A Structured Methodology for Interoperable Geographic Applications: The Case of the Hellenic Cadastre [1]

Nectaria Tryfona [2]
Department of Computer Science,
Aalborg University, Fredrik Bajers Vej 7E
DK-9220 Aalborg ∅st, DENMARK
tel: 45 96 358898 fax: 45 98 159889
tryfona@cs.auc.dk

Marinos Kavouras
Faculty of Rural and Surveying Engineering
National Technical University of Athens
9, H. Polytechniou Street, Zographou Campus
GR-15780 Athens, GREECE
tel: 30 1 7722731 fax: 30 1 7722670
mkav@survey.ntua.gr

Thanasis Hadzilacos
Computer Technology Institute
University of Patras
BOX 1122, GR-26110 Patras, GREECE
tel: 30 61 274496 fax: 30 61 222086
thh@cti.gr

Keywords: database design, conceptual modeling, logical modeling, interoperability, spatiotemporal data, data exchange, semantics exchange.

Acknowledgments

[1] Related to the subject of another research project of the National Technical University of Athens, funded by KTIMATOLOGIO SA, Greece.

[2] Supported in part by the Danish Technical Research Council through grant 9700780, the Danish Natural Science Research Council through grant 9400911, and the CHOROCHRONOS project, funded by the European Commission DG XII Science, Research and Development, as a Network Activity of the Training and Mobility of Researchers Programme, contract no. FMRX-CT96-0056.

A Structured Methodology for Interoperable Geographic Applications: The Case of the Hellenic Cadastre [1]

Abstract

The research reported in this paper uses as a test case an on-going project: the design and development of the cadastral system in Greece. The main focus is on the database design, specifically at the stages of conceptual and logical modeling. Issues such as modeling the intricate nature of space, as well as the need to exchange not only data formats but, also semantics captured by computer systems are faced. At the conceptual level we represent the ontologies of the application and their semantics, while at the logical level we provide the language to exchange them. Finally, commercial tools as well as models resulted from theoretical research are combined and tested for their applicability.

1. Introduction

Dealing with geographic applications has been a challenge for different and, sometimes, diverse research communities for many years. Theoreticians, on the one hand, provide solutions to issues such as complex data modeling problems and efficient algorithms, while on the other, practitioners go on with their own answers, mostly driven by an emerging need to build and support geographic applications. Nevertheless, Geographic Information Systems (GIS) is the meeting point of different but essential disciplines. Their successful combination is the real challenge.

This paper presents some of the decisions made and results achieved in parallel to another recently completed project: the definition of a standard for the incorporation of existing and new data in the Cadastral System of Greece (Kavouras 1997). The project, which involved scientists with different backgrounds, such as surveying engineers, cartographers and computer scientists, had very interesting results, which are in the process of being widely published. Furthermore, the objective of the project was related to new research directions, such as modeling of spatiotemporal data and processes, and information aspects of interoperability (exchange of data and their semantics) which are the subject of this paper.

The rest of the paper is organized as follows: Section 2 outlines some technical issues related to the Hellenic Cadastre Project. Section 3 describes the phase of database design of the cadastral system. The conceptual and logical modeling stages are further discussed, showing models, tools and techniques used for this purpose. An example taken from the cadastral application is designed in both stages. Section 4 deals with interoperability issues. More particular, it explains how the model used for the logical design stage can be used to exchange data and semantics among remote applications. Finally, Section 5 summarizes this work and gives the future plans.

2. The Hellenic Cadastre Project – Some Technical Issues

The Hellenic Cadastre is a very large and ambitious project with the objective to complete the non-existent cadastre of the entire Greek territory in a period of 15 years. The project is in its first phase with many pilot studies underway. There is an advantage of starting such a project now - the fact that, in many ways, the information system can be designed from scratch with the most modern and fresh views about property systems and without constraints of the past. The IT System's objectives as quoted by the Hellenic Cadastre Consult (Moropoulos 1997) are:

- Convenient efficient and effective access to the Cadastre for all citizens o Greece.
- A system that enables land dealings to be registered in an agreed acceptable time frame and at an acceptable cost.
- A database of reliable land information that is current, complete, accurate and recognised legally.
- Ready access to land information without the infringement of individual's rights.
- Cost effective maintenance of the database.
- Acceptance and understanding by the citizens of Greece of the operations and benefits of the Hellenic Cadastre.
- A business approach to the operations of the Cadastre, and the optimization of its commercial potential to the best interests of Greece and its citizens.

