Figure 3.3 illustrates different views of a "city", and a "river".



(a)



(b)

Figure 3.3: Different views of a “city”, and a “river”.

For reasons of simplicity, the relationship among an entity set, its position in space is depicted by a bold-margin rectangle (representing the geographic entity set) with a small triangle in the upper-left corner showing its dimensionality (whenever it matters). Fig. 3.4 depicts the equivalents of Fig. 3.3.

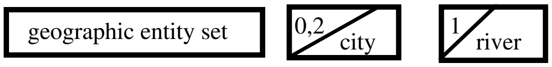


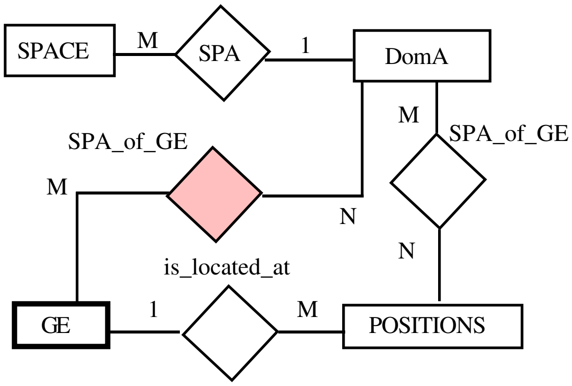
Figure 3.4: Geographic entities, their position and shape.

IV. Space-depending attributes are modeled as functions from $SPACE$ to the attribute domains.

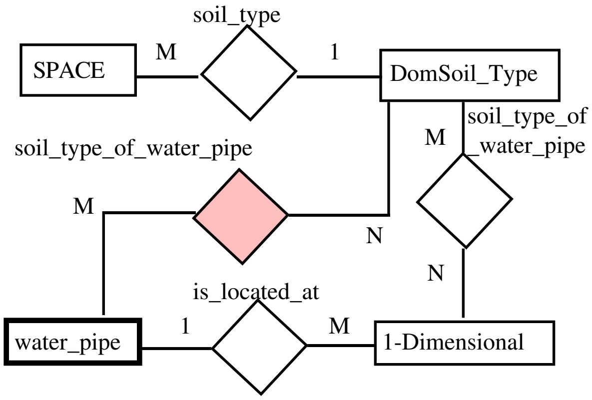
Let $SPA\colon SPACE \rightarrow DomA$ be a space-depending attribute ranging over $DomA$, and $g \in GE$ a geographic entity belonging to entity set GE with position $p \in POSITIONS$.

$$\begin{aligned} SPA(p) &= SPA \mid p = SPA \circ is\_located\_at_{GE}(g) = \\ &= SPA \mid is\_located\_at_{GE}(g) \end{aligned} \tag{2}$$

where $\mid$ and $\circ$ stand for functional restriction and composition respectively. By slightly abusing the notation (shaded diamonds in Fig. 3.5), we let $SPA(g)$ stand for $SPA \circ is\_located\_at_{GE}(g)$; in that way we extend SPA from an attribute of $SPACE$ to an attribute of entity set $POSITIONS$ and finally to an attribute of entity set GE (Fig. 3.5a).



(a)



(b)

Figure 3.5: Modeling space-depending attributes.

This is the integration of field (so called layer in commercial software) and entity (object) approaches. Thus, if SPA is “soil_type” and GE is the class of “water_pipe” then the “soil_type” of an underground “water_pipe” is “soil_type” restricted to the position of water_pipe (“soil_type_of_water_pipe”) (Figure 3.5b).

Alternatively, since SPA is an attribute of $SPACE$ (this is the reason why $DomA$ is an entity set without attributes), and object’s position belongs to it, and by introducing the 1:M relationship $belongs\_to$ between $SPACE$ and $POSITIONS$:

$$SPA(p) = belongs\_to\left(is\_located\_at_{GE}(p)\right) \tag{3}$$

Figure 3.6 illustrates this approach; in practice, the designer relates the space-depending attribute to the entity and not to space and then restricted it (the attribute) to entity’s position (represented by a shaded diamond).

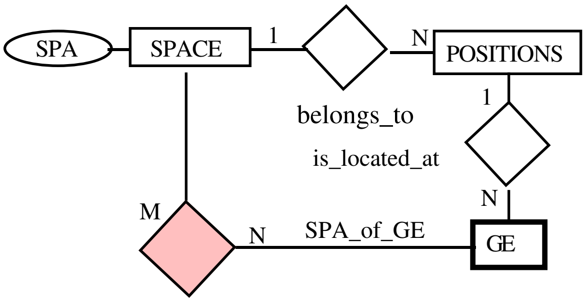


Figure 3.6: Spatial attributes as properties of space.

B. Extensions of ER

V. That “a network is an (ordered) set of network segments” differs from “a country being a set of people” in that the former grouping has a spatial dimension as well: the position of the network is the geometric union of the positions of its constituent segments -whereas nothing of the sort holds in the second case. To capture this additional semantic two new constructs are introduced:

spatial aggregation (*spatial_part_of*) and *spatial grouping* (*spatial_member_of*).