Three phases of the Hellenic Cadastre are envisaged:

- *Data Capture and evaluation*
- *Interim Cadastre establishment*
- *Hellenic Cadastre operation*

A pressing issue of the Cadastral Project is the development of an appropriate database on one hand, and of data specifications for the standardization and exchange of the undergoing data capture operations on the other. The database issue is central to this paper. The issue of standardization and data exchange has been tackled in another related but parallel research project (Kavouras 1997).

It must be also mentioned that the results achieved and presented here constitute a research viewpoint of the cadastral database problem. As it is always the case in a production environment, these results may differ from the actual future design and implementation of the cadastral database, since the Hellenic Cadastre may take into consideration other non-technical parameters/constraints (i.e., political, practical or economic).

3. Database design of the cadastral system

Our primal focus is the database design for the development of the cadastral system. It includes a set of engineering activities leading from the problem specification to the final system implementation. It is subdivided into several stages, which include:

- *Requirements analysis*.

At this stage the user describes in natural language the application domain as well as his/her requirements of the final system. Our input here came from a set of already existing electronic data, as well as discussions with the surveyors who know exactly the domain.

- *Conceptual modeling*.

It is an abstract representation of the application, focusing on data semantics. The Entity-Relationship model (ER) is the most known model used at this level. Section 3.1 describes this stage.

- *Logical modeling*.

It is closer to the implementation level, but still independent of specific characteristics of the target system; the relational model (Chen 1976) is often used here. Section 3.2 describes this stage.

- *Physical modeling and implementation.*

It is based on the requirements and techniques provided by the specific software.

The above described methodology is well-known and -for years- widely used in the computer science community. The main benefits of using it in the application development cycle are the following (Brodie 1982):

- *Portability*: since the design is platform- (and finally, vendor-) independent, whenever there is a requirement to change the software platform of the cadastral database system this can be done without major changes.
- *Expandability*: the design can be extended or modified independently of the first team of designers.
- *Correctness*: understanding the semantics and the special role of every component of the geographic database and representing them at a high level of abstraction lowers the risk of misunderstanding and misinterpreting user requirements.

These factors later translate into less time and cost, when dealing with these systems.

3.1 The Conceptual Design of the Cadastral Database

The main goal is to model the cadastral information (objects and procedures about land) in a way which (Tryfona and Hadzilacos 1995):

(a) uses no computer metaphors (such as record, process or relation) to represent the static and dynamic properties of the objects needed,

(b) is understandable to the user, i.e. the non-computer scientist person who has knowledge of the application domain, and

(c) is formal and complete, so that it can unambiguously and with no other user input be transformed into the logical data model.

The model used here is the GeoER (Hadzilacos and Tryfona 1997), an extension of the Entity Relationship (Chen 1976) to handle spatial peculiarities. GeoER includes special entity sets and relationships to express the *semantics of space*, *geographic entities' position*, entities' *space-varying attributes* and *spatial relationships*. Two new constructs have been added to express the spatial dimension of complex geographic entity sets: *spatial aggregation* and *spatial grouping*. Issues of time are also captured.

Figure 1 gives an expert of the cadastral system in GeoER. In this, terminology has been simplified for the sake of readability. Shaded components are parts of the GeoER model; not all of them are represented here.

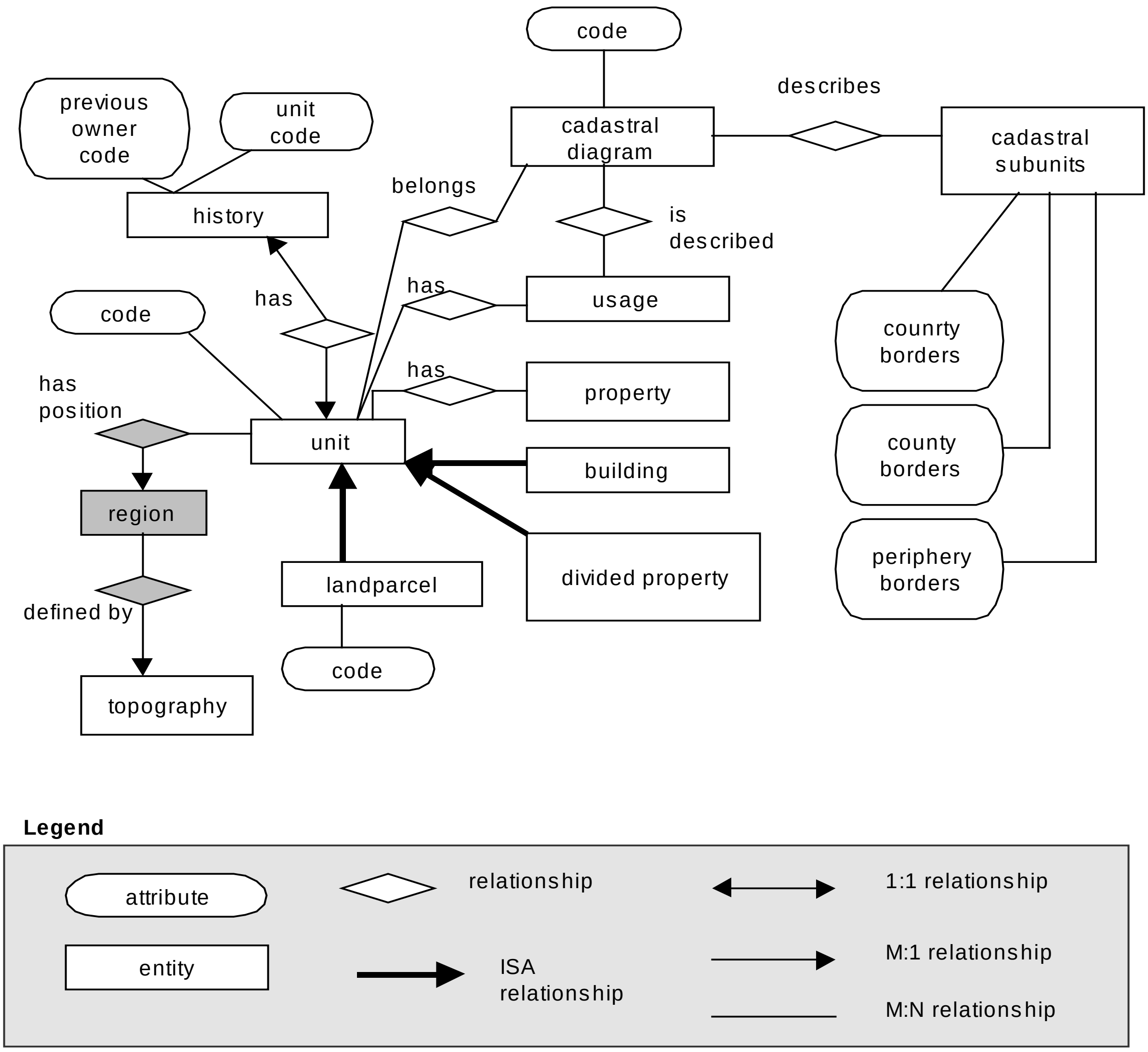

Figure 1: An excerpt of the conceptual design phase of the cadastral system in GeoER

3.2 The Logical Design of the Cadastral Database

The next step of the methodology involves the translation of the cadastral database schema to the logical one. The process is called logical design and is the intermediate stage from the semantic or conceptual schema to implementation. Logical data modeling leads to two schemata:

- the logical schema, which includes the definitions of *objects* (or entities), their *attributes* and their *relationships*, and
- the external schema, which is the integration of all user views.

For the logical design of our application we use the GeoRelational Data Model (GRDM) presented at (Hadzilacos and Tryfona 1996) (Tryfona and Hadzilacos 1998). GRDM is an extension of the relational, based on the spatial and temporal needs at this stage of design. It provides a language for the definition of:

(a) *relations*, used for non-spatial entities and relationships,

(b) *layers*, which represent space-varying attributes,

(c) *object classes*, which represent geographical entities that have characteristics from more than one layer, and

(d) *constraints* among objects or layers, which are used for topological and other spatial relationships. Objects and layers are combined in an orthogonal way, spatial constraints among objects are translated into spatial relationships, while relations among layers are obtained by using the layer algebra (Delis et. al. 1995).

Figure 2 represents portion of the GRDM syntax while Figure 3 represents our example in the GRDM.