- Let the aggregation (Spatial Part) $SP = C_1 \times C_2 \times \ldots \times C_n$, be a geographic entity set, and let $C_{i1},\ldots,C_{ik}$ be its geographic parts. We say that SP is a *spatial* aggregation (graphically represented by a "$\rightarrow$"), if and only if the position of sp is the geometric union of the positions of its geographic parts:

$$\forall sp \in SP\ \forall c_i\ \left(c_i \in C_i\right)$$
$$\left(sp = \langle c_1,\ldots,c_n\rangle \Rightarrow p\left(c_{i1}\right)\cup\ldots\cup p\left(c_{ik}\right) = p(sp)\right) \tag{4}$$

where $\cup$, $\langle c_1,\ldots,c_n\rangle$ and p are the operators for geometric union, aggregation and the function $is\_located\_at$ respectively. The dimension of the resulting entity set is equal to or higher than the higher dimension of the aggregated parts.

Figure 3.7a shows the graphic representation of spatial aggregation and Figure 3.7b the example of a "city" composed by "name", "country-name" (non-geographic entity sets) and "residential area", and "industrial area":

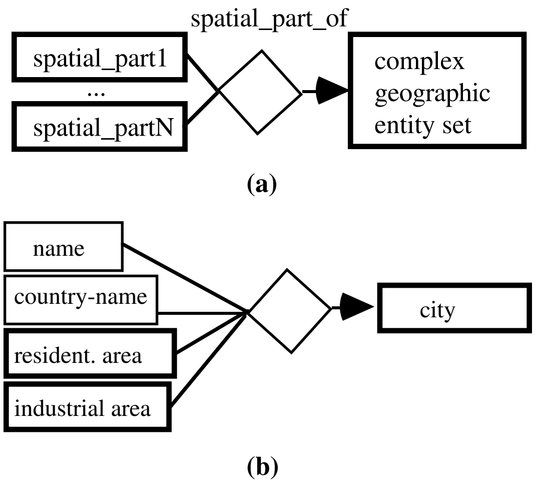


Figure 3.7: (a) spatial aggregation, (b) an example.

- Let SM (Spatial Member) be a grouping of geographic entity set C; we say it is a *spatial* grouping (graphically represented by a "$*$") if and only if for every instance of SM, say sm, the instances of C which form sm are topologically inside it:

$$\forall sm \in SM\ \forall c_i \in sm\Big(p(sm)\text{covers } p\left(c_j\right)\Big) \tag{5}$$

The dimension of the resulting entity set is equal to the dimension of the grouped members.

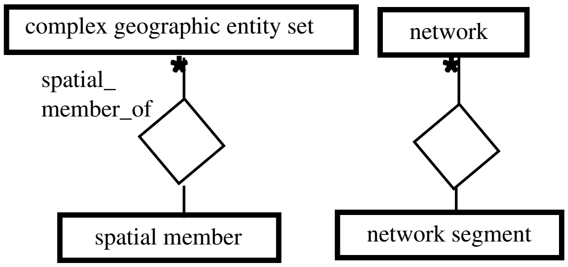


Figure 3.8: Spatial grouping: an example.

4. Example of Usage

In this section we present the level of conceptual design of an excerpt from a real application dealing with a Network Utility Management System [21] by using the Geo-ER Model. Consider the following description:

> "...A utility (underground) network is composed by links (line segments) and service reservoirs (points or regions). Its whole structure can be seen either as a linear or as a regional object. It is important to record the type of soil underground the network... Additionally, the network is related to the city it supplies with water..."

Based on the above, the *position* of the "network" is represented either by 1-Dimensional, or a 2-Dimensional entity set. *Parts* of the network are the "reservoir" and the "link". The "soil_type_of_the_underground_network" is a restriction of "soil_type" of the *space* to the position where the network exists. So, it is modeled as a space-depending attribute. Figure 4.1 illustrates this description in the Geo-ER Model.

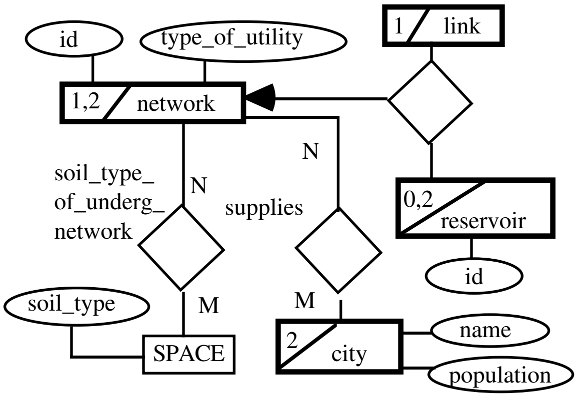


Figure 4.1: Using Geo-ER.