DEFINE RELATION rel_name
(attr_name$_1$, Domain$_1$, <KEY>),…,(attr_name$_k$, Domain$_k$, <KEY>)
<[START_V_TIME DATE,STOP_V_TIME DATE]>,
<[START_T_TIME DATE, STOP_T_TIME DATE]>

DEFINE LAYER n <layer_name>
<[START_V_TIME DATE,STOP_V_TIME DATE]>,
<[START_T_TIME DATE, STOP_T_TIME DATE]>
ATTR (attr_name$_1^n$, Domain$_1^n$, <UNIQUE>),…,
(attr_name$_{mn}^n$, Domain$_{mn}^n$, <UNIQUE>)
GEOMETRIC TYPE Geometric_type
<POSITIONING Coordinate_system>
<CONSTRAINT spatiotemporal_condition>

DEFINE OBJECT CLASS obj_class_name
<GEOMETRIC TYPE Geometric_type>
<SUBTYPE OF sup_obj_class>
<ON LAYERS layer_id$_1$,…,layer_id$_k$>
<WITH ATTRIBUTES attr_name$_1$,…,attr_name$_m$>
<CONSTRAINT constraint_name>

DEFINE CONSTRAINT constraint_name
ON {object_class_name | layer_id}
AMONG {obj_class$_1$,…,obj_class$_k$ | layer_id$_1$,…,layer_id$_k$}
AS spatiotemporal_condition_specification

Figure 2: An excerpt of the formal syntax of the GRDM Data Definition Language

4. A Model for Interoperating Systems

4.1 The rationale

During this project, we address the issue of exchanging semantics and data by using the GRDM model. An interoperable (or interoperating) system is one that allows remote systems seamless and transparent access to its functionality and data (Tryfona and Sharma 1996). It is a distributed yet unified environment, in which all supported models and processes are compatible. The heterogeneous remote systems are autonomous and “logically” connected providing meaningful exchange of data.

```
DEFINE LAYER 1 Cadastral_Diagram_Region
                                        /* a cadastral diagram consists of point, lines, regions */
ATTR area-id INT
GEOMETRIC TYPE REGION

DEFINE LAYER 2 Cadastral_Diagram_Line
ATTR line-id INT
GEOMETRIC TYPE LINE

DEFINE LAYER 3 Cadastral_Diagram_Point
ATTR Point-id INT
GEOMETRIC TYPE POINT

DEFINE LAYER 4 Landparcels              /* a landparcel is a unit */
ATTR lp-id INT
GEOMETRIC TYPE REGION
```

Figure 3: Part of the example (Figure 1) translated into GRDM

Providing a formality for the representation of applications, aids capturing semantics and handling important aspects of real-world entities represented in an information system (Sheth 1995). Moroever, because semantics play an important role in geographic applications (for example, "same" data may have different semantics and the opposite), it is critical to be able to exchange not only spatial data formats, but also semantics of the geographic application.

The conceptual design stage of the methodology in Section 3 allows us to reveal most of the semantics of the cadastral information.

On the other hand, based on experience gained from a specific application development (UtilNets 1994), as well as from general studies (Couclelis 1992), (Camara et. al. 1992), we came to the result that the modeling framework for the semantics must at least handle:

- the field- and object-based views of geographic space, and
- spatial relationships or constraints.

GRDM supports both elements; and therefore, was adopted to make systems interoperable. Figure 4 illustrates the rationale of interoperable cadastral systems from the database modeling viewpoint. Each remote system provides an interface for the spatial application built on top of a GIS. Each cadastral application communicates with the local data by using a local DataBase Management System (DBMS) and with remote data and processes through the GISs. GISs communicate via GRDM by providing a mechanism to translate their particular model to GRDM and vice versa (Tryfona and Sharma 1996).

It is important however, to notice that here we deal with semantics that can be captured by computer systems, and not with those depending on different interpretations based on linguistics, human behavior etc.

GRDM acts as an abstract level communication protocol among the remote systems. It is responsible for representing (a) user views, and (b) objects and operations, formally, without ambiguities, in a way understandable, adaptable and transferable to models that remote systems use. Its data definition and syntactic rules of an Object-Oriented specification, facilitates the homogenization of different representations of semantically equivalent data in remote systems at different levels of detail.