5. Conclusions

An extended ER model to capture spatial peculiarities at the conceptual level of geographic database design is presented. By modeling the concepts of position and space-depending attributes of geographic objects we provide an elegant way of dealing with the issues associated to conceptual modeling of spatial applications (details can be found in [19]):

- The field (layer) vs. object dichotomy corresponds to space-depending vs. geographic entity sets and the raster vs. vector choice to modeling space as a set homomorphic to Z^2 or R^2.

- Spatial relationships are reduced to algebraic or geometric integrity constraints on the entity set *POSITIONS* [11].
- Attribute generalization -an important aspect of scaling and map generalization- is achieved through algebraic transformations of the functions which represent space-depending attributes.
- Fuzzy points specified using pairs of probability distributions suffice for modeling uncertainty and fuzziness and fuzzy geometric figures in the entity set *POSITIONS* [10].

This proposal has been used for the design and development of an actual project [21] and small commercial systems. We have experienced a welcome acceptance from both users -who found it easy to understand- and developers -who were happy to adopt it.

Acknowledgments

The authors wish to thank Nikolina Renieri for the fruitful discussions on the concept of entities' position in space.

6. References

1. Abiteboul, S., and Hull, R., 1987, *IFO: A Formal Semantic Database Model*, ACM TODS, 4, pp. 525-565.

2. Abiteboul, S., Hull, R., and Vianu, V., 1995, *Foundations of Databases*, Addison-Wesley.

3. Chen, P.S., 1976, *The Entity-Relationship Model: Toward a unified view of Data*, ACM TODS, 1, pp. 9-36.

4. Couclelis, H., 1992, *People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate*, GIS. LNCS, 639.

5. Dijkstra, J., 1993, *On Complex Objects and Versioning in Complex Environments*, Proc. of ER'93, Texas, USA.

6. Egenhofer, M., and Herring, J., 1995, *A Mathematical Framework for the Definition of Topological Relationships*, Proc. of the 4th Int. Symp. on SDH, Switzerland.

7. Fernardez, R.N. and Rusinkiewicz M., 1993, *A Conceptual Design of a Soil Database for a Geographic Information System*, IJGIS (7, 6).

8. Hadzilacos, Th., 1995, *On Layer Based Systems for "Undetermined Boundaries"*, in "Objects with Undetermined Boundaries", Burrough and Frank (eds).

9. Hadzilacos, Th., and Stoumbou, P.M., 1995, *Conceptual Data Modeling for Prehestoric Excavation Documentation*, Proc. of the Int. Conference on Computer Applications in Archaeology Leiden, Holland.

10. Hadzilacos, Th., and Tryfona, N., 1994, *Logical Data Modeling for Geographic Databases*, IJGIS (10,2).

11. Hadzilacos, Th., and Tryfona, N., 1992, *A Model for Expressing Topological Integrity Constraints in Geographic Databases*. LNCS, 639, pp. 252-268.

12. Hull, R., and King, R., 1987, *Semantic Data Modeling: Survey, Applications and Research Issues*, ACM Computing Surveys, (19,3).

13. Kim, W., 1990, *Introduction to Object-Oriented Databases*, Cambridge, MA:MIT Press.

14. Jaescje, P., Oberweis, A., and Stucky, W., 1994, *Deriving Complex Structured Object Types for Business Process Modeling*, Proc. of ER'94, Manchester, UK.

15. Mylopoulos, J., 1991, *Conceptual Modeling and Telos*, DKBS-TR-91-3, University of Toronto.

16. Pelagatti, G., Belussi, A., Negri, M., Rossi, M., and Sbatella, L., 1991, *Design and Implementation of a Geographic Information System: The case of Regione Lombardia*, Proc. of the 2nd EGIS, Brussels, Belgium, April, 2, pp. 856-865.

17. Tryfona, N., 1994, *A Structured Database Methodology for Geographic Applications*, Ph.D Thesis, Department of Computer Engineering and Informatics, University of Patras.

18. Tryfona N., and Hadzilacos, Th., 1995, *Geographic Applications Development: Models and Tools at the Conceptual Leve*l, 3rd ACM Workshop in GIS, MD, USA.

19. Tryfona N., and Hadzilacos, Th., 1995, *Conceptual Data Modeling for Geographic Applications: Desiderata and Solutions*. TR-95.02.8, C.T.I., Patras, Greece.

20. Tryfona N., Pfoser, D., and Hadzilacos, Th., 1997, Modeling Behavior of Geographic Objects: An Experience with the Object Modeling Technique, Proc. of the 9th International Coneference on Advanced Information Systems and Engineering (CAiSE 1997), Barcelona, Spain, LNCS.

21. UtilNets, 1994, *Work Program of Brite-Euram Project 7120*, DG XII, European Union, Brussels, Belgium.

22. Worboys, M., 1994, *Object-Oriented Approaches to Geo-referenced Information*, IJGIS, 8,4, pp. 385-399.