We do not invite systems at remote locations to use our model. Each remote system (application) (Figure 4) may use a different model and language. Since GRDM captures all

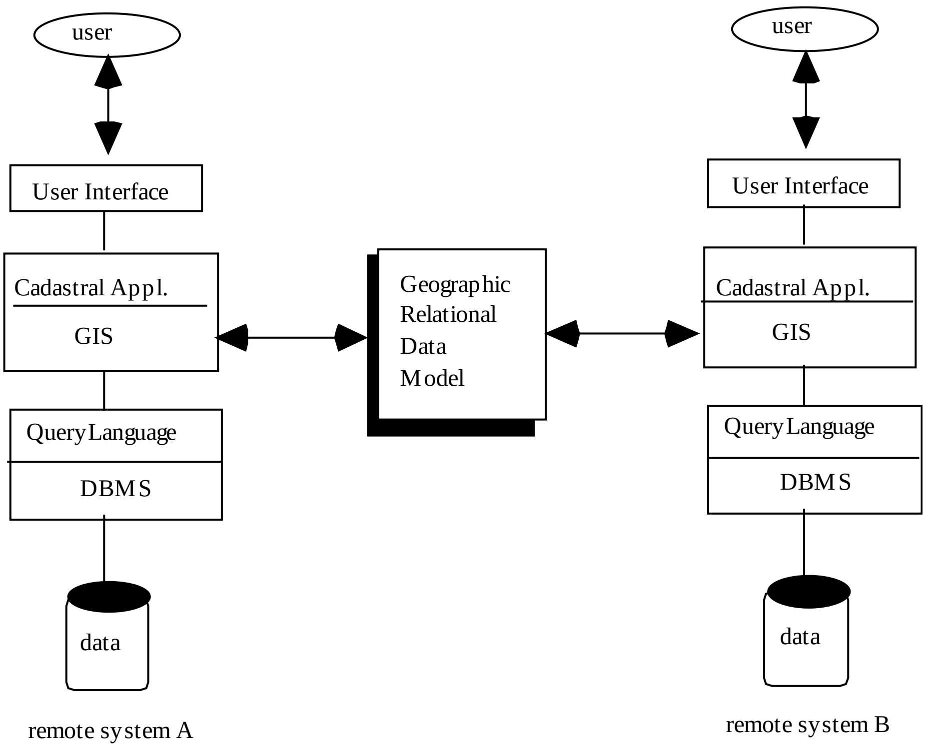

Figure 4: Interoperating Cadastral Databases

the possible modeling needs of geographic information, the models are just different syntactic versions of the same underlying spatial concepts encapsulated by GRDM; hence, they can be used interchangeably.

4.2 Example of Usage

We present a sample usage of the GRDM to show how objects, layers and spatial relationships can be combined to express excerpts of geographic information. The language of GRDM can be used to exchange the information. The basic goal - except from transferring the data (maps, records, etc.) - is to distinguish and transfer objects, relationships among them, space-varying attributes (i.e., fields, layers) and their semantics in an understandable way from one location to the other.

Consider the following scenario: An application "running" on remote system A "asks" for a portion of the cadastral application located in remote system B using the statements: "*identify sites suitable for a school. Good candidates must be within 1.25 Km (accessible) and located in regions designated as "residential"*.

This application requires the creation of a zone at the specified distance from motorways, which must be overlaid with sites that are residential. The result is a (virtual) layer representing candidate sites for a school. The result is derived from the overlay of two layers: one representing accessible but not noisy sites, and one representing the land use of our area of interest.

- DEFINE LAYER 1 LANDUSE
  ATTR (USAGE, STRING)
  GEOMETRIC TYPE REGION

- DEFINE LAYER 2 ROADS
  ATTR (ROAD_TYPE, STRING),
  (WIDTH, REAL)
  GEOMETRIC TYPE LINE
- DEFINE VIRTUAL_LAYER 3 AS COMPUTE SPATIAL
  (2, BUFF_ZONE_≤_1.25KM=BUFFER (1.25, ROAD_TYPE="MOTORWAY"))
- DEFINE VIRTUAL_LAYER 4 AS COMPUTE ATTRS
  (3,ZONE = $\begin{cases} true, if \text{ BUFF\_ZONE} \le 1.25Km = true \\ false, otherwise \end{cases}$)
- DEFINE VIRTUAL_LAYER 5 AS COMPUTE ATTRS
  (1,CANDIDATE_SITES = $\begin{cases} true, if \text{ USAGE ="Residential"} \\ false, otherwise \end{cases}$
- DEFINE VIRTUAL_LAYER 6 AS RECLASS OF (5, CANDIDATE_SITES)
- DEFINE VIRTUAL_LAYER 7 AS OVERLAY(4,6)
- DEFINE VIRTUAL_LAYER 8 AS COMPUTE ATTRS
  (7,PARK_SITES= $\begin{cases} true, if \text{ CANDIDATE\_SITES} = true \wedge \text{ZONE} = true \\ false, otherwise \end{cases}$)
- DEFINE OBJECT CLASS SCHOOL_SITE ON LAYER 8

Since remote systems “understand” GRDM, they can exchange objects, properties, layers, operations on layers and spatial integrity constraints.

5. Conclusions and Further Work

We present decisions made and results achieved during an on-going project: the design and development of the cadastral system in Greece. The main focus is on database design. We mainly deal with two stages: the conceptual and the logical. At the conceptual stage we use the GeoER Model to represent the semantics of the intricate nature of space. At the logical, we use the GRDM -an extension of the, mathematically, sound relational model. Both are based on the traditional and well-known techniques and extended to capture the spatial and temporal dimension of data. We also deal with the interoperability issue. At the conceptual level we capture the ontologies of a cadastral system and their semantics, while at the logical level we provide the language to exchange them. Our scope is to exchange not only data formats but also semantics.

Further plans include the full support of temporal aspects in both stages of design. Issues such as history of not only descriptive but also geometric information is a real challenge, as the volume of data that needs to be captured, stored and exchanged is enormous.

6. References

BRODIE, M.L., 1982, On the development of data models, in On Conceptual Modeling, Brodie, M.L., Mylopoulos, J. and Schmidt, W.J. (editors), (New York: Springer-Verlag).

KAVOURAS, M. (ed.), 1997, Definition of a standard for the exchange of digital cadastral data. Technical Report of a Research project of the National Technical University of Athens funded by KTIMATOLOGIO SA, Greece (in Greek).

CAMARA, G., FERITAS, U.M., SOUZA, R.C.M., and CASANOVA, M.A., 1992, SPRING: Procassemanto de Imagens e Dados Georefenciados. Proc. of V Brazilian Symp. on Computer Graphics and Image Processing (SIBGRAPI 92), pp. 233-242. (translation).

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

COUCLELIS, H., 1992, People Manipulate Objects (but cultivate fields): Beyond the Raster-Vector Debate in GIS. Lecture Notes in Computer Science, Springer-Verlag (639), Proceedings of the International Conference GIS: From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Pisa, Italy, September 1992.

DELIS, V., HADZILACOS, Th., and TRYFONA, N. 1994, An Introduction to Layer Algebra. Proceedings of the 6th International Symposium on Spatial Data Handling, Edinbrough, UK, Taylor and Francis (1995).

HADZILACOS, Th., and TRYFONA, N., 1996, Logical Data Modeling for Geographic Applications. International Journal of Geographic Information Systems 11,1.

HADZILACOS, Th., and TRYFONA, N., 1997. An Extended the Entity-Relational Model for Geographic Applications. ACM SIGMOD Record 26(3).

MOROPOULOS, N. 1997. The Hellenic Cadastre Information System. Internationl Federation of Surveyors - FIG Commission 3, International Seminar in GIS/LIS. Thessaloniki, November.

SHETH, A., 1995, Data Semantics: what, where and how?. Proceedings of the 6th IFIP working Conference on Data Semantics (DS-6). R Meersman and L. Mark (eds), Chapman and Hall, London, UK, 1996.

TRYFONA N., and HADZILACOS, Th., 1998, Logical Data Modeling of SpatioTemporal Applications: Definitions and a Model. Proceedings of the International Database Engineering and Applications Symposium'98 (IDEAS'98), IEEE Press, Cardiff, Wales (to appear).

TRYFONA. N., and SHARMA, J., 1996, On Information Modeling to Support Interoperable Spatial Applications. Lecture Notes in Computer Science 1080. Conference on Advanced Information Systems Engineering (CAISE) Crete, Greece, May 1996.

UTILNETS, [1994-1997]. Work Program of Brite-Euram Project 7120, GD XII, European Union, Brussels, Belgium. Research project of the Computer Technology Institute, Patras, Greece.